CONTROL PROCEDURES FOR SLOTTED ALOHA SYSTEMS THAT ACHIEVE STABILITY

Loren P. Clare

Rockwell International Science Center
1049 Camino Dos Rios
Thousand Oaks, California 91360

Abstract—**A class of slotted ALOHA dynamic control strategies is considered. These strategies are simple to implement and can yield lossless and stable operation for arbitrarily large user populations with aggregate arrival rates below e^{-1} packets/slot. An ergodicity analysis is given that provides conditions on the system parameters, such that any specified set of control parameters that satisfies the given conditions is guaranteed to yield stable performance. The system state is modelled as a two-dimensional Markov chain that incorporates the backlog (the number of packets awaiting retransmission) and the estimate of the backlog. The geometrical concepts are illustrated by figures corresponding to an example case. Simulation results are presented that compare alternative control schemes.**

I. INTRODUCTION

ALOHA is a fundamental technique for multiaccess communication and forms the basis of a number of major protocols in modern computer communications, such as CSMA, CSMA/CD, and many reservation schemes. Random access protocols, of which ALOHA is arguably the simplest, are appropriate for sharing a channel among a large population of users with bursty traffic. However, use of ALOHA can result in unstable behavior, causing low throughput and excessive delays, unless an adequate control procedure is employed.

Although a number of dynamic control schemes have been offered in the past (*e.g.*, [1]–[4]), only the scheme of Hajek and van Loon [5] appears to have been proven to be stable [6], where we define a system to be stable if it is ergodic for an arrival process that is independent of the system state. In this paper we consider a class of slotted ALOHA control procedures that include a number of schemes that can potentially offer superior performance compared to the procedures of Hajek and van Loon [5]. We develop a stochastic model and formulate constraints on the control system parameters that are proven to guarantee stability. The method of proof for this application is new. A complete queueing analysis may be made based on the system model [7], although we concentrate on the stability results only here.

In slotted ALOHA systems, users transmit information in the form of fixed-length packets. The users are synchronized to periodic instants of time, and packet transmissions must begin at one of these instants. The period, called a "slot," is equal to the packet transmission time. If multiple packets are transmitted simultaneously, a "collision" occurs and none is successfully received. If a newly transmitted packet fails, it joins the set of users with packets that have failed at least once and are awaiting successful retransmission. We refer to (the size of) this set as the backlog. When a user learns that his packet transmission failed, a decision must be made regarding when he should retransmit. We assume that this decision process can make use of feedback information that indicates whether none, one, or more than one packets were transmitted in the preceding slot.

We consider a class of slotted ALOHA schemes in which the retransmission control mechanism is a function of a single variable, representing an estimate of the backlog, that depends only on the previous estimate and the current feedback. This not only makes the mechanism simple to implement, but also implies that the vector process of the backlog and its estimate forms a Markov chain. This allows a characterization of the backlog process that can be used to determine stability.

A considerable number of definitions for stability have been suggested for the slotted ALOHA system, some of which apply to the finite population model and to the system with a static control [8], [9]. Fayolle, *et al.* [10] offered the definition of stability that is based upon the ergodicity of the backlog process, and showed that the infinite population slotted ALOHA system with a static control is unstable. Other authors have presented alternative proofs of nonergodicity (Kaplan [11], Rosenkrantz and Towsley [12]).

Fayolle, *et al.* [13] derived necessary conditions for a backlog-based slotted ALOHA system to be stable in the ergodic sense. They also considered the system in which the backlog was perfectly known to all users, and derived the "optimal" retransmission probability based on this knowledge. This nonrealizable system is useful because it is stable, easy to analyze and perform computations for, and can be used to establish bounds on performance.

The retransmission control procedure of Hajek and van Loon [5] is simple to implement and, as proved by Hajek [6], can be made stable for any arrival rate λ below e^{-1} with the appropriate choice of control system parameters. The control mechanism is such that the retransmission probability is updated directly based on the feedback information. This retransmission probability can be associated with a comparable "backlog estimate" (see equation (1) in the next section). In so doing, the "backlog estimate" update mechanism is primarily multiplicative, which is in contrast to the additive structure that we will consider.

Thomopoulos [15] has investigated the dynamic control procedure based on the minimum mean-squared error (MMSE) estimator of the backlog. He identified a simplified control scheme [16] that is the asymptotic limit of the MMSE estimator as the backlog tends to infinity; this simplified scheme is included in the class of control schemes we consider here. Thomopoulos has shown that the asymptotic MMSE scheme is stable [17] under a somewhat different definition than that used in this paper.

The approach in this paper is based on consideration of the drift of the underlying Markov chain. Based on geometrical reasoning, we construct a nonnegative-valued test function on the state space and demonstrate the test function drift is bounded strictly below zero. We conclude that the process is ergodic, *i.e.*, the system is stable. The remainder of this paper is organized as follows. In the next section, we define the control structure and system parameters. In Section III we derive the transition matrix for the Markov chain of the backlog and its estimate, and define the associated drift vector field. The stability analysis is presented in Section IV. Some numerical results are given in Section V. Finally, conclusions and suggestions for future work are given.

II. System Model and Control Structure

As indicated in the Introduction, slotted ALOHA is an inherently discrete-time system. We will refer to the t^{th} slot as occurring from time t to $t+1$, $t = 1, 2, 3, \ldots$, *i.e.*, time is measured in units of slots. We assume that each user is capable of buffering at most one packet at a time.

New packets arrive only at slot boundaries. The number of new packets (aggregated over all users) arriving at time t is denoted A_t, $t = 1, 2, \ldots$. We assume $\{A_t\}$ forms an independent and identically distributed sequence, and in particular, it is independent of the system state. Let λ [packets/slot] represent the mean of A_t and denote $a_i = P(A_t = i)$, $i = 0, 1, 2, \ldots$. Although the results to be presented are applicable to any such "white noise" arrival sequence, we will typically assume that A_t is Poisson; thus,

$$a_i = e^{-\lambda}\lambda^i/i! \quad , \quad i = 0, 1, 2, \ldots.$$

This is the "infinite user population" assumption, which arises from the fact that the superposition of a large number M of sufficiently independent point processes (each corresponding to an individual user) tends to be a Poisson process as $M \to \infty$; the "slotting" then yields the discrete-time process above.

A successful transmission occurs if and only if exactly one user transmits in that slot; we will also refer to this as a departure occurrence. We define the feedback F_t for the t^{th} slot as the following ternary-valued function of the number of users transmitting during the t^{th} slot: for $t = 1, 2, \ldots$,

$$F_t = \begin{cases} 0 & \text{if no user transmits in the } t^{\text{th}} \text{ slot,} \\ 1 & \text{if one user transmits in the } t^{\text{th}} \text{ slot, and} \\ \text{c} & \text{if two or more users transmit in the } t^{\text{th}} \text{ slot.} \end{cases}$$

We assume that F_t is known and available for use in the retransmission decision process for the $(t+1)^{\text{th}}$ (next) slot.

We assume an Immediate First Transmission protocol is employed, *i.e.*, as soon as a new packet arrives, it is transmitted. If a new packet fails, it joins the backlog. The backlog at time t, denoted B_t, is the number of packets that have been transmitted at least once but have not yet been successful. The total number of packets requiring transmission during slot t is $A_t + B_t$. Thus, B_t does not count a departure that just occurred, and it does not count any arrivals just about to occur at time t. A timing diagram is provided in Figure 1, where we introduce fictitious time intervals between system measurements for mathematical clarity.

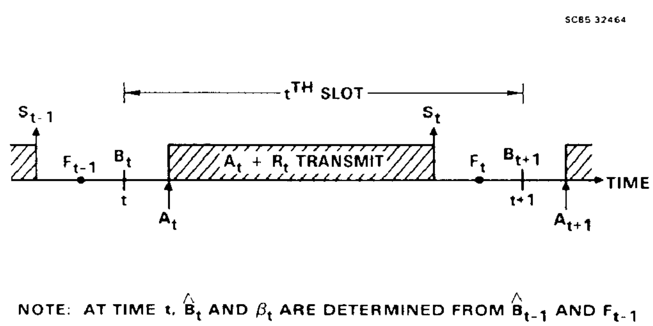

Figure 1. Timing diagram of system variables.

Each of the B_t backlogged packets will retransmit in the t^{th} slot with the <u>same</u> but independent probability, denoted by β_t, $t = 1, 2, \ldots$. We let R_t denote the number of backlogged packets that retransmit during the t^{th} slot. Then R_t is binomially distributed with parameters B_t and β_t, being the sum of B_t independent Bernoulli random variables. Note that F_t is a deterministic function of $A_t + R_t$, the number of transmissions in the t^{th} slot. The departure process is identified by defining the success indicator at time t as $S_t = I(A_t + R_t = 1)$, where I is the indicator function

$$I(\mathcal{E}) = \begin{cases} 1 & \text{if event } \mathcal{E} \text{ occurs, and} \\ 0 & \text{otherwise.} \end{cases}$$

We define the throughput at time t to be the probability of success $P(S_t = 1)$.

The system is controlled by the use of a backlog estimate. We let $\hat{B}_t$ denote the estimate (at time t) of the backlog at time t, $t = 1, 2, \ldots$. The retransmission probability is assumed to be a deterministic function of $\hat{B}_t$: $\beta_t = \beta(\hat{B}_t)$, and should be chosen to maximize the throughput. Conditioned on the backlog estimate $\hat{B}$, the throughput is

$$P(S=1 \mid \hat{B}) = E\big[\lambda e^{-\lambda}(1-\beta)^B + e^{-\lambda}B\beta(1-\beta)^{B-1} \mid \hat{B}\big].$$

In general, determination of the value of β that maximizes the throughput would require knowledge of the distribution of B conditioned on $\hat{B}$. If our estimate of B were perfect, then $\hat{B} = B$ almost surely and we may easily deduce that the optimal β is given by [13]

$$\beta(\hat{B}) = \begin{cases} (1-\lambda)/(\hat{B}-\lambda) & \text{if } \hat{B} \geq 1, \text{ and} \\ 1 & \text{if } 0 \leq \hat{B} < 1. \end{cases} \tag{1}$$

Although a perfect estimator is not possible, we nevertheless expect that (1) is a reasonable form for the function β. We therefore define the function β by (1) for the remainder of this paper. Note that (1) has been analytically extended to allow noninteger $\hat{B} \geq 0$.

The form of the backlog estimate update mechanism will be chosen such that each estimate depends only upon the previous estimate and the current feedback information. We express the update mechanism as follows:

$$\hat{B}_{t+1} = \hat{B}_t + U(F_t, \hat{B}_t). \tag{2}$$

We choose the update function U to have the following simple form:

$$U(f, \hat{b}) = \max(\hat{b}_{\min} - \hat{b}, u_f), \qquad \hat{b} \geq \hat{b}_{\min}, \quad f = 0, 1, c, \tag{3}$$

where $\hat{b}_{\min}$ is a lower boundary constraint on $\hat{B}$, and where u_0, u_1 and u_c are real constants that completely specify the control mechanism via (1)–(3). For later convenience in bookkeeping we restrict $\hat{b}_{\min} \geq 1$.

Unless there is a rational relationship between the control parameters u_0, u_1 and u_c, the state space for the component $\hat{B}$ is uncountable. An ergodicity analysis for the uncountable case seems possible using the results of Tweedie [18] for general state-space Markov chains; however, we will avoid the technical details that would be required, and make the assumption that the control parameters are rationally related. This is a minor constraint since any real number can be approximated arbitrarily closely by a rational number. We mention that it is possible to force the state space to be countable (*e.g.*, $\hat{B}$ integer) without sacrificing complete generality on the choice of the control parameters by use of a randomized update rule; see [7] for more details.

III. The Backlog Process and Drift

The system state is defined as the backlog B_t and the estimate of the backlog $\hat{B}_t$. The process $\{(B, \hat{B})\}$ is an irreducible time-homogeneous Markov chain by virtue of the control structure (1)–(3). We will use the following notation: for $b, b' \geq 0$ and $\hat{b}, \hat{b}' \geq \hat{b}_{\min}$,

$$\pi_t(b, \hat{b}) = P(B_t = b, \hat{B}_t = \hat{b}), \quad \text{and}$$

$$p(b', \hat{b}' \mid b, \hat{b}) = P(B_{t+1} = b', \hat{B}_{t+1} = \hat{b}' \mid B_t = b, \hat{B}_t = \hat{b}).$$

Because $(B, \hat{B})$ is a Markov chain, we have for $b' \geq 0$ and $\hat{b}' \geq \hat{b}_{\min}$ that

$$\pi_{t+1}(b', \hat{b}') = \sum_{b \geq 0, \hat{b} \geq \hat{b}_{\min}} \pi_t(b, \hat{b}) p(b', \hat{b}' \mid b, \hat{b}). \tag{4}$$

Thus, the system state process is completely characterized by the initial system state density π_1 and the transition matrix p.

We denote $r_n(b, \hat{b}) = P(R_t = n \mid B_t = b, \hat{B}_t = \hat{b})$ and $\langle \hat{b} \rangle = \max(\hat{b}, \hat{b}_{\min})$. For $\hat{b} \geq \hat{b}_{\min}$ we find that the only nonzero terms of the transition matrix are given by

$$p(b-1, \langle \hat{b} + u_1 \rangle \mid b, \hat{b}) = a_0 r_1(b, \hat{b}), \quad b \geq 1, \tag{5}$$

$$\begin{aligned} p(b, \hat{b}' \mid b, \hat{b}) = {} & a_0 r_0(b, \hat{b}) I[U(0, \hat{b}) = \hat{b}' - \hat{b}] \\ & + a_1 r_0(b, \hat{b}) I[U(1, \hat{b}) = \hat{b}' - \hat{b}] \\ & + a_0\big(1 - r_0(b, \hat{b}) - r_1(b, \hat{b})\big) I[U(c, \hat{b}) = \hat{b}' - \hat{b}], \\ & b \geq 0, \ \hat{b}' \geq \hat{b}_{\min}, \end{aligned} \tag{6}$$

$$p(b+1, \langle \hat{b} + u_c \rangle \mid b, \hat{b}) = a_1\big(1 - r_0(b, \hat{b})\big), \quad b \geq 0, \tag{7}$$

and for integer $m \geq 2$

$$p(b+m, \langle \hat{b} + u_c \rangle \mid b, \hat{b}) = a_m, \quad b \geq 0, \tag{8}$$

where for $b \geq 0$, $\hat{b} \geq \hat{b}_{\min}$, because R is binomial,

$$r_0(b, \hat{b}) = \big[1 - \beta(\hat{b})\big]^b \quad \text{and} \tag{9}$$

$$r_1(b, \hat{b}) = b\beta(\hat{b})\big[1 - \beta(\hat{b})\big]^{b-1}. \tag{10}$$

(Equation (6) is left in a form that permits $u_f = u_{f'}$ for distinct $f, f' \in \{0, 1, c\}$.) If either b or $\hat{b}$ is large then we may use the approximation [5]

$$r_0(b, \hat{b}) = e^{-\mu} \qquad \text{and} \tag{11}$$

$$r_1(b, \hat{b}) = \mu e^{-\mu}, \qquad \text{where} \tag{12}$$

$$\mu(b, \hat{b}) = b\beta = \frac{b(1-\lambda)}{\hat{b} - \lambda}. \tag{13}$$

The **drift** at state $(b, \hat{b})$, where $b \geq 0$ and $\hat{b} \geq \hat{b}_{\min}$, is defined as the mean conditional state differential:

$$d(b, \hat{b}) = E\big[(B_{t+1} - B_t, \hat{B}_{t+1} - \hat{B}_t) \mid B_t = b, \hat{B}_t = \hat{b}\big].$$

The drift d is a vector field and may be derived using (5)–(8). For our purposes it will suffice to assume b or $\hat{b}$ is large so that (11)–(13) applies, in which case we find (for Poisson arrivals)

$$\begin{aligned} d &= \big(d_1, d_2\big) \\ &= \Big(\lambda - (\lambda + \mu)e^{-\lambda-\mu}, u_0 e^{-\lambda-\mu} + u_1(\lambda + \mu)e^{-\lambda-\mu} \\ &\qquad + u_c[1 - e^{-\lambda-\mu} - (\lambda + \mu)e^{-\lambda-\mu}]\Big). \end{aligned} \tag{14}$$

An example of a drift vector field will be seen in Figures 2 and 3 of Section V.

If the control mechanism is chosen so as to be stable, then the stationary distribution of $(B, \widehat{B})$ exists, which we denote by

$$\pi_\infty(b, \hat{b}) = \lim_{t\to\infty} \pi_t(b, \hat{b}), \qquad b \geq 0,\ \hat{b} \geq \hat{b}_{\min}.$$

We may obtain $\{\pi_\infty(b, \hat{b})\}$ by iterating (4) until convergence occurs, or by solving the linear system of equations

$$\sum_{b,\hat{b}} \pi_\infty(b, \hat{b}) = 1 \quad \text{and} \tag{15}$$

$$\pi_\infty(b', \hat{b}') = \sum_{b,\hat{b}} \pi_\infty(b, \hat{b}) p(b', \hat{b}' \mid b, \hat{b}), \quad b' \geq 0,\ \hat{b}' \geq \hat{b}_{\min}. \tag{16}$$

IV. Stability Analysis

In this section we will generate criteria on the system parameters that will ensure stability. We begin by studying the properties of the drift components d_1 and d_2. We consider each component as a function of μ, $\mu \geq 0$, given by (13), since we will be interested only in the case where B or $\widehat{B}$ is large. The component

$$d_1(\mu) = \lambda - (\lambda + \mu)e^{-\lambda-\mu}$$

decreases monotonically to $\lambda - e^{-1}$ at $\mu = 1 - \lambda$, and then increases monotonically to λ as μ increases. We assume that $\lambda < e^{-1}$. Define μ_* and μ^* as the smaller and larger roots respectively of $d_1 = 0$. We can think of the state space as being divided into the two "bad" regions $\mu \leq \mu_*$ and $\mu \geq \mu^*$ and the "good" region $\mu_* < \mu < \mu^*$, since only in the "good" region will the backlog decrease.

Now consider the component

$$d_2(\mu) = u_c + (u_0 - u_c)e^{-\lambda-\mu} + (u_1 - u_c)(\lambda + \mu)e^{-\lambda-\mu}.$$

When μ is very small then the backlog estimate is too large (see (13)), so that we would like the estimate to be decreased, i.e., $d_2 < 0$. Similarly, we would like $d_2 > 0$ when μ is large. Thus we have the constraints

$$C1: \quad d_2(0) < 0, \quad \text{and}$$

$$C2: \quad d_2(\infty) = u_c > 0.$$

It is straightforward to check that d_2 has at most one critical point, and therefore $C1$ and $C2$ imply the existence of a unique $\mu > 0$ such that $d_2(\mu) = 0$; we denote this unique root of d_2 as μ'. To prevent the field lines of the vector field d from diverging to infinity, we must impose the following additional constraint:

$$C3: \quad \mu_* \leq \mu' \leq \mu^*.$$

The remaining constraints that will ensure stability will be developed with the following approach: We construct a nonnegative-valued test function (also called a Liapunov function) $L(B, \widehat{B})$, and to show the drift of the test function is strictly negative for all $(B, \widehat{B})$ outside a bounded region. More specifically, we will find a test function $L(B, \widehat{B})$ and derive constraints that imply that for all $b + \hat{b} > N$ we have

$$E\big[L(B_{t+1}, \widehat{B}_{t+1}) - L(b, \hat{b}) \mid B_t = b, \widehat{B}_t = \hat{b}\big] \leq -\epsilon \tag{17}$$

for some constant $\epsilon > 0$ and for some (large) number N. We then apply Theorem 6.1 of Tweedie [18] to conclude that the process is ergodic.

We look for a test function for which each level set encloses a bounded region, and where the level sets are nested inside one another, with smaller test function values corresponding to level sets that are closer to the origin. The test function will be chosen such that the process tends to drift from larger-valued level sets to smaller-valued level sets. Paraphrasing Kingman [20], the drift $d(b, \hat{b})$ must point into the region enclosed by the level surface through $(b, \hat{b})$.

The general technique of constructing a test function on a multidimensional state space has been used previously ([19], [20], [21]); however, none of these test functions is adequate for our needs here. For example, inspection of Figures 2 and 3 of the next section reveals that the Euclidean norm [19] will *not* satisfy our needs since the level sets for this function are partial circles with the origin at the center, and for large μ the tendency would be outward.

We instead choose a test function with level sets that are partial ellipses tangent to the lines $b = 0$ (a coordinate axis) and $\hat{b} = \lambda$. Such a test function is of the following form:

$$L(b, \hat{b}) = b + c_1(\hat{b} - \lambda) - \sqrt{4c_2 b(\hat{b} - \lambda)}, \tag{18}$$

where c_1 and c_2 are constants that satisfy

$$c_1 \geq c_2 \geq 0. \tag{19}$$

Computation of the Taylor series expansion of L given by (18) yields

$$\begin{aligned} L(b+\Delta b, \hat{b} + \Delta\hat{b}) &= L(b, \hat{b}) + \nabla L \cdot (b + \Delta b, \hat{b} + \Delta\hat{b}) \\ &+ O\left(\sqrt{\frac{\hat{b}}{b^3}}(\Delta b)^2 + \sqrt{\frac{1}{b\hat{b}}}\Delta b \Delta\hat{b} + \sqrt{\frac{b}{\hat{b}^3}}(\Delta\hat{b})^2\right), \end{aligned} \tag{20}$$

where the test function gradient ∇L is given by

$$\nabla L(b, \hat{b}) = \left(1 - \sqrt{c_2(\hat{b} - \lambda)/b},\ c_1 - \sqrt{c_2 b/(\hat{b} - \lambda)}\right). \tag{21}$$

Since the change in backlog in one slot is stochastically dominated by a Poisson variable A of fixed rate λ

$$|B_{t+1} - B_t| \prec A$$

and (unlike Hajek and van Loon's scheme [5]) the change in backlog estimate is bounded

$$|\widehat{B}_{t+1} - \widehat{B}_t| \leq \max_{f=0,1,c}(|u_f|)$$

it is straightforward to show that as long as $b + \hat{b} > N$ and N is chosen to be large, the higher order (than linear) terms of the expansion (20) are negligible. The condition for ergodicity (17) may therefore be replaced by

$$\nabla L(b,\hat{b})\cdot d(b,\hat{b}) \le -\epsilon, \quad \epsilon > 0, \quad b+\hat{b} > N, \tag{22}$$

where N is chosen sufficiently large that the linear approximation of L is valid as well as the Poisson approximation (11)–(13). This form of the condition for ergodicity was given by Kingman [20]. In words, at each point outside some bounded region, the drift d must be obtuse to the outward-pointing normal to the level set ∇L at that point. We note from (14) and (21) that both d and ∇L (and hence their dot product) may be considered as functions of the single variable μ given by (13).

There are a number of ways to choose the constants c_1 and c_2. We choose them so that the normal components $\nabla L = (n_1, n_2)$ satisfy $n_1 = 0$ at μ_* and $n_2 = 0$ at μ'. This will then guarantee that $\nabla L \cdot d < 0$ for all $\mu \le \mu^*$. The choice of c_1 and c_2 is then given by

$$c_1 = \frac{\sqrt{\mu_* \mu'}}{1-\lambda} \quad \text{and} \tag{23}$$

$$c_2 = \frac{\mu_*}{1-\lambda}. \tag{24}$$

(Note that (19) is implied by $C3$, (23) and (24).) The gradient of L is then given by

$$\nabla L = \left(1 - \sqrt{\frac{\mu_*}{\mu}},\ \frac{(\sqrt{\mu'} - \sqrt{\mu})\sqrt{\mu_*}}{1-\lambda}\right). \tag{25}$$

An example of such a test function will be seen in Figures 4 and 5 of the next section.

Because of (25), we are guaranteed that the system is stable if we can show

$$C4: \quad \nabla L \cdot d \le -\epsilon \quad \text{for some constant } \epsilon > 0, \quad \forall \mu > \mu^*.$$

In practice, the function $\nabla L \cdot d$ is quite smooth and it is therefore a simple matter to evaluate it over the entire domain and check that it remains negative. An example of such a computation will be provided by Figure 6 of the next section. Since one may not consider it mathematically rigorous to have a constraint that involves pointwise verification over a continuous domain of values, we now develop a finite set of sufficient conditions for $C4$. The sufficient conditions derived below are met for a nonempty range of system parameter values, but are not necessary for $C4$ to be satisfied for all choices of control parameters. We leave the rigorous treatment of other cases that do not meet the constraints below for later development.

We will find conditions that imply $\nabla L \cdot d$ is monotonically decreasing for $\mu > \mu^*$. We impose the following constraints, which will imply in particular that d_2 is monotonically increasing on $\mu \ge 0$:

$$C5: \quad u_c > u_1, \quad \text{and}$$

$$C6: \quad \frac{u_0 - u_1}{u_c - u_1} < \lambda.$$

We consider the derivative of $\nabla L \cdot d$:

$$\frac{d}{d\mu}(\nabla L \cdot d) = d_1 \frac{d}{d\mu} n_1 + n_1 \frac{d}{d\mu} d_1 + d_2 \frac{d}{d\mu} n_2 + n_2 \frac{d}{d\mu} d_2.$$

It is easy to show that, because of the monotonicity properties of d_1 and d_2,

$$\begin{aligned} d_1 \frac{d}{d\mu} n_1 + d_2 \frac{d}{d\mu} n_2 &= \frac{1}{2}\sqrt{\frac{\mu_*}{\mu}}\left[\frac{d_1(\mu)}{\mu} - \frac{d_2(\mu)}{1-\lambda}\right] \\ &\le \frac{1}{2}\sqrt{\frac{\mu_*}{\mu}}\left[\frac{\lambda}{\mu^*} - \frac{d_2(\mu^*)}{1-\lambda}\right], \quad \mu > \mu^*, \end{aligned}$$

which is nonpositive provided

$$C7: \quad d_2(\mu^*) \ge \frac{\lambda(1-\lambda)}{\mu^*}.$$

Now consider the remaining terms $n_1 \frac{d}{d\mu} d_1 + n_2 \frac{d}{d\mu} d_2$. Let

$$\begin{aligned} g(\sqrt{\mu}) =& (n_1 \frac{d}{d\mu} d_1 + n_2 \frac{d}{d\mu} d_2)/(\mu^{-\frac{1}{2}} e^{-\lambda-\mu}) \\ =& (\sqrt{\mu} - \sqrt{\mu_*})(\lambda + \mu - 1) \\ &+ \frac{\sqrt{\mu_*}}{1-\lambda}\sqrt{\mu}(\sqrt{\mu'} - \sqrt{\mu}) \\ &\times [u_1 - u_0 + (u_c - u_1)(\lambda + \mu)], \end{aligned} \tag{26}$$

i.e., treat the function $g(x)$ as a fourth-degree polynomial of $x = \sqrt{\mu}$. We may force this polynomial to be convex $\cap$ by forcing $\frac{d^2 g}{dx^2}$ to not have distinct real roots; this is found by the quadratic formula to be equivalent to

$$\begin{aligned} C8: \quad & 3\left[\frac{1-\lambda}{\sqrt{\mu_*}} + (u_c - u_1)\sqrt{\mu'}\right]^2 \\ & \le 8(u_c - u_1)[1 - \lambda + u_1 - u_0 + \lambda(u_c - u_1)]. \end{aligned}$$

The function $g(x)$ is then monotone decreasing provided $\frac{dg}{dx}(\mu^*) \le 0$, which is found to be equivalent to

$$\begin{aligned} C9: \quad 0 \ge\ & 2\sqrt{\mu^*}(\sqrt{\mu^*} - \sqrt{\mu_*}) + (\mu^* + \lambda - 1) \\ & + \frac{\sqrt{\mu_*}}{1-\lambda}\Big((\sqrt{\mu'} - 2\sqrt{\mu^*})[u_1 - u_0 + (u_c - u_1)(\lambda + \mu^*)] \\ & + 2\mu^*(\sqrt{\mu'} - \sqrt{\mu^*})(u_c - u_1)\Big). \end{aligned}$$

Finally, we may conclude that $g(x) \le 0$ for all $x \ge \sqrt{\mu^*}$ if

$$C10: \quad g(\sqrt{\mu^*}) \le 0,$$

where g is given by (26). This then completes the derivation of the constraints that prove stability.

We point out that the technique given in this section is constructive and can be applied to determine whether a *specified* set of control parameters yields stable performance for a given λ. This is in contrast to Hajek's proof [6]; his proof shows the existence of a stable control scheme with that control structure, but it is difficult to see how one could prove stability for a specific choice of control parameters. In particular, verification that the example scheme used in [5] is stable is not obvious.

V. Numerical Results

We illustrate the theory of the previous sections by selecting a specific dynamic control scheme. All figures in this section are based upon the control strategy of the form (1)–(3) in which the control parameters are chosen as $u_0 = 2 - e \cong -.7183$, $u_1 = 0$, and $u_c = 1$ (or more precisely, u_0 is chosen as a rational approximation to $2 - e$). The offered load is assumed to be $\lambda = .32$; since the system is stable this also represents the throughput. The particular choice of control parameters above is somewhat arbitrary, with u_1 and u_c chosen to be simple, and then u_0 chosen to yield $d_2(1 - \lambda) = 0$. This latter constraint is similar to that of Hajek and van Loon ([5], equation (4.2)).

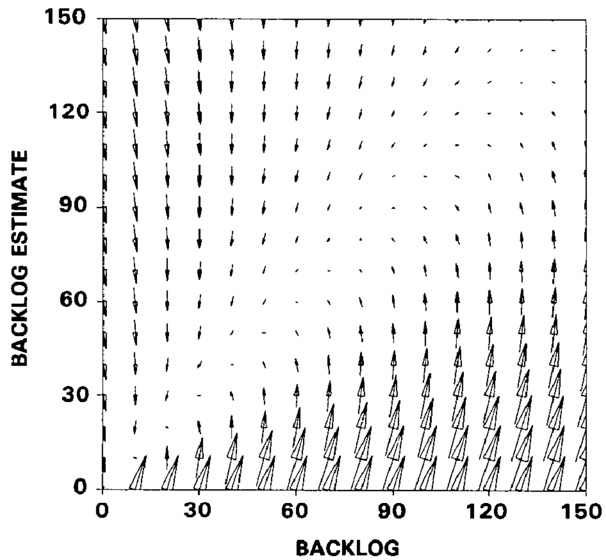

Figure 2. Drift vector field d. $\lambda = .32$, $u_0 = 2 - e$, $u_1 = 0$, and $u_c = 1$.

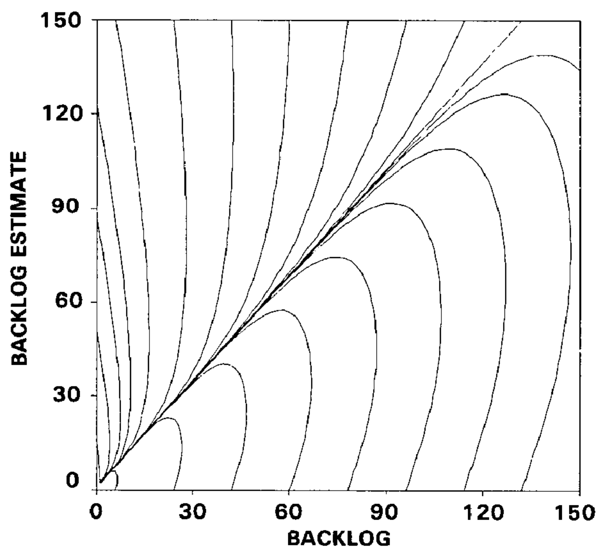

Figure 3. Drift field lines. $\lambda = .32$, $u_0 = 2 - e$, $u_1 = 0$, and $u_c = 1$.

An example of the drift vector field d is illustrated in Figure 2. The arrows in the figure indicate the direction of the drift at that point, and their size is proportional to the magnitude of the drift there. The field lines for this same case are illustrated in Figure 3 and are indicative of the average trajectories one might expect the process $(B, \widehat{B})$ to follow.

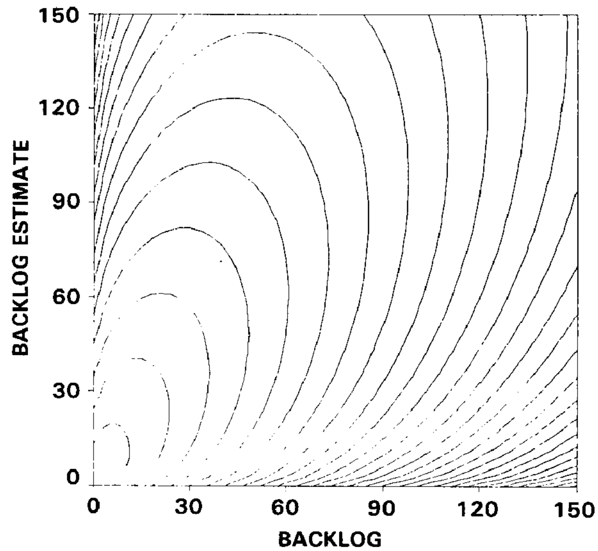

Figure 4. Test function level sets L =constant.

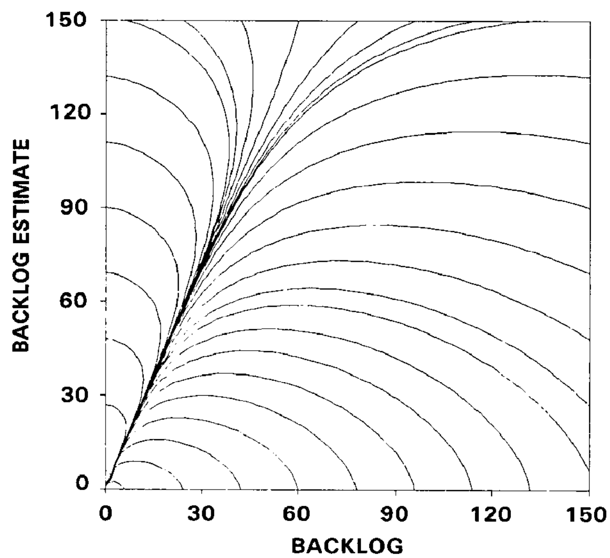

Figure 5. Test function gradient ∇L field lines.

Figure 4 illustrates the level sets for the test function (18) that arises from the use of (23) and (24) and the given parameter values. For this case one finds that $\mu_* \cong .2405$, $\mu^* \cong 1.304$, and $\mu' = 1 - \lambda = .68$. The gradient vector field ∇L consists of vectors that are normal to the level sets. Figure 5 illustrates the field lines for ∇L. The field lines of Figures 3 and 5 intersect at obtuse angles.

Figure 6 illustrates the test function drift $\nabla L \cdot d$ computed versus the parameter $\theta = \tan^{-1}((1-\lambda)/\mu)$, where in the figure θ is indicated in units of degrees. This simple graph demonstrates that the process is stable. It is straightforward to verify that this case satisfies all of the conditions $C1$–$C10$.

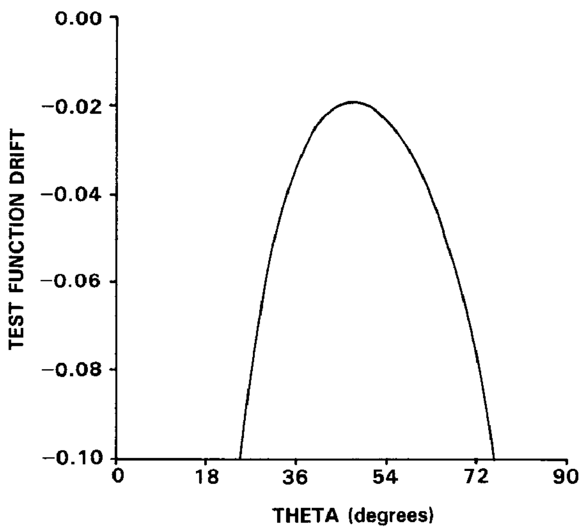

Figure 6. Test function drift $\nabla L \cdot d$ versus θ, $\theta = \tan^{-1}((1-\lambda)/\mu)$.

The problem of analytically determining the optimal choice of control parameters u_0, u_1 and u_c for a given traffic λ, where we wish to (for example) minimize the mean delay, appears to be quite difficult. Exact evaluation of the stationary distribution is possible, but computationally demanding. More efficient algorithms for this computation are being investigated, which should help in the numerical search for the optimal control values. Similar problems are present in finding the optimal scheme within the class of strategies of the type of Hajek and van Loon [5]. Thus a precise comparison is difficult.

Nevertheless, we have simulated what appear to be representative cases. We simulated the specific case presented by Hajek and van Loon [5], and compared the results to a number of stable cases of the type (1)–(3). All cases were run for a traffic load of $\lambda = .32$, and, as suggested by Hajek and van Loon, used $\hat{b}_{\min} = 2$. We include the example case considered in the figures, as well as a somewhat better performing case that also uses $u_1 = 0$. We have also simulated the asymptotic MMSE scheme of Thomopoulos [17], which is distinguished by using $u_0 = 0$. We also include a case with $u_0 = u_1$, corresponding to the "Collision/No Collision" binary feedback channel. Confidence intervals for the mean delay (total time in system) were generated using the regenerative properties of the process [22]. Each simulation was run for a duration of 100,000 slots. The results of the simulations are given in Table I. Because of the significant variance in the statistics, one cannot assert with certainty the relative merit of each scheme based on the results of Table I. However, it does suggest that schemes of the type (1)–(3) may yield somewhat better performance than the strategy of Hajek and van Loon.

(u_0, u_1, u_c)	Mean Delay	95% Conf. Int.
H & vL*	13.2	[10.9, 15.2]
$(2-\epsilon, 0, 1)$	12.3	[10.4, 14.4]
$(-.8, 0, 1.2)$	11.8	[10.2, 13.3]
$(0, -.664, .797)$**	12.0	[10.0, 13.6]
$(-.4, -.4, .9)$	11.7	[9.8, 13.3]

Table I. Simulation comparison: $\lambda = .32$; 100,000 slots.

* Hajek and van Loon's scheme ($\gamma = .3$ in [5])
** Thomopoulos' asymptotic MMSE scheme [16]

VI. Conclusions

We have presented a stability analysis for a broad class of dynamic control strategies for slotted ALOHA systems. This class includes schemes that are stable for throughputs $\lambda < e^{-1}$. The control structure is simple to implement and allows the underlying system process to be modelled as a two-dimensional Markov chain. Conditions on the control system parameters were generated that guarantee stability in the sense that the limiting distribution of the system process exists for an arrival process that is independent of the system state (*i.e.*, for an infinite population of users with aggregate arrival rate below e^{-1}). Stability was proven by constructing a test function that satisfies sufficient drift properties. Simulation comparisons of alternative dynamic control procedures were tabulated, and suggest that schemes of the type investigated here will yield performance superior to alternative strategies; however, further analytical and/or simulation work is needed. Determination of the optimal control parameter values for u_0, u_1, u_c and $\hat{b}_{\min}$, for a given throughput λ, is an open problem.

Directions for further work are abundant. For example, one may consider the performance for a more complicated form of update than (3). Relaxation of system assumptions, such as the assumption of single packet buffers per user, will also lead to interesting areas of research. Additionally, it would be fruitful to extend these results to CSMA and other ALOHA-based protocols.

References

1. 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, September 1975.
2. A. Segall, "Recursive estimation from discrete-time point processes," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 4, pp. 422–431, July 1976.
3. M. Gerla and L. Kleinrock, "Closed loop stability controls for S-ALOHA satellite communications," *Proc. Fifth Data Comm. Symp.*, Snowbird, Utah, pp. 2-10—2-19, September 1977.

4. N. B. Meisner, J. L. Segal, and M. Y. Tanigawa, "An adaptive retransmission technique for use in a slotted-ALOHA channel," ***IEEE Trans. Commun.***, vol. COM-28, pp. 1776–1778, September 1980.
5. B. Hajek and T. van Loon, "Decentralized dynamic control of a multiaccess broadcast channel," ***IEEE Trans. Automat. Control***, vol. AC-27, no. 3, pp. 559-569, June 1982.
6. B. Hajek, "Hitting-time and occupation-time bounds implied by drift analysis with applications," ***Adv. Appl. Prob.***, vol. **14**, pp. 502–525, September 1982.
7. L. P. Clare, "Delay analysis of stable slotted ALOHA systems," ***Proc. IEEE INFOCOM 86***, Miami, Florida, April 1986.
8. A. B. Carleial and M. E. Hellman, "Bistable behavior of ALOHA-type systems," ***IEEE Trans. Commun.***, vol. COM-23, pp. 401–410, April 1975.
9. L. Kleinrock and S. S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation," ***IEEE Trans. Commun.***, vol. COM-23, pp. 410-423, April 1975.
10. G. Fayolle, E. Gelenbe, and J. Labetoulle, "The stability problem of broadcast packet switching computer networks," ***Acta Informatica***, vol. **4**, pp. 49–53, 1974.
11. M. Kaplan, "A sufficient condition for nonergodicity of a Markov chain," ***IEEE Trans. Inform. Theory***, vol. IT-25, no. 4, pp. 470–471, July 1979.
12. W. A. Rosenkrantz and D. Towsley, "On the instability of the slotted ALOHA multiaccess algorithm," ***IEEE Trans. Automat. Control***, vol. AC-28, pp. 994-996, October 1983.
13. G. Fayolle, E. Gelenbe, and J. Labetoulle, "Stability and optimal control of the packet switching broadcast channel," ***J. Assoc. Comput. Machinery***, vol. **24**, no. 3, pp. 375–386, July 1977.
14. A. G. Pakes, "Some conditions for ergodicity and recurrence of Markov chains," ***Oper. Res.***, vol. **17**, pp. 1059–1061, 1969.
15. S. C. A. Thomopoulos, "Decentralized estimation and control in random access communication networks: Stability and performance analysis," ***Proc. IEEE INFOCOM '85***, Washington, D.C., pp. 246–254, March 1985.
16. S. C. A. Thomopoulos, "A certainty equivalence nonlinear separation control rule for random access channels: Delay analysis," ***Conference Record, IEEE GLOBECOM'85***, pp.5.2.1-5.2.5, Dec. 2–5, 1985.
17. S. C. A. Thomopoulos, "Decentralized control for random access channels: Stability analysis and performance evaluation," ***Proc. ACC'85***, Boston, pp. 580-585, June 19–21, 1985.
18. R. L. Tweedie, "Criteria for classifying general Markov chains," ***Adv. Appl. Prob.***, vol. **8**, pp. 737–771, 1976.
19. John Lamperti, "Criteria for the recurrence or transience of stochastic process. I," ***J. Math. Anal. and Appl.***, vol. **1**, pp. 314–330, 1960.
20. J. F. C. Kingman, "The ergodic behaviour of random walks," ***Biometrica***, vol. **48**, no. 3 and 4, pp. 391–396, 1961.
21. V. A. Malyšev, "Classification of two-dimensional positive random walks and almost linear semimartingales," ***Soviet Math. Dokl.***, vol. **13**, no. 1, pp.136–139, 1972.
22. M. A. Crane and D. L. Iglehart, "Simulating stable stochastic systems, II: Markov chains," ***J. Assoc. Comp. Mach.***, vol. **21**, no. 1, pp. 114-123, January 1974.