Decentralized dynamic control of a multiaccess broadcast channel

Bruce Hajek, T. van Loon

Institutions: University of Illinois at Urbana–Champaign

Published on: 01 Jun 1982 - IEEE Transactions on Automatic Control (IEEE)

Topics: Atomic broadcast, Retransmission, Broadcast radiation and Traffic intensity

Decentralized Dynamic Control of a Multiaccess Broadcast Channel

BRUCE HAJEK, MEMBER, IEEE, AND TIMOTHY VAN LOON

Abstract—**Retransmission policies are presented for the decentralized control of a multiaccess packet-switched broadcast channel. The policies have a simple recursive form yielding a Markov description of the system. Finite average delay is achieved for an infinite-population Poisson arrival model for any rate $\lambda < e^{-1}$.**

It is proposed that the goal of retransmission policies should be to maintain the traffic intensity at a nearly constant, optimum level. The policies we introduce achieve this goal by nearly decoupling the dynamics of the traffic intensity from the backlog fluctuations. Analysis and simulations show that the policies perform well, even when the channel feedback information is unreliable or incomplete.

I. Introduction

The fundamental nature of the time-slotted packet-switched multiaccess broadcast channel is that many remote users compete for the use of a common resource—namely, access to a central broadcast channel. If two or more users each transmit an information packet during a time slot, the packets "collide" and are not successfully broadcast. Such packets join the backlog of packets which must be rebroadcast at a later time. During each time slot, each user possessing a backlogged packet must decide whether or not to transmit the packet in that slot. Since the users are remote from each other, the decision making mechanism must be decentralized.

The information provided to each user is assumed to be the channel output, although in many applications such information will be incomplete or unreliable so that different users may receive different information about the channel status. Moreover, each user knows the history of his own transmissions, while he is not informed of the transmission history of other users. These factors cause different users to obtain distinct channel information. Thus, the multiaccess broadcast channel presents a nontrivial example of a decentralized estimation and control problem.

The purpose of this paper is to present a class of simple retransmission control policies for a multiaccess broadcast channel subject to an infinite population of users. Three important ideas are used in our selection of a control scheme. The first key idea is that the user's main objective should be to maintain a constant level of aggregate traffic. Our view in this paper is that the number of backlogged packets is relevant to a particular user only to the extent that the backlog size affects the channel traffic level. In other words, the product of the channel backlog times the individual user transmission probability is much more fundamental than the size of the backlog itself.

The second key idea of this paper is that there is a large class of retransmission control policies for which the dynamics of the total traffic level is nearly decoupled from the size of the channel backlog, as long as the channel backlog remains positive. In fact, when the size of the backlog is near n, the dynamics of the total channel traffic suffer disturbances of only order $0(1/n)$ due to the fluctuation of the amount of backlogged traffic.

The third key idea of this paper is to use simple first-order recursive retransmission policies. With this choice, a Markov chain with a two-dimensional state space is formed by the state of the control variable together with the size of the channel backlog. Furthermore, the local (or uncoupled) dynamics of the channel traffic intensity are governed by a simple discrete-state Markov chain.

As an application of the bounding technique related to drift analysis presented in [6], it is shown in [6] that stable throughput for any $\lambda < e^{-1} \approx 0.368$ can be achieved by the retransmission control policies introduced in this paper. Moreover, the Markov chain formed by the channel backlog and the control variable can be chosen to take on only countably many values, and then the chain is ergodic (in a strong sense) [6]. Thus, by truncating the state space to a finite set, the stationary probabilities and average delay can be computed to arbitrary accuracy as in [19]. Such a procedure could be justified by the work of Freedman [5]. Here we have chosen other analysis techniques.

Simulation results are given which indicate the effectiveness of the retransmission control policies introduced in this paper. Our retransmission control procedure even performs nearly as well in simulations, and still allows maximum stable throughput e^{-1}, when the users are only allowed to observe the channel output in every fifth slot. Other feedback information reductions are discussed in Section VII. Also, a modified form of the retransmission policies perform well even under severe uncertainties in the input traffic statistics. (See Section VI.)

The work of several other authors is relevant to our study. It is well known that the basic ALOHA multiaccess algorithm (i.e., constant retransmission probabilities) is unstable for the infinite user model unless some auxiliary control is used [7], [3]. Stability has been proved for several control schemes in which the retransmission probabilities are a function of the backlog [3], [4], [8]. However, these schemes cannot be implemented in a decentralized fashion.

Segall [13] obtained the exact recursive equations for estimating the channel backlog given observation of the channel output. For infinitely many users the recursive

Manuscript received October 31, 1980; revised July 13, 1981. Paper recommended by W. S. Levine, Chairman of the Networks and Transportation Systems Committee. This work was supported by the Naval Research Lab under U.S. Navy Contract N00014-80-C-0802.

B. Hajek is with the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801.

T. van Loon was with the Coordinated Science Laboratory, University of Illinois, Urbana, IL. He is now with TRW, Redondo Beach, CA 90278.

0018-9286/82/0600-0559$00.75 ©1982 IEEE

estimation equations are infinite dimensional. This fact and the fact that decentralized information patterns may be observed indicates the need to use suboptimal control policies.

Recently, Mikhailov [10] introduced recursive schemes similar to those presented here. His policies do not include ours, nor do ours include his, although there is a small nonempty intersection. A drift analysis indicates that Mikhailov's schemes yield a maximum stable throughput of 0.364 compared to $e^{-1} \approx 0.368$ for our schemes. We feel that our schemes allow more flexibility and provide more insight for the control and analysis of traffic dynamics. Also, the throughput of his schemes decreases appreciably if 0-1 feedback or similar reduced feedback information is used. (See Section VII.)

Lam and Kleinrock [8] introduced a decentralized retransmission control algorithm called CONTEST. Under this algorithm, the retransmission probabilities are based on traffic estimates obtained by viewing a "window" of time slots. Although this scheme appears difficult to analyze, it did perform well in simulations for a finite-user model [8]. Similar results were obtained by Ban-Tri-An and Gelenbe [1] and Seret and Macchi [15]. The control policies in [1] use (approximations to) the estimators of channel backlog found by Segall [13], whereas the policies in [15] are based on estimates of channel traffic.

An interesting analysis using the concept of delayed information patterns was considered by Schoute [12] and Varaiya and Walrand [18], but in their treatments colliding packets are not retransmitted—instead, a cost is assessed for collisions. Finally, tree algorithms based on the one introduced by Capetanakis, Tsybakov, and Mikhailov [2], [17], [9] also use only channel output information and achieve stable throughputs as high as 0.487. However, such algorithms are much more sensitive to reduced feedback information than are the algorithms presented here. Finally, we mention the works [11], [14], [16], [19] which contain Markov analysis of random access systems in different contexts than ours.

The remainder of this paper is organized as follows. In Sections II and III the infinite-user channel model and the control policies which we consider are introduced. In Section IV we study the local model which is obtained by holding the backlog at any (arbitrary) fixed positive level. In Section V we relate the local model to the original channel/transmission model. In Section VI a modified retransmission policy is suggested which is less sensitive to variations in input traffic statistics. Finally, in Section VII we consider situations in which the feedback information available to users is incomplete or unreliable.

II. The Channel Model and the Local Poisson Approximation

The packet-switched satellite broadcast channel model described in [7, sect. 5.11] will be assumed in this paper. We shall assume that there is an infinite population of users and that new packets continually arrive according to the points of a Poisson point process at an average rate of λ packets/slot. We assume that no particular user ever has more than one packet to transmit. Thus, no queues of packets are formed by a single user. The justification of this simplifying assumption is that if a user station has a group of packets to transmit, then the total access delay is essentially the same as that of the first packet successfully transmitted, for that packet could serve to reserve slots for the remaining packets. Time is normalized so that one packet/slot may be transmitted. For convenience we will not explicitly discuss transmission delays since they are easily incorporated into what we do and they could also be eliminated in a virtual sense through the use of interleaved channels. We are careful not to allow carrier sensing which could not be accomplished because of transmission delays.

Throughout this paper, t will be integer valued. Fix $\lambda > 0$, and let Y_t be the number of packets which first arrive during the slot ending at time t. Thus, the random variables $(Y_t)_{t\in Z}$ are independent with distribution Poisson (λ). Except in Section VI, we assume that the Y_t packets which arrive during slot $(t-1, t]$ are first transmitted in slot $(t, t+1]$, which is the first complete slot after their arrival. Let N_t denote the channel backlog (which does not include the new arrivals) at time t and let Z_t represent the channel output during slot $(t, t+1]$. Thus $Z_t = 0$, $Z_t = 1$, or $Z_t = e$, depending on whether zero, one, or more than one transmissions were attempted during slot $(t, t+1]$. Except in Section VII, we shall only consider retransmission control strategies in which, conditioned on the history of the channel up to time t, each user with a backlogged packet independently transmits the packet with the same probability f_t in the slot $(t, t+1]$. Thus the random sequence $(f_t)_{t\in Z}$ represents the retransmission control policy. The sequence (f_t) must satisfy $0 \leqslant f_t \leqslant 1$ and f_t must be a function of the channel output history $(Z_s\colon s<t)$ for each t. As indicated in Section VII, our retransmission policy need not require this much information.

The nonzero channel output and backlog transition probabilities are as follows.

Actual Channel Probabilities:

$$P[N_{t+1}=k, Z_t=0 \mid N_t=n, f_t=f] = e^{-\lambda}(1-f)^n \qquad \text{if } k=n$$

$$P[N_{t+1}=k, Z_t=1 \mid N_t=n, f_t=f] = \begin{cases} e^{-\lambda} nf(1-f)^{n-1} & \text{if } k=n-1 \\ \lambda e^{-\lambda}(1-f)^n & \text{if } k=n \end{cases}$$

$$P[N_{t+1}=k, Z_t=e \mid N_t=n, f_t=f] = \begin{cases} e^{-\lambda}\left[1-(1-f)^n - nf(1-f)^{n-1}\right] & \text{if } k=n \\ \lambda e^{-\lambda}\left[1-(1-f)^n\right] & \text{if } k=n+1 \\ \dfrac{\lambda^j e^{-\lambda}}{j!} & \text{if } k=n+j,\ j>1. \end{cases}$$

Denote the intensity of retransmitted traffic at time t by $\mu_t \triangleq N_t f_t$. Let $G_t = \lambda + \mu_t$ be the total traffic intensity at time t and let S_t be the expected throughput during slot $(t, t+1]$, given N_t and f_t. Thus

$$S_t \triangleq \lambda e^{-\lambda}(1-f_t)^{N_t} + e^{-\lambda}(1-f_t)^{N_t - 1} N_t f_t I_{\{N_t \neq 0\}}$$

where I_A denotes the indicator function of an event A.

A key in our analysis of the broadcast channel model is an approximation to the output and backlog transition probabilities which is based on the following proposition.

Proposition 2.1 (Poisson Approximation to Binomial Probabilities):

1) If $0<f<1$, $n\geq 0$, and $0<d<\frac{1}{2}$, then

$$0\leq e^{-nf}-(1-f)^n$$
$$\leq e^{-nf}\left[1-\exp\left(-nf^2\left(\frac{1}{2(1-f)}\right)\right)\right] \quad (2.1)$$
$$\leq e^{-n^d}+1-\exp\left(-n^{2d-1}\left(\frac{1}{2(1-n^{d-1})}\right)\right) \quad (2.2)$$

and if $f\leq\frac{1}{2}$, then

$$e^{-nf}-(1-f)^n\leq e^{-nf}(1-e^{-nf^2})\leq f. \quad (2.3)$$

2) If $0<f<1$, $n\geq 1$, and $0<d<\frac{1}{2}$, then

$$|nfe^{-nf}-nf(1-f)^{n-1}|$$
$$\leq nfe^{-nf}\left[1-\exp\left(-nf^2\frac{1}{2(1-f)}\right)\right]+nf^2e^{-(n-1)f} \quad (2.4)$$
$$\leq n^de^{-n^d}+1-\exp\left(-n^{2d-1}\frac{1}{2(1-n^{d-1})}\right)+\frac{4e}{n} \quad (2.5)$$

and if $f\leq\frac{1}{2}$, then

$$|nfe^{-nf}-nf(1-f)^{n-1}|\leq f^{-d}e^{-f^{-d}}+1-\exp(-f^{1-d})+f. \quad (2.6)$$

Note the right-hand sides of (2.2) and (2.5) tend to zero as $n\to\infty$ and the right-hand sides of (2.3) and (2.6) tend to zero as $f\to 0$. Proposition 2.1 is proved in Appendix A. Using Proposition 2.1, the following approximation to the channel output probabilities is obtained.

Approximate Channel Probabilities (Let $\mu=nf$, $G=\lambda+\mu$):

$$P[N_{t+1}=k, Z_t=0|N_t=n, f_t=f]=e^{-G} \qquad \text{if } n=k$$

$$P[N_{t+1}=k, Z_t=1|N_t=n, f_t=f]=\begin{cases} e^{-\lambda}\mu e^{-\mu} & \text{if } k=n-1\\ \lambda e^{-\lambda}e^{-\mu} & \text{if } k=n\end{cases}$$

$$P[N_{t+1}=k, Z_t=e|N_t=n, f_t=f]=\begin{cases} e^{-\lambda}[1-(1+\mu)e^{-\mu}] & \text{if } k=n\\ \lambda e^{-\lambda}[1-e^{-\mu}] & \text{if } k=n+1\\ \dfrac{\lambda^j e^{-\lambda}}{j!} & \text{if } k=n+j,\ j>1.\end{cases}$$

By Proposition 2.1, the error of these approximations approaches zero uniformly if either $f\to 0$ or $n\to\infty$. In addition $|S_t-G_te^{-G_t}|\to 0$ uniformly as either $f_t\to 0$ or $N_t\to\infty$ since $G_te^{-G_t}=P[Z_t=1|N_t, f_t]$ when the above approximate transition probabilities are used.

The approximate channel output and backlog transition probabilities given above would be exact if given (N_t, f_t) the conditional distribution of the number of retransmitted packets in slot $(t, t+1]$ is Poisson with mean μ_t. This approximation shall be called the *local* Poisson approximation, which is justified by the Poisson approximation to the binomial distribution. The local Poisson traffic approximation should not be confused with the Poisson traffic assumption under which the transmitted traffic stream is assumed to be a Poisson process (with constant intensity). An important observation regarding the approximate channel probabilities is that they depend on N_t and f_t only through the product $\mu_t=N_tf_t$. This fact will play a key role in the following sections.

III. Control Procedures Which Decouple the Traffic Dynamics from the Backlog

Certain retransmission control procedures cause the dynamics of the intensity of channel traffic to be nearly decoupled from the channel backlog and individual retransmission probabilities. This is desirable since, by the local Poisson approximation, the throughput intensity at time t is $S_t\cong G_te^{-G_t}$, which depends only on the traffic intensity and not directly on the backlog or individual retransmission probabilities. Our goal is to keep G_t near its optimum value, namely $G_t=1$, no matter what the backlog level N_t is, as long as $N_t\neq 0$. In this section we will first derive the desired form of the control strategy. Then the "local model" for the traffic intensity and its relationship with the overall channel model will be identified.

For simplicity, consider a first-order recursive retransmission control policy of the form

$$f_{t+1}=G(f_t, Z_t) \quad (3.1)$$

for each t for some function G: $[0,1]\times\{0,1,e\}\to[0,1]$. When this retransmission control rule is used, the pair process (N_t, f_t) forms a Markov chain with state space $Z_+\times[0,1]$. The transition probabilities for the chain are determined by the channel probabilities given in Section II (using the local Poisson approximation if desired) and by (3.1).

By (3.1), the intensity μ_{t+1} of the retransmitted traffic in slot $(t+1, t+2]$ satisfies

$$\mu_{t+1}=\frac{N_{t+1}}{N_t}\left[\frac{G(f_t, Z_t)}{f_t}\right]\mu_t \qquad \text{if } \mu_t>0. \quad (3.2)$$

Now, define $\Delta N_t=N_{t+1}-N_t$. Then

$$\Delta N_t=Y_t-I_{\{Z_t=1\}}$$

where Y_t is Poisson with mean λ and is independent of N_t. Thus

$$P[|\Delta N_t|\leq k|N_t=n]\geq P_\lambda(k) \qquad \text{for } k\geq 1,\ n\geq 0 \quad (3.3)$$

where $P_\lambda(\cdot)$ is the distribution function of a Poisson random variable with mean λ. Inequality (3.3) shows that ΔN_t is stochastically dominated, even when conditioned on $N_t = n$, uniformly in t and in n. Hence, given N_t, $\Delta N_t / N_t$ is "stochastically $0(1/N_t)$." Substituting this into (3.2) yields

$$\mu_{t+1} = \left(1 + 0\left(\frac{1}{N_t}\right)\right) \frac{G(f_t, Z_t)}{f_t} \mu_t. \tag{3.4}$$

The probability distribution of Z_t given (f_t, μ_t) depends only on μ_t if the local Poisson approximation is used. Thus, except for the term $0(1/N_t)$, the conditional distribution of the right-hand side of (3.4) given (f_t, μ_t) will depend on μ_t (but not otherwise on f_t) if and only if $G(f_t, z_t)/f_t$ does not depend on f_t. Equivalent conditions are $G(\alpha f, z) = \alpha G(f, z)$ for $\alpha > 0$ or $G(f, z) = a(z) f$ for some vector $\boldsymbol{a} = (a(0), a(1), a(e))$.

Due to the requirement that $0 \leq f_t \leq 1$ for all t, it is too much to ask that $G(\alpha f, z) = \alpha G(f, z)$ for all α, f. We shall thus consider estimators of the type $f_{t-1} = G(f_t, Z_t)$ such that $G(\alpha f, z) = \alpha G(f, z)$ if and only if αf and f are smaller than some $\beta > 0$. In fact the estimators examined in the following sections are of the form

$$f_{t+1} = \min\left(a(Z_t)^{\gamma} f_t, \beta\right) \tag{3.5}$$

for some positive constants γ, β, $a(0)$, $a(1)$, $a(e)$.

A more general class of retransmission control strategies have the form $f_{t+1} = G(Z_t, \cdots, Z_{t-k}, f_t, \cdots, f_{t-k})$. The vector process $V_t = (Z_t, \cdots, Z_{t-k}, f_t, \cdots, f_{t-k}, N_t)$, then forms a Markov chain. If the function G satisfies

$$G(z_0, \cdots, z_k, \alpha f_0, \cdots, \alpha f_k) = \alpha G(z_0, \cdots, z_k, f_0, \cdots, f_k)$$

whenever $\max(f_0, \cdots, f_k, \alpha f_0, \cdots, \alpha f_k) \leq \beta$ for some $\beta > 0$, then the analysis of the following sections carries over to this more general type of retransmission control policy.

In the remainder of this paper, unless otherwise specified, we shall always deal with the retransmission control strategy (3.5), where the constants γ, β, and $\boldsymbol{a} = (a(0), a(1), a(e))$ will be specified later.

For purposes of analysis it is convenient to define the new variable $\varphi_t = \ln(f_t(N_t \vee 1))$ where "$\vee$" denotes maximum. Thus when $N_t > 0$, $\mu_t = \exp(\varphi_t)$ is the retransmitted traffic intensity at time t. Since (f_t, N_t) and (φ_t, N_t) are functions of each other, (φ_t, N_t) forms a Markov sequence. The following paragraphs contain a brief digression on localization of a vector Markov process. The method will be applied to show that the local behavior of φ_t (and hence μ_t) is independent of N_t as long as N_t is not too small.

Localization of Vector Markov Chains: Suppose that (X_t, Y_t) is a Markov chain with state space $\mathfrak{X} \times \mathfrak{Y}$ and transition probability kernel

$$p_{XY}(dx\, dy | x_0, y_0), \qquad (x_0, y_0) \in \mathfrak{X} \times \mathfrak{Y}.$$

For any fixed $\bar{y} \in \mathfrak{Y}$, a new Markov chain $(X_t^{\bar{y}})$ with state space $\mathfrak{X}$ is defined as the Markov process associated to the transition probability kernel $p_{X\bar{y}}$ where

$$p_{X\bar{y}}(dx | x_0) = p_{XY}(dx \times \mathfrak{Y} | x_0, \bar{y}).$$

The new process $(X_t^{\bar{y}})$ is called the localization of the Markov process (X_t, Y_t) to the region where (Y_t) is identically equal to $\bar{y}$.

For most examples, the transition probabilities of $(X_t^{\bar{y}})$ depend on $\bar{y}$. If (Y_t) is "slowly varying" relative to (X_t) and if the transition probabilities $p_{X\bar{y}}$ are sufficiently regular with respect to $\bar{y}$, then the behavior of the process $(X_t^{\bar{y}})$ should closely resemble the behavior of the component (X_t) during those periods when y_t is close to $\bar{y}$.

The Local Model for Retransmitted Traffic: Fix an integer $n > 0$. The Markov chain (f_t^n) obtained by localizing the Markov chain (f_t, N_t) to $(N_t) \equiv n$ has the transition probabilities (using the local Poisson approximation)

$$f_{t+1}^n = \min\left(\beta, f_t^n \times \begin{cases} a(0)^{\gamma} \text{ with prob } e^{-G_t} \\ a(1)^{\gamma} \text{ with prob } G_t e^{-G_t} \\ a(e)^{\gamma} \text{ with prob } 1-(1+G_t)e^{-G_t} \end{cases}\right) \tag{3.6}$$

where $G_t = \lambda + n f_t$.

If $\varphi_t^n \equiv \log(n f_t^n)$, then $\exp(\varphi_t^n)$ represents the total retransmitted traffic intensity for the localized model. Further, (φ_t^n) is itself a Markov process with transition probabilities

$$\varphi_{t-1}^n = \min\left(\ln(n\beta), \varphi_t^n + \gamma \begin{cases} c(0) \text{ with prob } e^{-G_t} \\ c(1) \text{ with prob } G_t e^{-G_t} \\ c(e) \text{ with prob } 1-(1+G_t)e^{-G_t} \end{cases}\right) \tag{3.7}$$

where $c(i) = \log a(i)$ for $i = 0, 1$ or e. The process φ_t^n will be referred to as the local model (n) for the log retransmitted traffic intensity. The important point here is that the transition probabilities of φ_t^n do not depend on n, except through the term $\ln(n\beta)$ which will be discussed in the next section. We have thus shown that *when N_t is held at a constant level $n > 0$ (for example, by adding or subtracting blocked users at the end of each time slot as needed), then the total traffic intensity forms a Markov chain whose transition probabilities do not depend on n* except through the value of the upper boundary $\ln(n\beta)$. This fact, of course, depends on our use of the local Poisson approximation. If the actual transition probabilities are used, then the transition probabilities for (φ_t^n) will depend on n, but will converge rapidly to those of (3.7) as n increases. The Markov chain (φ_t^n) whose transition probabilities (3.7) do not depend on n will be referred to as the local model for the log intensity of retransmitted traffic.

Retransmitted Traffic—Global Model as Disturbed Local Model: For contrast, the Markov chain (N_t, f_t) or equivalently, the chain (N_t, φ_t) with $\varphi_t = \ln(f_t(N_t \vee 1))$, will be called the global model as opposed to the local model just described in which N_t is held fixed. Defining the random variables

$$\theta_t = \ln\left(f_t(N_{t-1} \vee 1)\right)$$

and

$$\xi_t = \ln(N_t \vee 1) - \ln(N_{t-1} \vee 1)$$

we see that

$$\varphi_t = \theta_t + \xi_t \tag{3.8}$$

and that the conditional distribution of θ_{t+1} given $(N_t, \varphi_t = \varphi)$ using the global model is the same as the conditional distribution of φ_{t+1}^n given $\varphi_t^n = \varphi$ using the local model. Hence, (3.8) represents φ_t for the global model as a perturbation of the corresponding variable for the local model, where the perturbation at time $t+1$ is $\xi_{t+1} = \Delta \ln(N_t \vee 1) \approx \Delta N_t / N_t \sim 0(1/N_t)$ for N_t large. We shall return to these ideas in the sections that follow.

Choice of the Parameter β: The parameters $c(0)$, $c(1)$, $c(e)$, and γ of our retransmission control procedure are chosen in the next section based on the local model. The choice of β is based on consideration of the global model. As seen from the recursive control equation (3.5), the constant β is the maximum value that the retransmission probabilities f_t can assume. A reasonable choice for β is $\beta = 1$, although we have found that a somewhat smaller value of β is preferable. In fact, during periods when $N_t = 0$, the users will not observe any collisions until two new arrivals occur in a single slot. The lengths of such idle periods thus have a geometric distribution with mean $(1-(1+\lambda)e^{-\lambda})^{-1}$, which is on the order of 20 slots for $\lambda < e^{-1}$. During these periods, since low channel traffic is being observed, the users' retransmission probabilities f_t will tend to drift to their maximum value β (see below). At the end of the "idle" period, there will be at least two users contending for the channel. Thus, a good choice for β is the retransmission level f which maximizes the throughput given that there are two users in contention; that is,

$$\beta = \arg\max_f \left\{ e^{-\lambda} 2f(1-f) + \lambda e^{-\lambda}(1-f)^2 \right\} = \frac{1-\lambda}{2-\lambda}. \tag{3.9}$$

This choice of β will be assumed throughout the remainder of this paper.

IV. Drift and Variance Analysis of the Local Model

The local model for the log intensity of the retransmitted traffic when N_t is fixed to be n is given by (3.7). The minimum in (3.7) simply reflects the fact that $\varphi_t^n = \ln(nf_t^n) \le \ln(n\beta)$ for all t. However, as n increases, the constraint $\varphi_t^n \le \ln(n\beta)$ becomes less crucial as will be seen below. Hence, in this section we study only the process $\hat{\varphi}_t$ with transition probabilities

$$\hat{\varphi}_{t+1} = \hat{\varphi}_t + \begin{cases} \gamma c(0) \text{ with prob } e^{-G_t} \\ \gamma c(1) \text{ with prob } G_t e^{-G_t} \\ \gamma c(e) \text{ with prob } 1-(1+G_t)e^{-G_t} \end{cases} \tag{4.1}$$

where $G_t = \lambda + \exp(\hat{\varphi}_t)$. Thus, $(\hat{\varphi}_t)$ has the same transition probabilities as the local model (φ_t^n), except the constraint to be less than $\ln(n\beta)$ is not imposed on $(\hat{\varphi}_t)$.

For numerical and theoretical purposes it is important to note that if $(c(0), c(1), c(e)) = (i\delta, j\delta, j\delta)$ for some constant $\delta > 0$ and integers i, j, k with greatest common divisor 1, then $\hat{\varphi}$ may be viewed as an irreducible aperiodic Markov chain with the discrete state space consisting of integer multiples of δ.

Choice of the Parameters $c(0), c(1), c(e)$: Define the quantities

$$m(x) = E[\Delta\hat{\varphi}_t | \hat{\varphi}_t = x]$$

and

$$\sigma^2(x) = \mathrm{var}[\Delta\hat{\varphi}_t | \hat{\varphi}_t = x]$$

where

$$\Delta\hat{\varphi}_t = \hat{\varphi}_{t+1} - \hat{\varphi}_t.$$

It is clear from (4.1) that

$$m(x) = \gamma(c(0), c(1), c(e)) \cdot b(x)$$

and

$$\sigma^2(x) = \gamma^2 \left(c(0)^2, c(1)^2, c(e)^2\right) \cdot b(x) - m(x)^2$$

where

$$b(x) = \left(e^{-G(x)}, G(x)e^{-G(x)}, 1-(1+G(x))e^{-G(x)}\right)$$

for $G(x) = e^x + \lambda$. The function m is the local drift and σ^2 is the local variance of the real-valued Markov chain $(\hat{\varphi}_t)$. Now, by the local Poisson approximation, the expected throughput in slot $[t, t+1)$ given $\hat{\varphi}_t$ is $S_t = G_t e^{-G_t}$ where $G_t = G(\hat{\varphi}_t)$, which is maximized when $G_t = 1$. Thus, it is reasonable to choose the constants $c(0), c(1), c(e)$ so that $\hat{\varphi}_t$ always drifts towards the value φ_{opt} which yields $G_t = \exp(\hat{\varphi}_t) + \lambda = 1$. That is, $\boldsymbol{c} = (c(0), c(1), c(e))$ should be chosen so that

$$\mathrm{sgn}(G(x)-1) = -\mathrm{sgn}(m(x)) \qquad \text{for all } x \in \mathbb{R} \tag{4.2}$$

where $\mathrm{sgn}(u) = +1, 0$ or -1 when $u > 0$, $u = 0$, or $u < 0$, respectively.

Proposition 4.1: Condition (4.2) is true if and only if $\boldsymbol{c} \ne 0$, $c(e) \le 0 \le c(0)$, and

$$c(e) = -(c(0) + c(1)) \left[\frac{e^{-1}}{1-2e^{-1}} \right].$$

(See Appendix B for proof.)

Proposition 4.1 shows that there are many choices for $\boldsymbol{c}$ which ensure that the traffic level drifts in the right direction. The conditions $c(e) \le 0 \le c(0)$ are intuitively pleasing, for they show that the retransmission probabilities should be decreased (respectively increased) when a collision (respectively empty slot) is observed. The constant $c(1)$ can be positive or negative. There is a unique choice of $\boldsymbol{c}$ (up to a multiplicative constant) which satisfies (4.2) and also satisfies $c(1) = 0$. It is

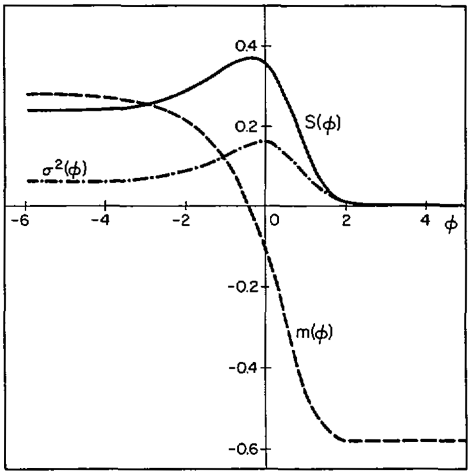


Fig. 1. Drift, covariance, and throughput intensity versus φ for local model, $\lambda = 0.32$.

$$c = (c(0), c(1), c(e)) = (0.418, 0, -0.582).$$

This choice of c will be used throughout the rest of this paper. Some choices of c which allow reduced feedback information are discussed in Section VII.

In Fig. 1, $m(\varphi), \sigma^2(\varphi)$, and throughput

$$S(\varphi) = (\lambda + e^{\varphi})\exp(-(\lambda + e^{\varphi}))$$

are plotted for $\gamma = 1$ and $\lambda = 0.32$. Note that $\hat{\varphi}$ tends to drift towards the value $\varphi_{\text{opt}} = \ln(1-\lambda)$ which maximizes the throughput level. Moreover, m is bound away from zero outside any neighborhood of its zero, $\ln(1-\lambda)$. Since the step sizes of $\hat{\varphi}$ are uniformly bounded, if $\hat{\varphi}$ is considered as a discrete state (see above) Markov chain, then there exists a unique stationary distribution and the time it takes for $\hat{\varphi}$ to get from one state to any other state has a distribution with a geometrically bounded tail [6].

The Role of the Parameter γ*:* A glance at (4.1) shows that the step sizes of the Markov chain $(\hat{\varphi}_t)$ are proportional to γ. If γ is fairly large, $\hat{\varphi}$ will quickly approach the vicinity of its optimal value, but will continue to fluctuate so that the expected value of throughput $S(\hat{\varphi}_t)$ when $\hat{\varphi}_t$ has its invariant distribution is somewhat less than $S(\varphi_{\text{opt}}) = e^{-1}$. If γ is very small, then $\hat{\varphi}$ approaches φ_{opt} more slowly, but on the average remains closer to φ_{opt} after the longer transient period. It is known (see [6, equations (5.11) and (5.12)]) that

$$\lim_{\gamma \to 0} \bar{S}_t = e^{-1}$$

where $\bar{S}_t = E_S[S(\hat{\varphi}_t)]$ denotes the expectation of $S(\hat{\varphi}_t)$ relative to the stationary distribution of $(\hat{\varphi}_t)$. Moreover, it can be shown that

$$\bar{S}_t = e^{-1} - \gamma^2 R + o(\gamma^2) \tag{4.3}$$

where

$$R = \left(-\frac{\sigma^2}{2m'}\right)\left(\frac{1}{2}S''\right)\bigg|_{\varphi = \varphi_{\text{opt}}} \approx 0.308(1-\lambda).$$

To understand (4.3), note that near $\varphi = \varphi_{\text{opt}}$, m, σ, and S are approximated by $\tilde{m}(\varphi) = m'(\varphi_{\text{opt}})(\varphi - \varphi_{\text{opt}})$, $\tilde{\sigma}(\varphi) = \sigma(\varphi_{\text{opt}})$, and $\tilde{S}(\varphi) = e^{-1} - \frac{1}{2}\tilde{S}''(\varphi_{\text{opt}})$. The diffusion with drift $\tilde{m}$ and diffusion term $\tilde{\sigma}^2$ is a translation of an Ornstein–Uhlenbeck process which thus has the stationary distribution $N(\varphi_{\text{opt}}, -\tilde{\sigma}^2/2m'(\varphi_{\text{opt}}))$. The approximation $e^{-1} - \gamma^2 R$ of $\bar{S}_t$ above is the expected value of $\tilde{S}(\varphi)$ with respect to this stationary distribution.

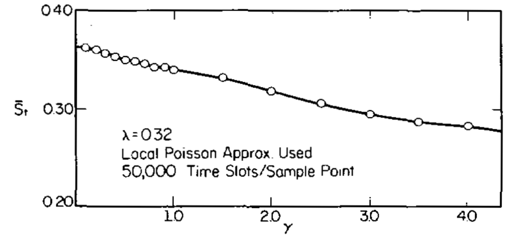


Fig. 2. Average throughput of local model versus γ; $\lambda = 0.32$.

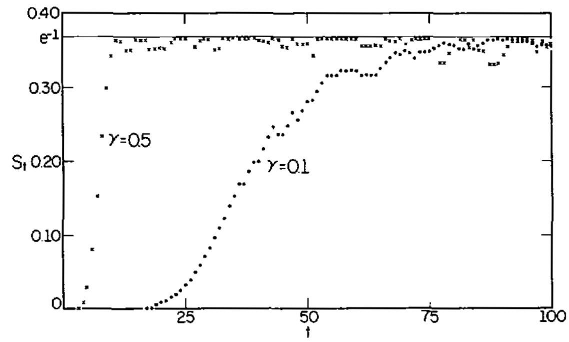


Fig. 3. Sample paths of throughput intensity using local model; $\lambda = 0.32$.

In Fig. 2 $\bar{S}_t$ is plotted as a function of γ. The results were obtained by computer simulation runs using 50000 time slots for each data point. Fig. 3 shows two sample paths of the throughput intensity $S_t = S(\hat{\varphi}_t)$ when $\gamma = 0.1$ and $\gamma = 0.5$ are obtained by computer simulation.

V. Stability and Packet Delay Under the Global Model

The average throughput rate $\bar{S}_t$ under the local model can be made greater than any fixed input rate $\lambda < e^{-1}$ by choosing γ small enough. The dynamics of the throughput intensity $S(\varphi_t)$ under the global model are the same as those of s_t under the local model, except for a disturbance of $0(1/N_t)$ as indicated by (3.8). Thus, when a large enough backlog occurs, the throughput rate $S(\varphi_t)$ will have a quasi-equilibrium distribution with a mean greater than any fixed input rate $\lambda < e^{-1}$ if γ is sufficiently small. This ensures that the backlog N_t will tend to drift downward. This is the intuition behind the following theorem, which is proved in [6]. In the theorem, the process (N_t, f_t) is considered as a Markov chain with the discrete state space

$$S = \{(n, \beta\Delta^k): n, k \text{ nonnegative integers}\}$$

where $0 < \Delta < 1$ and $a^{\gamma} = (\Delta^i, 1, \Delta^{-j})$ for some relatively prime integers i, j.

Theorem 5: For any input rate $\lambda < e^{-1}$, there is a $\gamma^* > 0$ such that if $0 < \gamma < \gamma^*$, then the following are true: 1) The

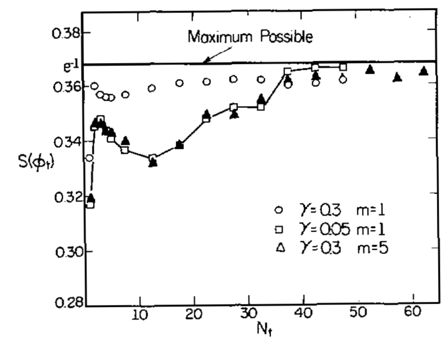


Fig. 4. Average throughput intensity versus backlog using local Poisson approximation to global model.

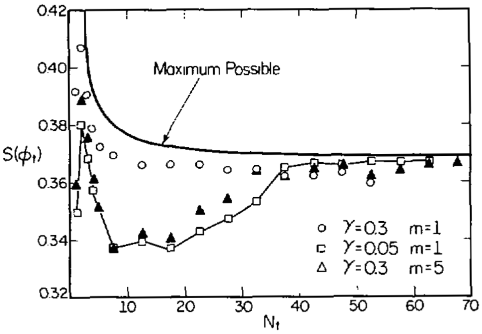


Fig. 5. Average throughput intensity versus backlog using actual global model.

discrete state Markov chain (N_t, f_t) is geometrically ergodic. (In particular, the chain is positive recurrent with an invariant probability measure π.) 2) $E_\pi[(N_k)^m]<\infty$ for all $m>0$, where E_π denotes expectation relative to π.

In view of this theorem, the next problem is to find the expected delay D sustained by a randomly arriving packet and to select γ to minimize D. By Little's result, if $\overline{N_t}$ denotes $E_\pi[N_k]$, then

$$D=\overline{N_t}/\lambda$$

so the study of D and the study of $\overline{N_t}$ are equivalent. Unfortunately, it is difficult to compute $\overline{N_t}$ or even find close bounds. Our approach here will be to combine computer simulation with analytical analysis.

To get a rough estimate of $\overline{N_t}$, the effect of the disturbance ξ_t in (3.8) on the throughput intensity $S(\varphi_t)$ under the global model should be analyzed as a function of the backlog N_t. In Figs. 4 and 5, the average of the throughput intensity $S(\varphi_t)$ is plotted as a function of the backlog N_t. The results were obtained by a computer simulation of the global model over a period of 50000 time slots. As indicated in Fig. 6, the sample average of $S(\varphi_t)$ was computed over sets of time slots for which the backlog was at specified levels. For example, Fig. 6 indicates that during 3877 out of 50000 time slots, the backlog was 2 packets, and the average of $S(\varphi_t)$ over these 3877 slots was 0.36006. Fig. 5 differs from Fig. 4 only in that the actual channel transition probabilities were used rather than the local Poisson approximations. The maximum throughput intensity that could be achieved if all users were informed of N is also indicated in Figs. 4 and 5. The parameter m

Backlog Level	m=1 Number of Occupations	m=1 $\bar{S}_t$	m=5 Number of Occupations	m=5 $\bar{S}_t$
0	13723	0.23236	8875	0.23236
1	3299	0.33689	2695	0.31966
2	3877	0.36006	3327	0.34662
3	3221	0.35766	2774	0.34649
4	2930	0.35612	2418	0.34377
5	2636	0.35600	1924	0.33652
5<N≤10	9355	0.35726	7863	0.33086
10<N≤15	5398	0.35982	6247	0.33196
15<N≤20	2095	0.36100	4339	0.33886
20<N≤25	1103	0.36147	3052	0.34905
25<N≤30	900	0.36265	2154	0.34966
30<N≤35	620	0.36241	1719	0.35580
35<N≤40	355	0.36037	908	0.36283
40<N≤45	282	0.36077	627	0.36356
45<N≤50	191	0.36188	621	0.36250
50<N≤55	15	0.34649	288	0.36518
55<N≤60			147	0.36225
60<N≤65			22	0.36487

Fig. 6. Average throughput intensity versus backlog using local Poisson approximation to global model with $\lambda=0.32$ and $\gamma=0.3$.

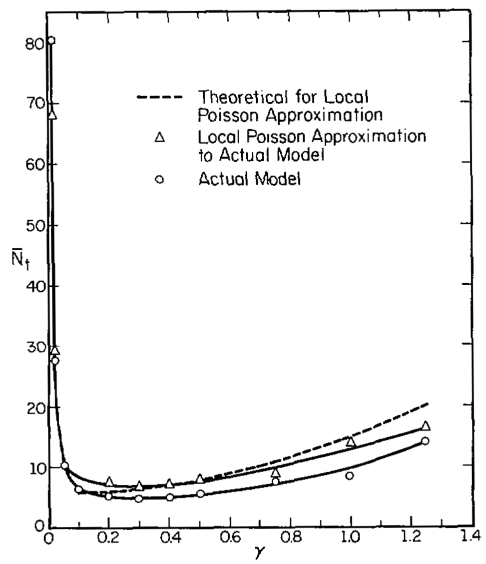


Fig. 7. Average backlog for global model versus γ. Computer simulation —50000 slots/point, $\lambda=0.32$.

indicated in the figures is discussed in the next section—for the model discussed here $m=1$.

For γ very small we conjecture that

$$\overline{N_t}\sim C_\lambda/\gamma \qquad \text{as } \gamma\to 0 \tag{5.1}$$

for some constant C_λ. The reason for this conjecture is that for γ small, the control procedure will be effective when and only when the size of the steps [which are $0(1/N_t)$] of the process ξ_t is comparable or smaller than the size of the steps of $\hat{\varphi}_t$ (which are proportional to γ). The relation (5.1) is apparent in Fig. 7 for $\lambda=0.32$ and $C_\lambda=0.6$. Fig. 7 was obtained by computer simulation of the global model for 50000 time slots for each data point.

For γ greater than about 0.3, the throughput intensity $S(\varphi_t)$ of the global model appears to be nearly the same for all $N_t\geqslant 2$. This is indicated by the curve for $\gamma=0.3$ in Fig. 4. The reason for this insensitivity to N_t is that the disturbance ξ_t is no more severe than the fluctuations of $\hat{\varphi}_t$ in the local model when the step size of $\hat{\varphi}_t$ is large enough. For γ greater than about 0.3, the throughput intensity $S(\varphi_t)$ is thus well approximated by the average $\bar{S}_t$ computed using the local model and shown in Fig. 2. This approximation leads to a model in which the throughput intensity is a constant independent of the backlog N_t as long as $N_t > 0$. Such a model was analyzed by Ferguson [4]. He was able to compute all moments of N_t with respect to the invariant distribution. His analysis is extended in Appendix C to yield an approximation of $\bar{N}_t$ as a function of γ when the intensity $S(\varphi_t)$ is approximated by a constant $\bar{S}$ whenever $N_t > 0$. This yields an approximation of $\bar{N}_t$ as a function of γ using the values of $\bar{S}_t$ versus γ as pictured in Fig. 2. The results of the calculations are plotted in Fig. 7.

It is apparent from Fig. 7 that the best choice of γ is approximately 0.3 and for this value of γ the average backlog is about 5 packets using the actual channel transition probabilities.

Given Theorem 5, the question arises as to what other recursive policies yield a stable system. Using the ideas of this paper, it is not difficult to establish the stability result claimed in [10] for the policies considered there. It is not practical to give a result that would cover all recursive estimators which yield stability, although the principles used in proving Theorem 5 can be used more generally as we will briefly describe. (Also, see [20].)

For a given recursive policy $f_{t+1} = G(f_t, Z_t)$ which may not be of multiplicative type, the local model (n), φ_t^n, can still be defined as in Section III, and the steady-state throughput intensity $\bar{S}_n = \liminf_{t\to\infty} E_n[S(\varphi_t)]$ can be defined. Roughly speaking, the channel will be stable if 1) $\lambda < \liminf_{n\to\infty} \bar{S}(n)$ and if 2) the statistics of local model (n) vary slowly enough with n for n large so that even when $n = N_t$ fluctuates as it does in the global model, the process φ_t still drifts to maintain average throughput intensity greater than γ. These conditions are readily tested for multiplicative policies since for such policies the statistics of the local model (n) depend on n only through the boundary $\ln(\beta n)$. [See (3.7).]

VI. A Policy Insensitive to Input Traffic Statistics

In practice the statistics of the input traffic process (Y_t) may significantly differ from Poisson statistics. Not only might each of the variables Y_t have a different distribution, the variables may be dependent. This may severely degrade the performance of the retransmission policy we have proposed since statistical changes in incoming traffic *directly* change the statistics of the total transmitted traffic. The reason for this direct influence is that our policy is an immediate-first-transmission (IFT) policy (notation suggested by Tobagi [16]). However, as we will show in this section, our policy (and the resulting analysis) can be readily converted to a delayed-first-transmission (DFT) policy which maintains stability over a wide class of input statistics with rate $\lambda < e^{-1}$.

Under the DFT policy, new packets join the backlog before their first transmission is attempted. Thus, the number of backlogged packets for the DFT policy satisfies

$$N_{t+1} = N_t + Y_{t+1} - I_{\{Z_t = 1\}}$$

where (Y_t) and (Z_t) are defined as before. Given (N_t, f_t), the throughput probability S_t and the total traffic intensity G_t satisfy

$$S_t = (1-f_t)^{N_t - 1} N_t f_t I_{\{N_t \neq 0\}}$$

and

$$G_t = \mu_t = N_t f_t.$$

Note that $|S_t - G_t e^{-G_t}| \to 0$ uniformly as $N_t \to +\infty$ or as $f_t \to 0$, by (2.2) and (2.3) as before. Consider the same rule (3.5) for updating retransmission probabilities as before. The parameters $a(0)$, $a(1)$, $a(e)$, and γ should be chosen as before, but a more appropriate choice for β is now $\beta = 1$. (See discussion of β at end of Section III.) Since $S_t \approx G_t e^{-G_t}$ as before, it is clear that the local model for the DFT policy is identical to the local model discussed in Section IV for the original IFT policy, at least when the arrival process (Y_t) is Poisson.

However, if, for example, the variables Y_t are independent but have a non-Poisson distribution, then the local model for the IFT policy changes, while the local model for the DFT policy just described is the same as it is for Poisson arrivals.

For example, suppose that $P(Y_t = 2) = \lambda/2$ and $P(Y_t = 0) = 1 - \lambda/2$. If N_t is large, then under the IFT policy the probability distribution of $Z_t \in \{0, 1, e\}$ is approximately

$$\left(\left(1-\frac{\lambda}{2}\right)e^{-\mu_t}, \left(1-\frac{\lambda}{2}\right)\mu_t e^{-\mu_t}, 1-\left(1-\frac{\lambda}{2}\right)(1+\mu_t)e^{-\mu_t}\right).$$

In particular, the throughput probability is $(1 - (\lambda/2))\mu_t e^{-\mu_t}$. This distribution of Z_t is different than when Y_t is Poisson—hence the transmitted traffic level need no longer drift to the optimum level, which is now $G_t = 1 + \lambda$ (so $\mu_t = 1$). Furthermore, even if the policy parameters are readjusted so that G_t drifts properly, the resulting maximum throughput is $(1-(\lambda/2))e^{-1}$, so in order for the channel to be stable it is required that $\lambda < (1-(\lambda/2))e^{-1}$, or $\lambda < 0.31$. On the other hand, the DFT policy maintains stability for this distribution of (Y_t) as long as $\lambda < e^{-1}$, and no readjustment of the constants $a(0)$, $a(1)$, and $a(e)$ is needed.

In certain cases, the IFT transmission scheme can achieve greater stable throughput than the DFT scheme. In fact, it is not hard to show that if (and only if) $\lambda < \frac{1}{2}$, then there exists a distribution for the arrival variables Y_t with $E[Y_t] = \lambda$ and $P(Y_t > 1) > 0$ such that the channel is stable under the original retransmission policy for some choice of parameters γ, $a(0)$, $a(1)$, and $a(e)$ *depending on the distribution of* Y_t.

Thus, if the policy parameters can be adjusted to depend on the arrival distribution, then neither the DFT nor the IFT policy always provides a larger maximum throughput than the other. However, the crucial point is that only the DFT policy can ensure stability (in the sense of Theorem 5.1) for any i.i.d. arrival process (Y_t) with $\lambda = E(Y_t) < e^{-1}$

and $E[e^{\epsilon Y_t}] < +\infty$ where $\epsilon > 0$ [6]. Furthermore, stability can be achieved by the DFT policy for all such arrival distributions without readjustment of the parameters $a(0)$, $a(1)$, $a(e)$, and γ (γ must be sufficiently small, depending only on λ and ϵ). The DFT policy is also relatively insensitive to dependence among the variables Y_t.

VII. Control Under Feedback Information Limitations

It has been assumed in the previous sections that each user can find out whether zero, one, or greater than one packet was transmitted in a given slot. It is convenient to call this feedback channel information. In many communication situations such feedback information may be incomplete or may contain errors which may be different for different users. For example, fading channels, noise, interference, or certain spread spectrum modulation techniques can make it difficult for receivers to distinguish between one packet transmission or more than one, or to distinguish between noise and packet transmissions. The control policy design and analysis ideas presented in this paper readily extend to cover many such situations.

One assumption stated earlier is that during slot $[t, t+1)$ all backlogged users transmit with the same probability f_t. However, due to the channel feedback problems considered above, the users may not be able to all maintain identical retransmission probabilities. Instead, user α retransmits with probability $f_t^{(\alpha)}$ in slot t. It might also be desirable to let $f_t^{(\alpha)}$ depend on the user α in order to give some users higher priority. This could easily be accomplished by assigning a larger value of the control constant η to users with higher priority.

Allowing $f_t^{(\alpha)}$ to depend on α has little effect on our retransmission procedure. This is because the retransmitted traffic intensity

$$\mu_t = \sum_{\alpha:\ \text{user } \alpha \text{ backlogged}} f_t^{(\alpha)}$$

has the same dynamics as before:

$$\mu_{t+1} \approx a(Z_t)^{\gamma} \mu_t$$

as long as all users continue to use the same retransmission probability update scheme (3.5). In particular, the channel traffic intensity dynamics still decouple from the backlog fluctuations when the backlog is moderately large.

To understand the effect of incomplete channel feedback information suppose that for some $m \geqslant 1$ the channel output is observed and the retransmission probabilities are updated using (3.2) only once every m slots. We shall see that performance is not severely degraded. The transition probabilities and hence also the average throughput intensity for the *local* model do not depend on m. The effect of $m > 1$ in the *global* model is that the rate of retransmission probability adjustment is slower by a factor of m when compared to the case $m = 1$, while the backlog fluctuates at about the same rate for all m. The effect is to increase the disturbance imposed on the local model due to backlog fluctuations by a factor of m. This effect is only significant when the backlog is small enough that backlog fluctuations limit the retransmission control performance.

The net effect of fixing $m = m_0 > 1$ versus $m = 1$ can best be summarized as follows. If the backlog is small, then the average throughput for $m = m_0$ will be about the same as for $m = 1$, but with γ replaced by γ/m. If the backlog becomes large enough, the average throughput should be about the same for $m = m_0$ and $m = 1$ (with the same γ). This behavior for $m_0 = 5$ is nicely illustrated in Figs. 4–5. The curves for $(m = 5, \gamma = 0.3)$ are close to the curves for $(m = 1,\ \gamma = 0.05)$ for $N_t \leqslant 35$ and they are close to the curves for $(m = 1, \gamma = 0.3)$ for $N_t \geqslant 40$.

The amount of feedback information available might be reduced in other ways. For example, it may not be possible for a user to distinguish between one and more than one transmission in a slot (unless the user transmitted a message himself). In other words, each user can only detect whether or not a slot was empty. Now by Proposition 4.1, two possible choices of $\boldsymbol{c} = (c(0), c(0), c(e))$ for which (4.2) is satisfied are

$$\boldsymbol{c}' = (0.462, -0.269, -0.269)$$

and

$$\boldsymbol{c}'' = (0.209, 0.209, -0.582).$$

Since the last two entries of $\boldsymbol{c}'$ are equal, the resulting retransmission rule can be implemented even if the stations can only distinguish between an empty and a nonempty slot. Similarly, if the retransmission rule corresponding to $\boldsymbol{c}''$ is used, then the stations need only distinguish collisions from noncollisions. (This has been called "0–1 feedback" [10].) In each case, stable throughput for any $\lambda < e^{-1}$ can be achieved.

The preceding analysis deals with restricted feedback information. Now consider instead situations in which the feedback information contains errors. In our framework, the effect of channel feedback errors should be analyzed in the local model since the occurrence of such errors may be closely linked to the intensity of channel traffic, but should be otherwise independent of the channel backlog. The effect of such errors could be quantified by seeing how much they cause $\bar{S}_t$ to decrease ($\bar{S}_t$ is the average throughput intensity under the local model). The system performance should degrade gracefully as the frequency of errors increases because of the apparent stability of the local model. These ideas and the use of estimation techniques in cases of severe feedback channel limitations pose important problems for future research.

Appendix A
Poisson Approximation

Since $e^{-f} \geqslant 1 - f$ for $f \geqslant 0$, $e^{-nf} - (1-f)^n \geqslant 0$ for $f \geqslant 0$. On the other hand, using the power series

$$\ln(1-f) = -f - \frac{1}{2}f^2 - \frac{1}{3}f^3 - \cdots$$

which is absolutely convergent for $|f| < 1$ yields that

$$
\begin{aligned}
e^{-nf}-(1-f)^n &= e^{-nf}[1-\exp(n(f+\ln(1-f)))] \\
&= e^{-nf}\left[1-\exp\left(-nf^2\left(\frac{1}{2}+\frac{1}{3}f+\frac{1}{4}f^2+\cdots\right)\right)\right] \\
&\leq e^{-nf}\left[1-\exp\left(-nf^2\frac{1}{2}(1+f+f^2+\cdots)\right)\right] \\
&= e^{-nf}\left[1-\exp\left(-nf^2\frac{1}{2(1-f)}\right)\right]
\end{aligned}
$$

for $0<f<1$. This proves (2.1).

The right side of (2.1) is a product of two factors, each of which is less than 1 and is monotonic in f. Hence, for any fixed f_0,

$$
e^{-nf}-(1-f)^n \leq \begin{cases} e^{-nf_0} & \text{if } f \geq f_0 \\ 1-\exp\left(-nf_0^2\left(\dfrac{1}{2(1-f_0)}\right)\right) & \text{if } f \leq f_0. \end{cases}
$$

Inequality (2.2) follows from this for the choice $f_0=n^{d-1}$. The first inequality in (2.3) is an easy consequence of (2.1) and the second follows by maximizing $e^{-xf}(1-e^{-xf^2})$ over $0 \leq x \leq \infty$ for fixed f.

To prove (2.4) we note that

$$
\begin{aligned}
|nfe^{-nf}-nf(1-f)^{n-1}| &= nf|e^{-nf}-(1-f)^{n-1}| \\
&\leq nf|e^{-nf}-(1-f)^n|+nf^2(1-f)^{n-1}
\end{aligned}
$$

and then apply inequality (2.1) to each of the terms on the right. Finally, inequalities (2.5) and (2.6) can be proved using (2.4) by the same technique used to prove (2.2).

Appendix B
Proof of Proposition 4.1

Equation (4.2) is equivalent to

$$
\operatorname{sgn}(G-1) = -\operatorname{sgn} f(G) \qquad \text{for } 0<G<+\infty \qquad \text{(B.1)}
$$

where

$$
f(G)=\left(e^{-G}, Ge^{-G}, 1-(1+G)e^{-G}\right)\cdot \boldsymbol{c}.
$$

By defining $f(+\infty)=c(e)$, f becomes continuous on the extended interval $[0,+\infty]$. If (B.1) is true, then

$$
\begin{aligned}
&0 \leq f(0)=c(0) \\
&0 \geq f(+\infty)=c(e) \\
&0 = f(1) = (e^{-1}, e^{-1}, 1-2e^{-1})\cdot \boldsymbol{c}
\end{aligned}
$$

so the conditions of Proposition 4.1 are necessary.

To prove the sufficiency, assume the conditions of Proposition 4.1 are true. Note that

$$
f'(G)=e^{-G}(-1,1-G,G)\cdot \boldsymbol{c}.
$$

Now $f'(1)=-c(0)+c(e)\leq 0$ since $c(0)\geq 0 \geq c(e)$. In fact, $f'(1)<0$ since $\boldsymbol{c}\neq 0$. Furthermore, since $(-1,1-G,G)\cdot\boldsymbol{c}$ is an affine function of G, f' has, at most, one zero in $(0,+\infty)$. Also, $f(1)=0$ and $f(0)\geq 0 \geq f(+\infty)$. These facts imply (B.1), and Proposition 4.1 is proved.

Appendix C

In this appendix we shall find an approximate expression for the average backlog $\bar{N}_t$ when the throughput intensity is given by

$$
S_t = \begin{cases} \lambda e^{-\lambda} & \text{if } N_t=0 \\ \bar{S} & \text{if } N_t>0 \end{cases}
$$

for a constant $\bar{S}>0$. Some quite general equations can be used.

Random Walk with Boundary and Steps Bounded Below by -1: Suppose that $(X_k\colon k\geq 0)$ is a Markov chain on $Z_+=\{0,1,\cdots\}$ such that

$$
P[X_{k+1}=i \mid X_k=j] = \begin{cases} u_{i-j} & \text{if } j>0 \\ v_i & \text{if } j=0 \end{cases} \qquad \text{(C.1)}
$$

where

$$
\sum_{i=-1}^{\infty} u_i = \sum_{i=0}^{\infty} v_i = 1.
$$

Let

$$
p_n^{(k)} = P[X_k=n].
$$

Then

$$
p_n^{(k+1)} = \sum_{i=-1}^{n} p_{n-i}^{(k)}u_i + p_0^{(k)}(v_n-u_n) \qquad n\geq 0. \qquad \text{(C.2)}
$$

Define the generating functions

$$
\begin{aligned}
P^{(k)}(z) &= \sum_{n=0}^{\infty} p_n^{(k)} z^n \\
U(z) &= \sum_{n=-1}^{\infty} u_n z^n \\
V(z) &= \sum_{n=0}^{\infty} v_n z^n.
\end{aligned}
$$

Then (C.1) becomes

$$
P^{(k+1)}(z) = P^{(k)}(z)U(z) + p_0^{(k)}(V(z)-U(z)).
$$

If a stationary distribution exists, then since the chain is irreducible and aperiodic, $p_n^{(k)} \to p_n$ for each n, where $(p_n\colon n\geq 0)$ is the stationary distribution. So $P^{(k)}(z)\to P(z)$ if $|z|<1$ where $P(z)=\sum_{n=0}^{\infty}p_n z^n$. So if a stationary distribution exists, it satisfies

$$
P(z)=P(z)U(z)+p_0(V(z)-U(z)).
$$

If the distribution (u_n) is not degenerate so that $U(z)\not\equiv 1$, we can solve for $P(z)$:

$$
P(z) = \frac{V(z)-U(z)}{1-U(z)}p_0.
$$

Since $\lim_{z\to 1}P(z)=1$, we find by l'Hopitals' rule that $p_0=-\bar{u}/(\bar{v}-\bar{u})$, assuming that the means $\bar{u}=\sum k u_k$ and $\bar{v}=\sum k v_k$ exist. Let $\bar{x}=\sum k p_k$ be the expected value of X_t with respect to the invariant measure. Then $\bar{x}=$

$\lim_{z\to 1} P'(z)$. Now, using l'Hopitals' rule and the fact $\lim_{z\to 1} U''(z) = \overline{u^2} - \bar{u}$, etc., yields that

$$\bar{x} = \frac{\bar{v}\,\overline{u^2} - \overline{v^2}\,\bar{u}}{-2\bar{u}(\bar{v} - \bar{u})} \tag{C.3}$$

assuming that the first and second moments $\bar{u}, \bar{v}, \overline{u^2}, \overline{v^2}$ of the distributions (u_k) and (v_k) are finite.

Average Backlog Assuming Constant Traffic Rate During Busy Periods: Consider a random access situation in which the retransmitted traffic intensity is a constant $\mu > 0$ whenever $N_t > 0$. Then the backlog process N_t is a Markov chain which, under the local Poisson approximation, has transition probabilities which are of the form (C.1) with

$$v_k = \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{k!} & k \geq 2 \\ (1+\lambda)e^{-\lambda} & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$u_k = \begin{cases} \dfrac{\lambda^k e^{-\lambda}}{k!} & k \geq 2 \\ \lambda e^{-\lambda}(1 - e^{-\mu}) & k = 1 \\ \lambda e^{-(\lambda+\mu)} + e^{-\lambda}(1 - \mu e^{-\mu}) & k = 0 \\ \mu e^{-(\lambda+\mu)} & k = -1. \end{cases}$$

Elementary calculations yield that

$$\bar{v} = \lambda - e^{-\lambda} \tag{C.4}$$

$$\overline{v^2} = \lambda + \lambda^2 - \lambda e^{-\lambda} \tag{C.5}$$

$$\bar{u} = \lambda - \bar{S} \tag{C.6}$$

$$\overline{u^2} = \lambda + \lambda^2 + \frac{\mu - \lambda}{\mu + \lambda}\bar{S} \tag{C.7}$$

where $\bar{S} = (\lambda + \mu)\exp(-(\lambda + \mu))$ is the throughput intensity when $N_t > 0$. Using these values in (C.3) yields the mean of N_t under the invariant distribution.

Approximate Average Backlog Assuming Constant Throughput Intensity During Busy Periods: The throughput at time t in the general model is $S_t = (\lambda + \mu_t)\exp(\lambda + \mu_t)$. If S_t is held at a fixed level $\bar{S}$, the traffic level μ_t is not completely determined. However, if $\bar{S}$ is near e^{-1}, then μ_t must be near 1 so it is reasonable to replace μ by 1 in the right-hand side of (C.7). If (C.4)–(C.7) are then substituted into (C.3) an approximation of $\overline{N_t}$ is obtained in terms of λ and $\bar{S}$ which is especially good for $\bar{S}$ near e^{-1}. The exact value of $\overline{N_t}$ under the local Poisson assumption is obtained when $\bar{S} = e^{-1}$, for then $\mu = 1$.

REFERENCES

[1] B.-T. An and E. Gelenbe, "Etude et simulation de reseaux de diffusion sous controle," *RAIRO Informatique/Comput. Sci.*, vol. 11, pp. 301–321, 1977.

[2] J. I. Capetanakis, "Tree algorithms for packet broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-25, Sept. 1979.

[3] G. Fayolle, E. Gelenbe, and J. Labetoulle, "Stability and optimal control of the packet switching broadcast channel," *J. Ass. Comput. Mach.*, vol. 24, pp. 375–386, July 1977.

[4] M. J. Ferguson, "On the control, stability, and waiting time in a slotted ALOHA random-access system," *IEEE Trans. Commun.*, Nov. 1975.

[5] D. Freedman, *Approximating Countable State Markov Chains.* San Francisco, CA: Holden-Day, 1971.

[6] B. Hajek, "Hitting time and occupation time bounds implied by drift analysis with applications," *Advances Appl. Probability*, Sept. 1982, to be published.

[7] L. Kleinrock, *Queueing Systems, vol. 2: Computer Applications.* New York: Wiley, 1976.

[8] S. S. Lam and L. Kleinrock, "Packet switching in a multiaccess broadcast channel: Dynamic control procedures," *IEEE Trans. Commun.*, vol. COM-23, pp. 891–904, Sept. 1975.

[9] J. L. Massey, "Collision-resolution algorithms and random-access communications," in *Multi-User Communications*, G. Longo, Ed. New York: Springer-Verlag, 1981.

[10] V. A. Mikhailov, "A random access algorithm for a broadcast channel" (Abstract, in Russian), in *5th Int. Symp. Inform. Theory*, vol. 3, Tbilisi, Moscow, 1979, pp. 83–86.

[11] I. Rubin and M. Louie, "A class of hybrid TDMA/tree-random-access schemes for multi-access communication channels," in *Abst. Papers, 1981 IEEE Inform. Theory Symp.*, Santa Monica, CA, Feb. 1981.

[12] F. Schoute, "Decentralized control in packet switched satellite communication," *IEEE Trans. Automat. Contr.*, vol. AC-23, Apr. 1978.

[13] A. Segal, "Recursive estimation from discrete-time point processes," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 422–431, 1976.

[14] J. W. Sennott and L. I. Sennott, "A queueing model for analysis of a bursty multiple-access communication channel," *IEEE Trans. Inform. Theory*, vol. IT-27, May 1981.

[15] D. Seret and C. Macchi, "Controle adaptatif d'un reseau à diffusion de paquets," in *Flow Control in Computer Networks*, J. L. Grange and M. Gien, Eds. New York: IFIP, North-Holland, 1979.

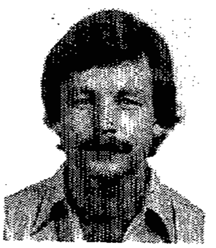
[16] F. A. Tobagi, "Analysis of a two-hop centralized packet radio network, Part I: Slotted ALOHA," *IEEE Trans. Commun.*, vol. COM-28, Feb. 1980.

[17] B. S. Tsybakov and U. A. Mikhailov, "Slotted multiaccess packet-broadcasting feedback channel," *Problemy Peredachi Informatsii*, vol. 14, pp. 32–59, Oct.–Dec. 1978.

[18] P. Varaiya and J. Walrand, "Decentralized control in packet switched satellite communication," *IEEE Trans. Automat. Contr.*, vol. AC-24, Oct. 1979.

[19] J. E. Wieselthier and A. Ephremides, "A new class of protocols for multiple access in satellite networks," *IEEE Trans. Automat. Control*, vol. AC-25, Oct. 1980.

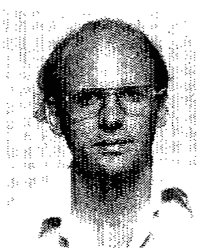
[20] R. L. Cruz, "Protocols for multi-access channels with continuous entry and noisy feedback," M.S. thesis, Lab. Inform. Decision Syst., M.I.T., Cambridge, MA, to appear.



Bruce Hajek (M'79) was born in Elmhurst, IL, on August 20, 1955. He received the B.S. degree in mathematics with highest honors in 1976, the M.S. degree in electrical engineering from the University of Illinois, Urbana, in 1977, and the Ph.D. degree in electrical engineering from the University of California, Berkeley, in 1979.

During the Summer of 1975 he was a Summer Research Student in the Division of Applied Mathematics, Brookhaven Laboratory, New York. Since August 1979 he has been an Assistant Professor with the Department of Electrical Engineering and a Research Assistant Professor with the Coordinated Science Laboratory, University of Illinois, Urbana. His research interests include multiple-user communication theory (e.g., random access and network routing problems), information theory, random fields and other stochastic processes, and estimation and stochastic control.

Dr. Hajek was a three-year National Science Foundation Graduate Fellow from 1976 to 1979. He was a winner of the 1973 U.S.A. Mathematical Olympiad. He won the Bennet Prize for the Outstanding Senior in Mathematics at the University of Illinois. He is a member of Phi Kappa Phi and the American Mathematics Society.



Timothy van Loon was born in Denver, CO, on April 12, 1953. He received the B.S. degree in mathematics/physics from the University of Southern Colorado in 1975, the M.S. degree in mathematics from Iowa State University, Ames, in 1977, and the M.S. degree in electrical engineering from the University of Illinois, Urbana, in 1981.

Mr. van Loon participated in the summer student research program at Argonne National Laboratory, IL, in 1979. Since August 1981 he has been with TRW in Redondo Beach, CA.