Heavy Traffic Analysis of AIMD Models

Eitan Altman*and Harold J. Kushner†

July, 2003, Revised, March 2004

Abstract

We study heavy traffic asymptotics of many Additive Increase Multiplicative Decrease (AIMD) connections sharing a common router in the presence of other uncontrolled traffic, called "mice". The system is scaled by speed and average number of sources. With appropriate scalings of the packet rate and buffer content, an approximating delayed diffusion model is derived. By heavy traffic we mean that there is relatively little spare capacity in the operating regime. In contrast to previous scaled models, the randomness due to the mice or number of connections is not averaged, and plays its natural and dominant role. The asymptotic heavy traffic model allows us to analyze buffer and loss management policies of early marking or discarding as a function of the queue size and/or the total input rate and to choose a nearly optimal function via use of an appropriate limiting optimal control problem, captures the essential features of the physical problem, and can guide us to good operating policies. After studying the asymptotics of a large number of persistent AIMD connections we also handle the asymptotics of finite AIMD connections whose number varies as connections arrive and leave. The data illustrate some of the advantages of the approach.

Keywords: AIMD models, FTP analysis, heavy traffic analysis, approximating delay-diffusions, nearly optimal controls

1 Introduction

Background and motivation One of the most active research areas in networking in recent years has been the modeling and analysis of AIMD traffic; e.g., [1, 2, 3, 4, 6, 10, 15, 16, 17, 18, 20, 21]. When considering a single connection and modeling all other connections through an idealized loss process, simple

*The work of this author was partially supported by the Euro NGI network of excellence. INRIA, 2004 Route des Lucioles, B.P. 93, 06902 Sophia-Antipolis Cedex, FRANCE, eitan.altman@sophia.inria.fr

†The work of this author was supported by contracts DAAD-19-02-1-0425 from the Army Research Office and National Science Foundation Grant ECS 0097447, Division of Applied Mathematics, Brown University, Providence RI 02912, USA, hjk@dam.brown.edu

mathematical formulas for the connection's throughput can be obtained; e.g., [1, 6, 18, 21]. However, it is important in practice to understand the interaction of competing random numbers of connections and the associated system randomness that determines both the throughput as well as the losses suffered by the various connections. One approach is through a fixed point argument; see, e.g., [4]. If the loss rates over the nodes (or links) traversed by the connections are sufficiently small and can be assumed to be additive, an alternative framework can be used where the throughputs of TCP are obtained as the solution of a convex optimization problem and where the loss probabilities are obtained as the Lagrange multipliers [16, 21].

Although the methodologies in these references can be useful due to their simplicity, no dynamical systems description is provided; hence the actual "processes" do not appear, and it is very difficult to add dynamical (say, queue and packet rate dependent) controls to the formulation. The way that packet losses affect individual sources and the consequent effects on the full system are not modeled explicitly, and it is difficult to analyze the oscillations or instabilities that might be caused by delays. They cannot provide a sample-path or transient analysis. Models including some of these features appear in [2, 5] under simplified assumptions on the protocol's behavior (e.g., an assumption in [2] that loss probabilities do not depend on rates, or an assumption in [5] that all connections simultaneously lose a packet when the buffer is full). In order to analyze more complex systems that include buffer management, early marking or discarding, and the impact of the delay in the feedback loop, an alternative line of research has emerged based on fluid models using delay differential equations methodology; see e.g. [8, 17, 21]. In [20], a fluid model of the form of a delayed ordinary differential equation is obtained as a limit of a sequence of suitably scaled physical systems, as the number of connections and the speed of the system grows, and where the randomness is due to the varying number of non controlled connections. However, there is no randomness in the limit model. More detail on the relations between [20] and our work appears in Section 7.

Our goal is to analyze heavy traffic approximating models for multiplexing between AIMD and non controlled traffic, where the losses are a consequence of the actual underlying physical processes, as well as to determine good controls for buffer and loss management. The limit model should retain the main effects of the randomness of the physical processes, which determines the essential features of the buffer and loss processes. The limit model is not deterministic, but it is much simpler to handle than the original discrete stochastic system, and (as seen through numerical examples) it allows us to get good controls for buffer and loss management.

The basic ideas. As with many models for TCP, we will use a "stochastic fluid" model for describing the transmission process; i.e., rather than work explicitly with the widow size; we work with the number of packets that are allowed to be sent per unit time. We consider a model for AIMD traffic in the operating region where the system is near capacity. The analysis will be "asymptotic," as the system grows in speed. In particular, the bandwidth (speed of the router) as well as the mean number of users will be roughly proportional to a parameter n, which is to go to infinity. The analysis will be of the so-called heavy traffic (HT) type [13], which has been of considerable help in studying many complex queueing systems that would be intractable otherwise. Several formulations of the demand process are given. In all cases, there are a certain number of controlled users of the order of n, each having a lot of data to transmit. These share the channel with a large and randomly varying number of users with smaller amounts of data. These are commonly referred to as "mice." They are in the system for too short a time to be controlled, but might take a substantial (40% or more) of the total capacity. While each of the mice (resp., each of the controlled users) has identical statistical properties, this is only for convenience in the numerical analysis: Any number of classes can be handled.

The packets created by the various users enter the system in some random order, then are sent to a buffer via various links, from which they are transmitted. If the buffer capacity is exceeded, then a packet is said to be "lost." Until noted otherwise, the round trip delay α is the same for all AIMD users. The timing of the various rates are as seen at the buffer (not at the sources). They depend on the feedback sent from the buffer α units of time ago, which reached the source t_1 units of time ago, was then acted on and affected the rates at the input to the buffer t_2 units of time later, where $t_1 + t_2 = \alpha$.

We wish to identify a region of operation which is "near capacity" for large n, and a scaling under which the stochastic effects are apparent. One approach to asymptotic analysis is via a fluid model (e.g., [20]). These tend to average or eliminate the effects of stochastic variations in the number of users, mice, data rates, etc. But we are more concerned with demonstrating the actual random processes of losses and buffer content in terms of the random processes of arrivals, data levels, etc.

We are guided by the scaling used for heavy traffic models, as in [13]. There are two related aspects to being "near capacity." One is the difference between the mean packet creation rate and the speed of the system, and the other concerns the buffer size. Suppose that the total mean rate of arrival of packets to the buffer is vn. In order for the system to be in the heavy traffic regime, the speed of exiting the buffer would have to be slightly greater than vn, but not so much faster that the buffer is virtually empty almost all of the time. If the arrival process is the superposition of many independent users, then (loosely speaking) the standard deviation of the "randomness" would be $O(\sqrt{n})$. This suggests that if the system is near capacity at that time, then both the buffer size and the extra capacity would be $O(\sqrt{n})$. If either the buffer or the extra capacity are of a larger order, then the buffer level (scaled by $1/\sqrt{n}$) would go to zero as $n \to \infty$, and there would be no observable packet loss. These are the usual orders in heavy traffic analysis [13]. The amplitude scaling will be $1/\sqrt{n}$.

The heavy traffic regime is one important region of operation, one where small changes in the rates will have major consequences for buffer overflow (i.e, lost packets) and queueing delay. One can view the system as starting much below capacity, with a lower packet rate, and with the rates increasing until capacity is almost reached, at which point the control mechanisms are activated. Our analysis is confined to the time that the system is in this heavy traffic regime. In the comments at the end of Section 3, we will argue formally that the heavy traffic regime is very natural, and that a well regulated system will eventually find itself there. We will also argue formally that one should do the control problem with delays by allowing the controls to depend on the path over an appropriate time interval.

The controls. There are two classical types of rate control for each user. The first (the AI in AIMD) is the usual simple slow and steady linear increase in the allowed rate of packet creation when there are no buffer overflows. As noted above, in the heavy traffic regime, the number of controlled users is proportional to n on the average, and the excess capacity is $O(\sqrt{n})$. This suggests that the cumulative effect of the first type of control should be a rate increase of $O(\sqrt{n})$ over all controlled users, which implies a rate increase of $O(1/\sqrt{n})$ per user. If it were larger, the system would experience very serious packet losses in short order. Thus we suppose that there is a constant c such that the rate per source increases by $c/\sqrt{n}$. This is the correct order in the heavy traffic regime. See comments at the end of Section 3, where we conjecture that a well regulated system will eventually find itself in this situation.[1] It will turn out that the cumulative effects of this control and of the buffer overflow controls are of the same order. The second type of control (the MD part) is the usual multiplicative decrease when there is a lost packet.

To improve the performance, we also use another type of control, called a *preemptive control*, by which packets are selected at random to be "marked" as they enter the buffer. The chance of being selected depends on the buffer state and/or its input rate, and is a control function to be chosen. (Early discarding or marking has become very popular since it was proposed and deployed in the well known RED buffer management [7, 21].) The selection probability will increase when the system nears a dangerous operating point. There are two choices of how to handle the marked packets. Either they are deleted so that no acknowledgment is sent, or they are not deleted, but "modified" acknowledgments are sent back [19]. In either case, the source rate is decreased as though the packet were lost. This control, which anticipates the possibility of lost packets in the near future, can actually reduce the queueing delay as well as the rate of overflow considerably, with minimal cost in lost throughput; numerical data will illustrate this point. In either case, the use of the preemptive control helps avoid oscillations or instability due to the effects of bursts of lost packets caused by the delays. Here, we work with the second option, and do not delete the selected packets.

Outline of the paper. A general model for the mice is discussed in the next section. Two properties are paramount. One concerns the asymptotic (scaled) total number of packets that have been transmitted by them over any time interval. The other concerns the current rate of creation of packets. The assumptions are intuitively reasonable. To emphasize this, we discuss one particular exam-

[1]One could change the model, using fewer sources, each with a higher rate, and allow an accordingly faster increase in the AI control. The analysis would be similar.

ple in detail, starting from more "physical" assumptions. It is supposed (as is commonly done) that the mice enter with a fixed packet rate (possibly random among the individuals), but that they are in the system for a relatively short time, are not controlled and do not retransmit lost packets.

In Section 3, we consider the case where there are just n controlled users, analogously to the setup in [20]. Each of them has a very large (infinite, here) amount of data to be sent, and is subject to rate control. However, the randomness of the mice process has a significant effect on the total throughput, since it is a major cause of lost packets (buffer overflows), and the consequent rate control. The limit model is a delayed stochastic differential equation with boundary reflection. Note that a delayed reflection term arises. Section 4 considers various extensions of the basic model of Section 2, including the case where there is no buffer and where the rate for the controlled users changes randomly, perhaps due to reinitializations; this can be useful to model a sequence of TCP connections that are opened consecutively by the application layer, as is the case in the HTTP/1.1 version.

Section 5 deals with the case where the controlled users appear at random, each with a random amount of data to be sent, and vanish when their data has been transmitted. This introduces additional randomness, which (in the asymptotic limit) shows up via the addition of new Wiener processes in the dynamics for the rate process. Data that show some of the advantages of the approach and how to use it effectively are in Section 6.

2 The Model for the Mice

Recall that we use the name "mice" to describe any set of sources whose transmission rates are uncontrolled and with a relatively small number of packets/source. Various cases where the number of packets goes to infinity as $n \to \infty$ are covered by the assumptions. We suppose that the total rate at which mice packets are being put into the buffer at time t is $a_m n + \sqrt{n}\xi^n(t)$, where $a_m > 0$ and $\xi^n(\cdot)$ is a random process such that $\int_0^t \xi^n(s)ds$ converges weakly to a Wiener process $w_m(\cdot)$, with variance σ_m^2. More specifically (where $\Rightarrow$ denotes weak convergence),

$$\frac{(\text{total number of mice packets by } t) - n a_m t}{\sqrt{n}} = \int_0^t \xi^n(s)ds = w_m^n(t) \Rightarrow w_m(t), \tag{2.1a}$$

$$\begin{gathered} \frac{\text{mice rate}(\cdot) - n a_m}{n} = \frac{\xi^n(\cdot)}{\sqrt{n}} \Rightarrow \text{"zero" process}, \\ \sup_n E \sup_{s \le t} \left| \int_0^s \xi^n(\tau)d\tau \right| < \infty, \quad \text{each } t > 0. \end{gathered} \tag{2.1b}$$

(2.1a) says that the total mice packet rate is the sum of a "fluid" component and a part that is essentially independent over short and disjoint intervals. It is motivated by the central limit theorem. Owing to the complicated way that packets from different users are scrambled in transmission, it might be hard to say more, or to specify the "mice" model more explicitly. The sizes of the individual mice can grow with n, but slower than $O(n)$. All that we require is that (2.1) hold. An interesting specific example of a mice process is given next.

Example of a "mice" model. Consider the following example, which was one of the motivations for the general conditions above. The example is meant to be illustrative, and does not exhaust the possibilities. Suppose that the mice arrive as a Poisson process with rate $\lambda_m n$, with each arrival having an exponentially distributed (and independent among arrivals) amount of packets, with mean v_m/μ_m. The packets are put into the system at a rate v_m. The number of active mice at any time is $N_m^n(t)$, which satisfies

$$dN_m^n(t) = n\lambda_m dt - \mu_m N_m^n(t)dt + dM_m^n(t),$$

where $M_m^n(\cdot)$ is a martingale with quadratic variation process $\int_0^t [n\lambda_m + N_m^n(s)\mu_m]ds$. Let us work with the stationary processes. Then $N_m^n(\cdot)/n$ converges weakly to the process with constant values λ_m/μ_m. The rate at which mice packets arrive is $N_m^n(t)v_m$. Write $N_m^n(t) = n\lambda_m/\mu_m + \sqrt{n}\eta_m^n(t)$. Then

$$d\eta_m^n(t) = -\mu_m \eta_m^n(t)dt + dM_m^n(t)/\sqrt{n}.$$

The process $M_m^n(\cdot)/\sqrt{n}$ converges weakly to a Wiener process $\tilde{w}_m(\cdot)$ with variance $2\lambda_m$. The process $\eta_m^n(\cdot)$ converges weakly to $\eta_m(\cdot)$, where $d\eta_m(t) = -\mu_m\eta_m(t)dt + d\tilde{w}_m(t)$. The "noise part" of the arrival rate process for the mice satisfies

$$\xi^n(s) \equiv \frac{v_m N_m^n(\cdot) - n v_m \lambda_m/\mu_m}{\sqrt{n}} \Rightarrow v_m \eta_m(\cdot) \equiv \xi(\cdot).$$

Note that (2.1b) holds.

The variance of (scaled mice packet rate at t)$/\sqrt{n}$ is, asymptotically, $v_m^2\lambda_m/\mu_m$. The (scaled packet rate) correlation function is this times $e^{-\mu_m t}$. For high speed systems, both μ_m and v_m are large, while the ratio v_m/μ_m (the mean number of packets per mouse) is "moderate." In this case, to show that (2.1a) holds "approximately," write (neglecting the initial condition),

$$\eta_m(t) = \int_0^t e^{-\mu_m(t-s)} d\tilde{w}_m(s),$$

$$\begin{aligned}\int_0^t \xi(s)ds &= v_m \int_0^t \int_0^s e^{-\mu_m(s-\tau)} d\tilde{w}_m(\tau)ds \\ &= \frac{v_m}{\mu_m}\tilde{w}_m(t) - \frac{v_m}{\mu_m}\int_0^t e^{-\mu_m(t-s)} d\tilde{w}_m(s).\end{aligned}$$

The dominant part is the Wiener process. Thus, in (2.1a), $a_m = v_m\lambda_m/\mu_m$ and the variance of the Wiener process is $\sigma_m^2 = 2\lambda_m[v_m/\mu_m]^2$. The stationary variance of the error process (the last term on the right) is $(v_m^2/\mu_m^2)\lambda_m/\mu_m$. For large μ_m and moderate σ_m^2 the error process is close to the "zero" process, in that it converges weakly to it as $\mu_m \to \infty$.

We could also suppose, alternatively, that the individual mice send their packets all at once, but they are interleaved randomly with those from other sources along the way; then we come even closer to (2.1).

3 n Controlled Users, Each With Infinite Backlog

In this section, there are a fixed number, namely n, of controlled users, with each having a very large (infinite here, for modeling simplicity) amount of data to be sent. Let $r_i(t)$ denote the rate for controlled source i at time t, and suppose that there are positive a_i such that $a_0 \leq r_i(0) \leq a_1$, so that no single source dominates. Thus $\int_0^t r_i(s)ds$ is the total number of packets generated by controlled source i by time t. Define $\bar{r}^n(t) = \sum_{i=1}^n r_i(t)/n$, and $v_1 = \bar{r}^n(0)$, $v_2 = \sum_i [r_i^n(0)]^2/n$, and $\rho^n(t) = \left[\sum_i r_i(t) - nv_1\right]/\sqrt{n}$. Thus $\sqrt{n}\rho^n(t)$ is the rate at time t, centered about the initial mean rate nv_1. The analysis commences at the point at which the HT regime is entered.[2] The service rate (channel speed in packets per second) is assumed to be $C^n = nv_1 + a_m n + b\sqrt{n}, b > 0$, which covers the mean requirements (for both persistent connections as well as the mice process) and gives an excess (over the mean requirements) of $b\sqrt{n}$. The buffer size is $B\sqrt{n}$. These are the correct orders in HT analysis [13]. If the buffer or spare capacity were of a larger order, then the number of buffer overflows, asymptotically, would be zero.

When the buffer overflows (i.e., a packet is lost), that packet is assumed to come at random from the various users, in proportion to their individual current rates of packet creation: The various users (mice and controlled) would send their packets in some order, and the order would be more or less scrambled in the course of transmission, so that buffer overflows can be assigned at random to the various users.

As noted in the introduction, the standard multiplicative decrease control is activated by lost packets. I.e., there is some constant $\kappa \in (0, 1)$ such that, if the dropped packet at time $t - \alpha$ was from connection i, then the rate $r_i(t-)$ at $t-$ is changed to $r_i(t) = (1 - \kappa)r_i(t-)$.

The "preemptive" control. The performance would be improved if the sources were also signaled to reduce their rates as the buffer level or total input rate increases, but before actual buffer overflow. The type of control, called the *preemptive control,* attempts to do just this, analogously to what is done in the RED system. It selects packets on arrival, either at random or in some deterministic way according to the chosen control law. For notational simplicity, we suppose that the selection is done randomly. The probability that a packet entering the buffer at time t is selected is $u(t)/\sqrt{n}$, where $0 \leq u(\cdot) \leq u_{max} < \infty$ is a measurable control function, and is to be selected. The selected packets could be deleted as if there was an overflow. A preferable alternative, which we use, does not delete the packets, but returns a modified acknowledgment,

[2] **More on this later.**

which is used to reduce the flow at the source, similarly to what would happen if the packet were actually lost [19]. Let $\mathcal{F}_t^n$ denote the minimal σ-algebra that measures the systems data to time t. Then $u(\cdot)$ is $\{\mathcal{F}_t^n, t < \infty\}$-adapted. We suppose that there is a $\kappa_1 \in (0,1)$ so that if a packet from source i is selected at time $t-\alpha$, then $r_i(t) = (1-\kappa_1)r_i(t-)$. This preemptive control is to be chosen by the system designer and, when suitably selected, it can have a major beneficial effect on the overall operation.

Buffer input-output equations. We have

$$\begin{aligned}\rho^n(t) = \rho^n(0) + ct - &[\text{overflow control effects}] \\ &-[\text{preemptive control effects}],\end{aligned} \tag{3.1}$$

Let $x^n(t)$ denote $1/\sqrt{n}$ times the number of buffered packets at time t. Then

$$x^n(t) = x^n(0)\ + [(\text{total input - total output - overflow) by } t]\,/\sqrt{n}.$$

If the buffer is not empty, then its output rate is C^n. For modeling purposes, it is convenient to use this output rate all the time, even if the buffer is empty. Then we must correct for the "fictitious" outputs when the buffer is empty. This is done by adding an "underflow" correction term $L^n(t)$ (which is the number of fictitious outputs sent when the buffer is empty) as is usual in heavy traffic analysis [13]. Let $U^n(\cdot))$ denote $1/\sqrt{n}$ times the buffer overflow. Now, using the definition of $C^n, \rho^n(\cdot)$, and the mice model (2.1a), we can write

$$x^n(t) = x^n(0) + \int_0^t [\rho^n(s) - b + \xi^n(s)]\,dt - U^n(t) + L^n(t). \tag{3.2}$$

The limit dynamical equations. The following theorem gives the HT limits, and identifies the limit control system. Define $\hat{u}^n(t) = \int_0^t u^n(s)ds$.

Theorem 3.1. *Assume the mice model* (2.1)*, that* $C^n = nv_1 + a_m n + \sqrt{n}b$*, and that* $\sup_n |\rho^n(0)| < \infty$*. Then the sequence* $\{x^n(\cdot), \rho^n(\cdot), \hat{u}^n(\cdot), w_m^n(\cdot), L^n(\cdot), U^n(\cdot)\}$ *is tight in the Skorohod topology. For any weakly convergent subsequence, there is a process* $u(\cdot)$ *such that the weak sense limit* $(x(\cdot), \rho(\cdot), \hat{u}(\cdot), w_m(\cdot), L(\cdot), U(\cdot))$ *satisfies*

$$d\rho(t) = cdt - v_2\Big[\frac{\kappa}{v_1 + a_m}dU(t-\alpha) + \kappa_1 u(t-\alpha)dt\Big], \tag{3.3}$$

$$x(t) - x(0) = \int_0^t [\rho(s) - b]\,ds + w_m(t) + L(t) - U(t), \tag{3.4}$$

where $\hat{u}(t) = \int_0^t u(s)ds$*. Let* $\mathcal{F}_t$ *denote the minimal* σ*-algebra that measures* $(x(s), \rho(s), w_m(s), u(s-\alpha), L(s), U(s), s \le t)$*. Then* $w_m(\cdot)$ *is an* $\mathcal{F}_t$*-Wiener process with variance* σ_m^2*,* $0 \le u(t) \le u_{max}$*, and* $u(t)$ *is* $\{\mathcal{F}_t, t < \infty\}$*-adapted.*

Comment on the limit equations. Equations (3.8) and (3.9) are suggestive even for more general models. They capture much of the essence of the AIMD and the preemptive control mechanisms, and retain the fundamental role of the randomness, all for an aggregated and scaled system. Equations (3.3) and (3.4) identify the correct limit control system. The asymptotic effects of the overflow and preemptive controls are in the given form. The control is admissible in that it is a delayed nonanticipative (with respect to the Wiener process $w_m(\cdot)$) function satisfying the appropriate bounds.

Proof. It follows from the proof of the reflection mapping theorem in [13, Theorems 3.4.1, 3.5.1] that there is a constant C such that, for each $0 \le T_0 < T < \infty$,

$$(L^n(T)-L^n(T_0))+(U^n(T)-U^n(T_0)) \le C \sup_{T_0\le t\le T}\left[x^n(T_0)+\int_{T_0}^t [\rho^n(s)+\xi^n(s)]\,ds\right]. \tag{3.5}$$

By the assumption on $\rho^n(0)$, $\sup_n E\sup_{s\le t}\rho^n(s) < \infty$ for each t. By this, the second line of (2.1b), and (3.5), we have $\sup_n EU^n(t) < \infty$. Thus the number of buffer overflows on any bounded interval is $O(\sqrt{n})$. Thus, since the association of overflow with source is random, we can neglect the possibility that any one source will have more than one overflow on any finite interval.

The Lipschitz condition in (3.5) and the tightness criterion in [13, Theorem 2.5.6] or [11, Theorem 2.7b] assures that the sequence $\{x^n(\cdot), \rho^n(\cdot), U^n(\cdot), L^n(\cdot)\}$ is tight in the Skorohod topology. The sequence $\{\hat{u}^n(\cdot)\}$ is obviously tight since $0 \le u^n(t) \le u_{max}$. The fact that some arguments are delayed is irrelevant.

We next approximate the overflow control effects in (3.1). Suppose that there is a single overflow at time $t-\alpha$. I.e., $\sqrt{n}dU^n(t-\alpha) = 1$. Let $I_i^n(t-\alpha)$ denote the indicator function of the event that the overflow is associated with controlled source i. Then $r_i(t) = r_i(t-)(1-\kappa I_i^n(t-\alpha))$ and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n [r_i(t)-r_i(t-)] = -\kappa\sum_{i=1}^n r_i(t-)I_i^n(t-\alpha)dU^n(t-\alpha). \tag{3.6}$$

The user with the lost packet is selected at random, with the probability that controlled user i is selected being (its rate divided by the total rate, all at $t-\alpha$)

$$f_i^n(t-\alpha) = \frac{r_i(t-\alpha)}{\sum_j r_j(t-\alpha)+na_m+\sqrt{n}\xi^n(t-\alpha)}. \tag{3.7}$$

Use (3.7) to center (3.6) about the conditional mean (given the $r_j(t-\alpha)$, $\xi^n(t-\alpha)$, and that $dU^n(t-\alpha) > 0$), and rewrite the right hand term of (3.6) as

$$-\kappa\sum_{i=1}^n r_i(t-)\frac{r_i(t-\alpha)dU^n(t-\alpha)}{\sum_j r_j(t-\alpha)+na_m+\sqrt{n}\xi^n(t-\alpha)} + dM_1^n(t), \tag{3.8}$$

where $M_1^n(\cdot)$ is the martingale

$$\int_0^t \kappa\sum_{i=1}^n r_i(s-)\left[f_i^n(s-\alpha)-I_i^n(s-\alpha)\right]dU^n(s-\alpha).$$

By the random association of buffer overflow to user, we can show that

$$E|M_1^n(t)|^2 = O(1)E\sum_{s\le t}|dU^n(s)|^2 = O(1/\sqrt{n})EU^n(t).$$

Hence $M^n(\cdot)$ converges weakly to zero, and the left side of (3.8) can be used for (3.6), as $n \to \infty$.

It was seen that we can neglect the possibility that any one source is associated with more than one overflow on any finite interval. Thus, in evaluating the left side of (3.8), we can suppose that $r_i(t-\alpha) = r_i(t-)$. Using this and (2.1b), and dividing each part of the term

$$\frac{\sum_i r_i(t-)r_i(t-\alpha)}{\sum_j r_i(t-\alpha) + na_m + \sqrt{n}\xi^n(t-\alpha)}$$

by n, we see that it converges weakly to the constant process, with values $v_2/[v_1 + a_m]$, as $n \to \infty$. The above computations imply that, as $n \to \infty$, the buffer overflow control term in (3.1) is well approximatable by $(\kappa v_2/[v_1 + a_m])U^n(t-\alpha)$.

Now, we turn our attention to approximating the effects of the preemptive control. Redefine $I_i^n(t)$ to be the indicator of the event that a packet selected at time t came from controlled source i. Define $R^n(t-\alpha) = \sum_j r_i(t-\alpha) + na_m + \sqrt{n}\xi^n(t-\alpha)$, the total (unscaled) packet arrival rate at time $t-\alpha$. Let $J^n(t)$ denote the number selected by t. Then

$$\frac{\kappa_1}{\sqrt{n}} \sum_i [r_i(t) - r_i(t-)] = \frac{\kappa_1}{\sqrt{n}} \sum_i r_i(t-) I_i^n(t-\alpha) dJ^n(t-\alpha). \tag{3.9}$$

The mean rate at which packets are selected at time $t-\alpha$ is

$$\frac{u(t-\alpha)}{\sqrt{n}} R^n(t-\alpha). \tag{3.10}$$

We can model the random selection times as the jump times of a jump process with conditional jump rate $u^n(t)R^n(t)/\sqrt{n}$ at time t. Thus on any finite interval there are only $O(\sqrt{n})$ selections, and the event that more than one comes from the same source can be neglected. Thus, if a selection at $t-\alpha$ comes from source i, we can suppose (without loss of generality) that $r_i(t-) = r_i(t-\alpha)$. The rest of the development is similar to that for the effects of the overflow control, but with $J^n(\cdot)$ replacing $U^n(\cdot)$. Thus, by centering $I_i^n(t-\alpha)$ at its conditional mean $f_i^n(t-\alpha)$, we have the representation of (3.9) as

$$\frac{\kappa_1}{\sqrt{n}} \sum_i \frac{r_i(t-)r_i(t-\alpha)}{R^n(t)} dJ^n(t-\alpha) + dM_p^n(t), \tag{3.11}$$

for a martingale $M_p^n(\cdot)$. The quadratic variation of the martingale is $O(1/\sqrt{n})$, hence it converges weakly to zero as $n \to \infty$. Now centering dJ^n about its conditional mean yields the approximation to the left hand term of (3.11) as

$$\frac{\kappa_1}{\sqrt{n}} \sum_i \frac{r_i(t-)r_i(t-\alpha)}{R^n(t)} \frac{u^n(t-\alpha)R^n(t-\alpha)}{\sqrt{n}} dt + dM_q^n(t), \tag{3.12}$$

where $M_q^n(\cdot)$ is a martingale whose quadratic variation is also $O(1/\sqrt{n})$, hence it is asymptotically negligible.

By what has been said,

$$v_2\kappa_1 \int_0^t u^n(s-\alpha)ds$$

approximates the effects of the preemptive control for large n. Now, with these asymptotic representations for the effects of the controls, we see that the limit of any weakly convergent subsequence of $\{x^n(\cdot), \rho^n(\cdot), \hat{u}^n(\cdot), U^n(\cdot), L^n(\cdot), w_m^n(\cdot)\}$, satisfies (3.3) and (3.4). The Wiener property of $w_m(\cdot)$ is just the assumption (2.1). The fact that it is an $\{\mathcal{F}_t, t < \infty\}$-Wiener process is proved using standard methods; for example see [13, Theorem 6.1.2]. The limit $\hat{u}(\cdot)$ is absolutely continuous with respect to Lebesgue measure, with derivative bounded by u_{max}; hence the asserted process $u(\cdot)$ exists. ■

Cost functions and nearly optimal controls for the physical system. In order to assure good performance of the AIMD connections, the buffer management would implement control $u(\cdot)$. The quantities to penalize in the cost are queueing delay (measured by $x(\cdot)$), the loss of throughput due to the control (measured by $-\rho(\cdot)$), and buffer overflow (measured by $U(\cdot)$)[3]. Let us work with a discounted cost criterion, where $\beta > 0$ can be as small as we wish, $c_0 > 0$, and the $k_i(\cdot) \geq 0$ are Lipschitz continuous:

$$W(u) = \beta E \int_0^\infty e^{-\beta t} \left([k_1(x(t)) - k_2(\rho(t))] \, dt + c_0 dU(t) \right). \tag{3.13}$$

The possibility that the $k_i(\cdot)$ are nonlinear can be useful, since (e.g.) we might wish to heavily penalize long queues, but not be too concerned with short queues.

Using the methods of heavy traffic analysis for controlled problems [13], it can be shown that the optimal costs for the physical problem converge to the optimal cost for the limit problem. If the delay is zero, then the optimal control for the limit problem is of the switching curve type: $u(x, \rho)$ takes the maximum value on one side of a switching curve and is zero on the other, and the switching curve is smooth. The switching curve character for $\alpha = 0$ follows from a formal examination of the Bellman equation for the optimal value, since the control appears linearly in the dynamics and does not appear in the cost. The smoothness was implied by the numerical computations. See, for example, Figure 6.1. Such switching optimal controls are nearly optimal for the physical system for large n. We note that the cost (3.13) is well defined, since it can be shown that $E|\rho(t)| + EU(t) \leq a_1 + a_2 t$, for some $a_i \geq 0$.

We shall also consider an ergodic cost criterion

$$\gamma(u) = \lim_{T\to\infty} E \frac{1}{T} \left[\int_0^T \left(k_1(x(t)) - k_2(\rho(t)) \right) dt + c_0 U(T) \right] \tag{3.14}$$

[3]Penalizing buffer overflow may be important for several reasons. First, if the mice correspond to real time applications, then these applications will suffer due to losses. Secondly, the AIMD themselves may correspond to real time applications which are "TCP friendly", in which case lost packets are typically not retransmitted. Losses due to overflow then again degrade the quality of the communication.

At present, there is little theory concerning stability or ergodicity theory for delayed reflected diffusions such as ((3.3), (3.4)), or ((3.4), (5.5)) for the model of Section 5. If the delay is zero then, for any feedback control $u(\cdot)$, stability can be shown and the model ((3.4), (5.5)) can be shown to have a unique invariant measure; see, e.g., [13, Chapter 4]. In the numerical computations (where zero delay was always used), we were always able to compute an optimal control for the ergodic cost criterion (with cost and control well approximated by those for the discounted problem for small β), and both stability and convergence to the stationary distribution under the optimal (or other reasonable) controls were apparent.

Comments on the heavy traffic regime. First, we comment on the control $u(t-\alpha)$ in (3.3). Suppose that there is a cost function of either the type (3.13) or (3.14). Owing to the delay, the optimal control $u(t)$ will not be simply a function of $(x(t), \rho(t))$, but rather a function of the path segment $\{x(s), \rho(s), t-\alpha \leq s \leq t\}$ [9, 22]. Although there is some progress with numerical methods for computing and approximating optimal $u(\cdot)$ when $\alpha > 0$, good algorithms are not yet available. Controls $u(t)$ that depend just on the value of $(x(t), \rho(t))$ are subject to oscillations. Those that depend as well on the recent past can avoid oscillations, since they "remember" recent control values and can select the current value accordingly. It is likely that using the full potential of the delay dynamical system can improve the operation and keep the system in the heavy traffic regime. Following are some conjectures. They are reasonable, but unproved at this time.

Suppose that the system starts far from the heavy traffic regime, and that the "slow" increase in packet rate is initially $\bar{c}$ for *each* of the n users. Then the total increase in the rate for the persistent connections is $n\bar{c}$. Capacity will be reached quickly. One can try to pose a control problem with another preemptive control, where losses are penalized heavily. One expects that (roughly speaking) at a time t when the average rate reaches a level where approximately α units of time later, it will be within $O(\sqrt{n})$ of capacity, this control starts to act, and selects packets for the modified acknowledgment. The packet rate into the buffer will keep increasing until the rate reducing effects of the feedback reach the buffer α units of time later. After time t, since the control "knows" the recent state values, it knows how many packets it has already selected, and adjusts new selections accordingly. The $\sqrt{n}$ level is used because that represents the effects of the randomness. The control will start to act when asserted since otherwise there will be large losses. It is reasonable to expect that such behavior would avoid oscillations and bring the system to the heavy traffic regime, where an additional fine control can be exerted. Similar comments apply to the case where the delay depends on the user. We conjecture that a fuller development and exploitation of control theory when the controls are delayed will have a major impact.

4 Extensions of the Model of Section 3

No Buffer. Suppose that there is no buffer, so that if the total current packet rate exceeds the channel speed, then the excess packets are rejected. The forms of the input processes and channel speed (service rate) are as in the last section, but in lieu of (2.1), we assume that $\xi^n(\cdot)$ converges weakly to a process $\xi(\cdot)$, as in the example in Section 2. Since there is no buffer to overflow, the "reject" process $U^n(\cdot)$ needs to be defined. Define

$$y^n(t) = \left[C_n - \left(\sum_i r_i(t) + a_m n + \sqrt{n}\xi^n(t)\right)\right] \Big/ \sqrt{n} = \left[b - \rho^n(t) - \xi^n(t)\right], \tag{4.1}$$

the scaled difference between the channel speed and input packet rate at t. Then the scaled number of rejected packets is

$$U^n(t) = \int_0^t \left[y^n(s)\right]^- ds = \int_0^t \left[\xi^n(s) + \rho^n(s) - b\right]^+ ds \tag{4.2}$$

Suppose that the correlation time of $\xi^n(\cdot)$ is short (e.g., large μ_m in the special mice model) of Section 2. Then a law of large numbers argument can be used to show that $\xi^n(t)$ can be "integrated out" of (4.2), in that, as $n \to \infty$ and the correlation time goes to zero, the integrand can be replaced by the average over $\xi^n(t)$. This simplifies the expression for $U^n(\cdot)$, and the limit equation for the scaled and centered rate process $\rho^n(\cdot)$ is

$$\dot{\rho}(t) = cdt - v_2\Big[\frac{\kappa}{v_1 + a_m}\dot{U}(t-\alpha) + \kappa_1 u(t-\alpha)dt\Big], \tag{4,3}$$

where $\dot{U}(t) = E_\rho\left[\xi(t) + \rho(t) - b\right]^+$, and the expectation is over the $\xi(t)$. Then we have a deterministic limit, which is not a priori obvious. The randomness due to the mice in the arrival process does not appear explicitly in (4.3), but it affects the value of the expectation that yields the overflow rate $\dot{U}(t)$.

Let us look a little more closely at the example in Section 2. As $\mu_m \to \infty$, we would also have that $v_m \to \infty$, to keep the total mean data per mouse from going to zero. If we suppose that $\lambda_m(v_m/\mu_m)$ is bounded, then the stationary variance of $\xi(\cdot)$ would be $O(v_m)$, which implies that the excess capacity factor b would have to be $O(\sqrt{v_m})$, if large loss rates are to be avoided. This is useful scaling information. It is hardly surprising, since we no longer have the buffer to "integrate" the mice process, and we must deal directly with the large variations in the $\xi^n(\cdot)$ instantaneous rate process. In order to avoid huge losses, the excess capacity must be some large constant times the standard deviation of this process.

Finally, we note that the value of an optimal control at time t for the limit process (which will be applied at time $t+\alpha$) need only depend on $\rho(s), t-\alpha \le s \le t$. Since such controls are nearly optimal for the physical process in heavy traffic, we see that an nearly optimal control can depend only the rate $\rho^n(\cdot)$ on $[t-\alpha, t]$, and not on the more rapidly changing "mice" process $\xi^n(\cdot)$.

Analogous results holds for the model of Section 5, but there the randomness due to the arrival and departure processes of the controlled users remains in the limit. As for the case of concern above, only the $\xi^n(t)$ process would be "integrated out."

Random $r_i(0)$. In the rest of this section, we suppose that there is a buffer, as in the problem of Section 3. Suppose that the initial values of the rates are random, identically distributed, and mutually independent, with $Er_i(0) = v_1$ and $E[r_i(0)]^2 = v_2$. Then all the asymptotic results continue to hold.

Randomly changing rates. In some internet applications, where a user sends a sequence of consecutive TCP connections, the rate of transmission is reinitialized for each new TCP transfer (e.g. HTTP/1.1). We next propose a model of which this scenario is a special case. Suppose that the users change the packet transmission rates at random, and each with rate λ_0. The new rates (which are uniformly bounded) are chosen randomly with the same first two moments. More precisely, there are mutually independent Poisson processes $P_i(\cdot)$ all with rate λ_0. When $P_i(\cdot)$ jumps, the rate for user i is replaced. The set of replacements, over all users and time, is mutually independent, and independent of all other "driving" processes. Let q denote the canonical rate replacement, and define $v_1 = Eq$, $v_2 = Eq^2$, $\bar{v}_2 = E[q - v_1]^2 = v_2 - v_1^2$, and $Q^n(t) = \sum_i r_i(t)$. Then

$$dQ^n(t) = \sqrt{n}cdt - [\text{effects of controls}] - \lambda_0 \left[Q^n(t) - nv_1\right] dt + dM_r^n(t), \quad (4.3)$$

where the martingale $M_r^n(\cdot)$ can be shown to have quadratic variation process

$$\lambda_0 \int_0^t \sum_i E\left[r_i(s) - q\right]^2 ds = \lambda_0 n \int_0^t \left[\frac{\sum_i r_i^2(s)}{n} - \frac{2v_1 \sum_i r_i(s)}{n} + v_2\right] ds, \quad (4.4)$$

where the expectation is over q only. Recall the definition $\rho^n(t) = [Q^n(t) - nv_1]/\sqrt{n}$. It can be shown that $\rho^n(\cdot)$ is tight and that the limit of any weakly convergent subsequence satisfies

$$d\rho(t) = cdt - v_2\Big[\frac{\kappa}{v_1 + a_m} dU(t-\alpha) + \kappa_1 u(t-\alpha)dt\Big] - \lambda_0 \rho(t)dt + dw_r(t), \quad (4.5)$$

where the Wiener process $w_r(\cdot)$ has variance $2\lambda_0 \bar{v}_2$. The limit system equations are (3.4) and (4.5). The control is as in Theorem 3.1.

Delay depending on the user. Up to now, all users had the same delay. The general theory can handle user-dependent delays. Suppose that user i has delay $\alpha_i \le D < \infty$. Let the buffer overflow at time s with a packet from user i. The information will reach user i at time $s + t_{1,i}$. Thus, at time t, user i receives information concerning overflows at time $t - t_{1,i}$, and its response reaches the buffer $t_{2,i}$ units of time later, with $t_{1,i} + t_{2,i} \equiv \alpha_i$. This leads to $dU^n(t-\alpha)$ in the ith summand in (3.8) being replaced by $dU^n(t-\alpha_i)$. To simplify matters in this brief presentation, first suppose that all initial rates are equal: $r_i(0) = v_1$.

Then, for large n, the main term in (3.8) is approximately

$$
\begin{aligned}
& -\kappa \sum_{i=1}^{n} r_i(t-) \frac{r_i(t-\alpha_i) dU^n(t-\alpha_i)}{\sum_j r_j(t-\alpha_i) + n a_m + \sqrt{n}\xi^n(t-\alpha_i)} \\
\approx \quad & -\frac{\kappa v_1^2}{a_m + v_1} \frac{1}{n} \sum_{i=1}^{n} dU^n(t-\alpha_j).
\end{aligned}
$$

More succinctly, with $\beta^n(\cdot)$ being a measure with mass $1/n$ at α_i, write

$$
\frac{v_1^2}{n} \sum_{i=1}^{n} dU^n(t-\alpha_i) = v_1^2 \int_0^D dU^n(t-\alpha)\beta^n(d\alpha).
$$

Suppose that the distribution of delays $\beta^n(\cdot)$ converges weakly to a distribution $\beta(\cdot)$. Then the $dU(t-\alpha)$ in (3.3) is replaced by $\int dU(t-\alpha)\beta(d\alpha)$. All else remains the same. If the $r_i(0)$ are not all equal, then redefine $\beta^n(\cdot)$ to have weight $r_i^2(0)/n$ at α_i, suppose that $\beta^n(\cdot) \Rightarrow \beta(\cdot)$, and replace the right side of the last expression by $\int_0^D dU(t-a)\beta(d\alpha)$. Details of the proof are omitted.

5 A Stochastic Process of Finite AIMD Connections

In the model of Section 3, the number of users is fixed at n. Now, we consider a model where the controlled users arrive independently and randomly and leave at random, with the arrival process independent of the mice process. New users come from an unlimited population, with (Poisson) arrival rate λn. Each new user comes with an exponentially distributed number of data packets, each with mean v_1/μ, and independent of the mice process and arrival times.[4]

With this model, as with the previous ones, the buffer overflows (i.e., packet losses) are created by the physical process and not imposed. Note that the mean amount of data in a new source does not depend on n. The parameter n scales

[4]Exponential distribution of interarrival times and session duration are more appropriate for telephone calls than for data connections. Thus this model is expected to be more useful for VoIP applications that use TCP friendly mechanisms to regulate their rate. The "exponential" assumptions can be helpful even for the data connections for some preliminary dimensioning purposes.

Non exponential distributions can be handled as well, with an increase in the dimensionality of the limit model. For example, a k-stage Erlang model would require a k-dimensional process to represent the rate process. The mathematical development and results are similar. This higher dimensionality is a handicap for numerical computations, say via the Markov chain approximation method [12], or a pathwise approximation method. But it is not a serious handicap for simulation. Indeed, simulating the approximating limit model is substantially simpler than simulating the physical process, when there are very many users.

Experimentation with the basic model can lead to insights that are useful for more general cases. For example, numerical results for the basic model with no delay indicate that threshold controls, based on the rate only, provide good approximations to the values obtained by optimal controls. This observation provides a basis for getting good controls, which would be very hard to compute otherwise, for more general large size systems.

the system speed and mean number of users only.[5] The source (i.e., the user) stays "active" until all data is sent, and then disappears. Time is still measured at the buffer and the mice model is (2.1). For simplicity, suppose that the initial rate of each new controlled source is v_1.

First suppose that there are no controls (constant transmission rate from each source) and buffer overflows are not retransmitted. Then the packets are sent from each active source to the buffer at a rate v_1. The mean time that a source is active is $1/\mu$, and the total rate at which the sources drop out at t is $\mu N^n(t)$, where $N^n(t)$ denotes the number of active sources. The (stationary) mean number of sources in the system is $n\lambda/\mu$. Hence, the analog of the channel speed C^n of Section 3 is $C^n = v_1 n[\lambda/\mu] + a_m n + b\sqrt{n}$, where, again, $b\sqrt{n}$ denotes the excess capacity over the mean rate $n[v_1\lambda/\mu + a_m]$. On departure of a user, its rate v_1 is lost.[6] We suppose that $1/\mu$ is large enough relative to the delay α so that there is enough time for many round trips.

Now suppose that the input rates from the non-mice sources are actually controlled. There are several approaches that one can take for the source departure process. One approach supposes that the departure rate (of an AIMD connection) is μ, and does not depend on the current packet transmission rate for the source. Then the lost packet rate if connection i leaves is $r_i(t)$. This situation arises when the AIMD connections correspond to real time applications that have a dynamic compression rate (which is then "TCP friendly"). In these applications, lost packets are not retransmitted (the possibility of lost packets might be anticipated in the coding). For simplicity in the development, this is the approach that will be taken. [7]

The dynamics and limit for the rate process. The details are similar to those in Section 3, except for the treatment of the randomness due to the arrivals and departures for the controlled users, and we will concentrate on this point. Write $N^n(t) = n\hat{N} + \sqrt{n}\nu^n(t)$, $\hat{N} = \lambda/\mu$. Since the user arrival process is Poisson and the departure rate per user is constant,

$$dN^n(t) = \lambda n dt - \mu N^n(t)dt + dM_a^n(t) - dM_d^n(t). \tag{5.1}$$

Here $M_a^n(\cdot)$ is the martingale associated with the arrival process and has quadratic variation process $n\lambda t$, and $M_d^n(\cdot)$ is the martingale associated with the departure process and has quadratic variation process $\mu\int_0^t N^n(s)ds$. For simplicity, suppose that $N^n(\cdot)$ is stationary. It follows from this, (5.1), and the cited values

[5] The rate of arrivals of new users can be a smaller order of n, and then they would each have an amount of data that would depend on n. E.g., rate of arrival $O(\sqrt{n})$, with data $O(\sqrt{n})$. In this case the rate of work on each source is $O(\sqrt{n})$, so that the average sojourn in the system is still $O(1)$.

[6] Strictly speaking a source should not depart until an acknowledgment of its last transmission has been received. But our approximation to the actual departure rule has little effect, since the order of lost packets is still $O(\sqrt{n})$, and μ is large.

[7] An alternative approach replaces the value of μ by a time varying quantity to reflect the fact that even if the service rate per source changes the total amount of data per source doesn't. For example, if the allowed data rate for an AIMD connection is cut in half due to an increase in the number of sources, then the value of the connection departure rate for that source should be cut in half. The mathematical development of this situation is much harder.

of the quadratic variations, that the sequence $N^n(\cdot)/n$ converges weakly to a process with constant value $\hat{N} = \lambda/\mu$, as $n \to \infty$. Also, $\nu^n(\cdot)$ satisfies

$$d\nu^n(t) = -\mu\nu^n(t)dt + [dM_a^n(t) - dM_d^n(t)]/\sqrt{n}. \tag{5.2}$$

The quadratic variation of the scaled martingale term in (5.2) is $\lambda t + \mu\int_0^t N^n(s)ds/n$, which converges weakly to $2\lambda t$. The sequence $\nu^n(\cdot)$ converges weakly to $\nu(\cdot)$, where

$$d\nu(t) = -\mu\nu(t)dt + dw(t), \tag{5.3}$$

where $w(\cdot)$ is a Wiener process with variance 2λ.

Returning to the rate process, write $\sum_i r_i(t) = Q^n(t) = n\hat{R} + \sqrt{n}\rho^n(t)$, where $\hat{R} = v_1\lambda/\mu$. The process $Q^n(\cdot)$ satisfies

$$\begin{aligned} dQ^n(t) &= \lambda v_1 n dt - \mu Q^n(t)dt + c\sqrt{n}dt \\ &\quad -[\text{effects of overflow and preemptive controls}] + v_1 dM_a^n(t) - dM_{d,1}^n(t), \end{aligned} \tag{5.4}$$

where $M_{d,1}^n(\cdot)$ is the martingale associated with the "rate departure" process and it has quadratic variation process $\mu\int_0^t \sum_i r_i^2(s)ds$. This, divided by n, converges weakly to the process with values $v_1^2\lambda t$, as $n \to \infty$. Finally, following the procedure used in the proof of Theorem 3.1, it is not hard to show that $\rho^n(\cdot) \Rightarrow \rho(\cdot)$, where

$$\begin{aligned} d\rho(t) = -\mu\rho(t)dt + [\lambda/\mu]cdt - \frac{v_1^2\kappa[\lambda/\mu]}{v_1[\lambda/\mu] + a_m} dU(t-\alpha) \\ +v_1^2\kappa_1[\lambda/\mu]u(t-\alpha)dt + v_1 dw. \end{aligned} \tag{5.5}$$

Approximations to the optimal via the limit model. The limit system equations are (3.4) and (5.5). The comments made after Theorem 3.1 concerning the convergence of the optimal costs for the physical problem to that for the limit also hold here.

6 Numerical Data: Optimal Preemptive Controls

It is not possible at present to compute optimal policies when there is a delay in the control (although there is promising work being done on the development of numerical algorithms), so we set $\alpha = 0$. The results still shed light on the system behavior when the delay is small relative to the time constant in (3.3) or (5.5).

Numerical results were obtained for the optimal control and costs for the model of Section 5 with the cost function being either (3.13) or (3.14), with $k_1(x) = c_1 x$, $k_2(\rho) = c_2\rho$. The results for the two cost functions were nearly the same when $\beta \leq .02$, and the ergodic case will be described. The numerical method was the Markov chain approximation method [12], which is the most

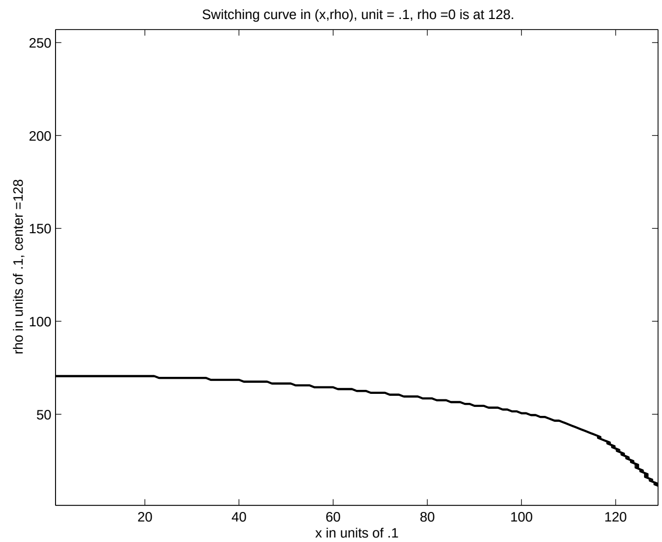

Figure 1: A switching curve; no delay.

versatile current approach for controlled reflected diffusions. Only a few details can be given here. Use $c = 1, \lambda/\mu = 4, b = 1, v_1 = 1.5, a_m = 4, \kappa = \kappa_1 = .5$, $\sigma_m^2 = 4$. Since there is no delay the control is a function of $(\rho(t), x(t))$. We used the bound $0 \leq u(x, \rho) \leq 1$. The buffer capacity is $12.8\sqrt{n}$ packets, and $c_0 = 100, c_1 = 1, c_2 = 5$, reflecting our desire to penalize lost packets most heavily. The mice account for about 40% of the traffic and the system is quite "noisy," since the variances of the Wiener processes driving (x, ρ) are $(9, 4.5)$.

The optimal preemptive controls are determined by a switching curve: $u(x, \rho) = 0$ below the curve and equals its maximum value above the curve. The curve obtained for our example in the asymptotic regime is given in Figure 6.1. As we see, in (x, ρ) space, the curve is initially (for small x) almost a straight line with a slightly decreasing slope as x increases. As the buffer fills up, the slope becomes sharply more negative, as expected. The optimal cost for the problem with preemptive control was about 1/10th of that without. In general, The values of the cost components (stationary mean values of $x(t), \rho(t)$, and $\lim_{t\to\infty} EU(t)/t$) are more significant than the optimal cost, since they give us information on the tradeoffs. Optimal control is not of interest for its own sake, but rather for the information provided on good design, and tradeoffs among the cost components as the weights change.

For the uncontrolled problem, the sum of the buffer overflow rate for all users was $5.35\sqrt{n}$, vs. $0.28\sqrt{n}$ under the optimal control for the given cost coefficients. The mean queue was virtually full for the uncontrolled case, compared to an average of one-third full under optimal control. The total input rate for the controlled users was reduced by an average of $0.36\sqrt{n}$ under optimal preemptive control, compared with an increase of $6.3\sqrt{n}$ with no control. Thus to get an improvement in overflow of about 20 times cost a fractional reduction in the throughput of $(6.3 + 0.36)/[(v_1\lambda/\mu) + a_m]\sqrt{n}) = 0.666/\sqrt{n}$.

If the buffer size is increased, its average percentage occupancy is about the same (queue size is not weighted heavily), the average $E\rho$ increases, and the average overflow rate does not change dramatically (e.g., doubling the buffer only halves the overflow, under our parameters). The optimal system adapts to an increased buffer size mainly by increasing the average flow, keeping the queue size roughly in proportion to the buffer size, an interesting fact in itself. Of course, a larger weight on x will reduce the average queue size.

These numbers illustrate the type of tradeoffs that are possible. One pays for reduced overflow by reduced packet rate. But the packet rate is reduced only where it does the most good. The tradeoffs vary with the cost coefficients. To use the method effectively, one makes a series of runs, varying the coefficients c_i This yields a set of possible tradeoffs between the competing criteria. In each case, the tradeoff is under an optimal control. The approach to the use of numerical methods and heavy traffic approximations is similar to what was done for the problem of input control of a multiplexer system in [14]. A comparison with threshold controls shows that the effects of the optimal control can be well approximated by a threshold control depending on ρ only, for appropriate values of the threshold. The cost components for the no control, optimal, and threshold cases are summarized in Table 1. If the threshold controls are activated only when the buffer exceeds some modest level, their performance is even better. Keep in mind that the described optimal control and costs are for a very heavy weight on overflow.

Table 1. Cost components.

under run type	buf overflow/$\sqrt{n}$	Ex	$E\rho$
no cont.	5.35	11.92	6.35
opt. cont.	.28	4.4	-.36
thresh $\rho = 0$	.69	7.6	1.46
thresh $\rho = -1$	.48	6.4	.98
thresh $\rho = -3$	.33	4.9	.2

7 Appendix: Comparison With a Fluid Model

Reference [20] also concerned a limit approximation for large systems and justified the use of a delayed deterministic differential equation as an approximation for a certain class of problems. Since there are major differences between that work and this, apart from the different scaling, and since that paper is the main other current work on the use of limit-delay equations for AIMD models, a brief discussion of some of the differences is worthwhile.

In the basic model of Section 3, capacity (i.e., bandwidth) scales linearly with n, and so does the number of sources. The packet rate for each source is $O(1)$. Our general approach also allows the possibility that the number of sources grows more slowly with n, with the packet rate per source growing accordingly faster. While there are no explicit capacity constraints in [20], it is clear that the bandwidth (BW) is proportional to their n^2, and we use this fact below. They use a fixed number of connections of the order of $\sqrt{\mathrm{BW}}$ (and no analog of the models of Sections 4 and 5), each sending packets at rate $O(\sqrt{\mathrm{BW}})$. The number of mice connections grows linearly with $\sqrt{\mathrm{BW}}$, and so does the rate of each mouse. Time is divided into "decision intervals" of length $O(1/\sqrt{\mathrm{BW}})$, and the rates are (perhaps unrealistically) averaged over these successive intervals before feedback and decisions. This averaging over $O(\sqrt{\mathrm{BW}})$ packets before feedback effectively eliminates the randomness due to the mice. We work closer to system capacity where the effects of random variations are greater, and it is the true instantaneous randomness that causes the losses and activates the controls.

The total overall rates of increase of the packet rate due to the slow additive control is the same here and in [20]. In [20] the "slow constant rate of increase of the packet rate" of each connection in the nth model (the one corresponding to n TCP connections) increases by $1/n$ per each time slot, so that in terms of real time the total rate of increase does not depend on n. Thus the total rate of increase is of the order of $\sqrt{\mathrm{BW}}$, as in our case. In our model, the packet loss of each AIMD source is random and determined by the loss process associated with that source. This is in conformance with the objectives of buffer management schemes [7]. In [20], in contrast, all AIMD sources have the same instantaneous dynamics, hence identical losses. An important advantage of the work in [20] is that the model, being deterministic, is much simpler. Hence, under its assumptions, one can more conveniently explore some of the effects of delays.

References

[1] E. Altman, K. Avratchenkov and C. Barakat, "A stochastic model of TCP/IP with stationary random losses", *ACM SIGCOMM 2000*.

[2] F. Baccelli and D. Hong, "A.I.M.D, Fairness and Fractal Scaling of TCP Traffic" Technical Report, April 2001, RR-4155, INRIA Rocquencourt, France, 2001.

[3] S. Ben Fredj, T. Bonald, A. Proutiere, G. Regnié and J. W. Roberts, "Statistical bandwidth sharing: a study of congestion at flow level", *SIGCOMM'*, 2001.

[4] T. Bu and D. Towsley, "Fixed point approximation for TCP behaviour in an AQM network", *ACM SIGMETRICS*, June 2001.

[5] P. Brown, "Resource sharing of TCP connections with different round trip times", *IEEE Infocom*, Mar 2000.

[6] V. Dumas, F. Guillemin and P. Robert, "A Markovian analysis of AIMD algorithms", *Advances in Applied Probability*, 34(1) 85-111, 2002.

[7] S. Floyd and V. Jacobson, "Random Early Detection gateways for Congestion Avoidance" *IEEE/ACM Transactions on Networking*, 1(4):25–39, 1993.

[8] C. Hollot, V. Misra, D. Towsley and W.-B. Gong, "A control theoretic analysis of RED" IEEE INFOCOM, 2001

[9] A. Ichikawa, "Quadratic control of evolution equations with delays in control" *SIAM J. Control and Optim.* 20:645–668, 1982.

[10] P. Kuusela, P. Lassila, J. Virtamo, "Stability of TCP-RED Congestion Control", in proceedings of ITC-17, Salvador da Bahia, Brasil, Dec. 2001, pp. 655-666.

[11] T.G. Kurtz. *Approximation of Population Processes*, volume 36 of *CBMS-NSF Regional Conf. Series in Appl. Math.* SIAM, Philadelphia, 1981.

[12] H.J. Kushner and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer-Verlag, Berlin and New York, 1992: Second edition, 2001

[13] H.J. Kushner. *Heavy Traffic Analysis of Controlled Queueing and Communication Networks.* Springer-Verlag, Berlin and New York, 2001.

[14] H.J. Kushner, D. Jarvis, and J. Yang. Controlled and optimally controlled multiplexing systems: A numerical exploration. *Queueing Systems*, 20:255–291, 1995.

[15] T.V. Lakshman and U. Madhow, "The performance of TCP/IP for networks with high bandwidth-delay products and random loss", *IEEE/ACM Transactions on Networking*, Jun 1997.

[16] S. H. Low, "A Duality Model of TCP and Queue Management Algorithms", ITC Specialist Seminar on IP Traffic Measurement, Modeling and Management, September 18-20, 2000, Monterey, CA. To appear IEEE/ACM Trans. on Networking, 2003.

[17] Laurent Massoulie, "Stability of distributed congestion control with heterogeneous feedback delays", IEEE Transactions on Automatic Control 47(2002) 895-902.

[18] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP throughput: A simple model and its empirical validation", *ACM SIGCOMM*, Sep 1998.

[19] K. K. Ramakrishnan, S. Floyd, and D. Black, “The Addition of Explicit Congestion Notification (ECN) to IP” RFC 3168, Proposed Standard, September 2001,
available at ftp://ftp.isi.edu/in-notes/rfc3168.txt

[20] S. Shakkottai and R. Srikant. ”How good are deterministic fluid models of internet congestion control.” In *Proc., IEEE INFOCOM, 2002*, New York, 2002. IEEE Press.

[21] R. Srikant. *The Mathematics of Internet Congestion Control*. Birkhäuser, Boston, 2003.

[22] R. B. Vinter and R. H. Kwong. ”The infinite time quadratic control problem for linear systems with state and control delays: An evolution equation approach.” *SIAM J. Control and Optim.* 19: 139–153, 1981.