Mean FDE Models for Internet Congestion Control Under a Many-Flows Regime

Sanjay Shakkottai and R. Srikant

Abstract— **Congestion control algorithms used in the Internet are difficult to analyze or simulate on a large scale, i.e., when there are large numbers of nodes, links and sources in a network. The reasons for this include the complexity of the actual implementation of the algorithm and the randomness introduced in the packet arrival and service processes due to many factors such as arrivals and departures of sources and uncontrollable short flows in the network. To make the analysis or simulation tractable, often deterministic fluid approximations of these algorithms are used. These approximations are in the form of either deterministic delay differential equations, or more generally, deterministic functional differential equations (FDEs). In this paper, we ignore the complexity introduced by the window-based implementation of such algorithms and focus on the randomness in the network. We justify the use of deterministic models for proportionally-fair congestion controllers under a limiting regime where the number of flows in a network is large.**

Index Terms— **Delay-differential equations, fluid model, Internet congestion control, many-flows asymptotics, proportional fairness**

I. INTRODUCTION

There has been a lot of recent work on decentralized end-to-end congestion control algorithms for the Internet. These are based on ECN marking with the goal of building a low-loss, low-queueing-delay network. The control algorithms are designed on the premise that each user has a utility function, which the user is trying to maximize, while the network is simultaneously trying to maintain some sort of fairness amongst various users. In the algorithms proposed, the network tries to achieve it's goal by *marking* packets during congestion (see [1], [2]). The notion of fairness (from the network's point of view) which has been used is weighted *proportional fairness* (see [3]). Through appropriate choice of the weights, other fairness criteria such as minimum potential delay fairness can be realized. If we interpret the utility function of the user as the users' willingness to pay for bandwidth, and suppose that the price paid by the user is proportional to the number of marks received, then, a weighted proportionally fair scheme leads to *same price per unit bandwidth* paid by any user for utilizing some resource in the network. The algorithms proposed have a *decentralized* implementation to achieve the network and user objectives simultaneously. In this paper,

we focus on the case where all users have the same utility function of $\log(x)$, leading to a proportionally fair sharing of the bandwidth. It is easy to extend the results to the case of weighted proportional fairness as we outline in Section IV-D.

There have been various delay differential equation models for Internet congestion control. The generic model of such a system consists of a flow, which sends data at rate $x(t)$, a router which signal congestion to flows by marking packets, and a receiver which detects the marks and informs the flow to increase or decrease it's transmission rate. Associated with the router is a marking function, which marks the a fraction of the flow as a function of the total arrival rate. The larger the fraction is, the more aggressively the flow "backs off". Such scenarios have been studied ([3], [4], [2], [5]), and in the absence of delays, differential equation models of such systems have been shown to converge. With delays, bounds have been derived on the behavior of the flow ([6]). However, in a realistic scenario, we have short flows which do not adapt, thus causing "noise" at the router. Further, the marking function could base its decision on more than the instantaneous arrival rate. A formal justification of how delay differential equations models correspond to "real" systems does not seem to exist in current literature.

In this paper, we study "noisy" congestion control algorithms and show that the deterministic differential equations that have been studied earlier in the literature are appropriate limits in a many flows regime.

Related work includes that of [7], where a stochastic approximations (see [8]) based model is considered. The authors study the rate process for additive increase, multiplicative decrease (AIMD) algorithms. Under the assumption of small gains (Δ and β defined in Section II-A are small), they show that an asynchronous implementation of a generic AIMD converges to an ordinary differential equation. However, the small gain assumption is not valid in practice. Further, this approach leads to a fluid limit which does not capture the oscillations due to delayed feedback. In fact, it is known that the rate control algorithms do not always converge when delays are present (see [9], [10], [6], [11], [12]). We believe that a justification of the delay-differential equation model arises from a many-flows approximation, where we scale the capacity of the system along with the number of flows. We will show that in this regime, the delay-differential equations are suitable limits of such a system. Our approach can be thought of as a functional-differential equation analog of "averaging" used for ordinary differential equations [13], [14].

Sanjay Shakkottai is with the Wireless Networking and Communications Group, Department of Electrical and Computer Engineering, The University of Texas at Austin, email: shakkott@ece.utexas.edu.

R. Srikant is with the Department of Electrical and Computer Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, email: rsrikant@uiuc.edu.

A. Organization of the Paper and Summary of the Main Results

The rest of paper is organized as follows:

- In Section II, we present two basic models of congestion control. The models differ in the way that congestion information is generated by routers in the Internet. In one model, congestion is measured in terms of the arrival rate at the router and in the other model, congestion is measured based on the queue length at the router.

- In Section III, we present a set of sufficient conditions for a sequence of stochastic delay-difference equations to converge to a deterministic functional differential equation.

- In Section IV, the result in Section III is shown to be applicable directly to the average rate per flow in a single-link network accessed by many flows with identical delays, subject to feedback provided by a rate-based marking scheme. The randomness in this scheme can be viewed as arising due to implementation dynamics that cannot be precisely modeled in the equations describing the rate-control algorithm and due to the presence of uncontrollable or unresponsive flows in the system. However, a caveat in applying this result is that we have to assume that the rate transmitted by each flow is not constrained to be non-negative. We then show that, under reasonable initial conditions for the rate of each flow and when the number of flows is large, the rate of each source remains positive, thus showing that the constraint is redundant. We also show that, after a rate allocated to each individual flow, converges to the average rate per user, thus showing fairness to each individual user. In Section IV-C, these results are extended to $[0, \infty)$. Again, the crucial part of the proof is in establishing that non-negativity constraint on each flow's rate is not violated for reasonable initial conditions. We make the assumption that the limiting delay differential equation is exponentially stable. While local stability has been established recently [11], whether the system is globally, exponentially stable is an open question.

- In prior sections, we had assumed that the router had the ability to mark a specified fraction of the arriving flow. In Section V, we relax this assumption to include random marking where the fraction of the flow that is marked is random. We also point out the difficulties in extending the results of this section to the time interval $[0, \infty)$.

- In Section VI, we prove convergence results over a finite time interval where the marking is done based on the dynamics of queue whose buffer is of the order of the round-trip feedback delay. Again, we establish the fact that non-negativity constraints of the arrival rates of each flow and the queue are not violated for suitable, yet practical, initial conditions, thus allowing us to apply the results of Section III.

- In Section VII, we show how the results of Section III can be applied to networks with adaptive marking [2] and to networks with multiple delay classes and arbitrary topologies. However, in the case of adaptive marking, we have not been able to establish that the rates of individual flows will remain non-negative for reasonable initial conditions. In the case of a network with multiple delay classes and/or a general topology, the non-negativity of each flow's rate can be established for reasonable initial conditions only if the number of delay classes is small and the number of nodes in the network is small. Obtaining more general results is an open problem.

- In Sections IV through VII, we only consider congestion control algorithms derived from a penalty function form of a resource allocation convex program introduced in [15]. In Section VIII, we discuss the extension of these results to congestion controllers derived from a dual formulation of the convex program [16], [17].

II. MODELS OF CONGESTION CONTROL

In this section, we describe two widely-used models of congestion control. We will use these models, and variants of these basic models, throughout the rest of the paper.

A. Rate-Based Marking with Arrival Noise

We consider a sequence of systems (indexed by n), where the first system is the following. A single congestion-controlled flow accesses a single link of capacity c, and the system is assumed to evolve in discrete time-steps. At each time $i = 0, 1, \ldots, T$, the user adapts its transmission rate x_i depending on the feedback it receives from the router. The router marks some amount of the flow it receives, and this amount is proportional to the user transmission rate. In practice, for a packet-based system, such marking could be implemented using ECN marks (see [1]). For a fluid model such as ours, we assume that some volume of the fluid is marked (see [3], [4], [2]). The fraction of fluid marked is determined by means of a marking function, about which we will discuss more later. We assume that there is a round trip delay between the flow and router of $d \in \mathcal{Z}$. Thus, the rate at time $i + 1$ depends on the amount marked at the router half a round trip back, which in-turn, depends on the user transmission rate a further half round-trip time back. Thus, we can describe the evolution of the user transmission rate by

$$x_{i+1} = (x_i + \Delta - \beta x_{i-d} p(x_{i-d} + \hat{e}_{i-d}))^+$$

where Δ, β are positive constants which determine the rate at which flow increases or decreases it's transmission rate; $p(.)$ is the marking function, and $\hat{e}_i$ is a "noise" process. We can interpret the above equation as though two flows are accessing the router: the first is the congestion-controlled flow, whose data transmission rate is represented by $\{x_i\}$ and the other is an uncontrolled flow, possibly generated by some other short-duration flows, popularly known as web-mice, passing through the link, which is represented by the sequence $\{\hat{e}_i\}$. The sequence $\{\hat{e}_i\}$ is assumed to be a stochastic process, with $E(\hat{e}_1) = a > 0$. Let $\tilde{e}_i = \hat{e}_i - a$. We assume that the "noise" process $\tilde{e}_i$ is a bounded, stationary-ergodic zero-mean stochastic process for $i \geq 0$; and for $i \leq 0$, $\tilde{e}_i = \tilde{e}_0$. The mean initial conditions (i.e., for $t \leq 0$) is given by sampling $\theta(t), -T \leq t \leq 0$, where $\theta(t)$ is a non-negative, bounded,

Lipschitz continuous trajectory. Thus, the initial conditions for above system are given $x_{-i} = \theta(-i) + \tilde{e}_0$.

Finally, we comment on the marking function itself. This function is based on the total data rate accessing the router and determines the fraction of flow to be marked, and satisfies satisfies the following criteria:

Assumption 2.1: We assume that $p(.)$ satisfies

 (i) $0 \le p(x) \le 1$
 (ii) $p(x) = 0$ for $x < 0$. Further, there exists a δ satisfying $0 \le \delta \ll \Delta$ such that $|\beta p(x)| < \delta/a$ for $x < a$.
 (iii) $p(x)$ is an increasing function.
 (iv) $p(x)$ is Lipschitz continuous.

∎

The first property is obvious, as the marking function represents the fraction of flow marked. To understand the second property, we first note that a, the mean arrival rate of the uncontrolled flows, is typically less than 25% of the link capacity. Thus, condition (ii) expresses the intuitive reasoning that, if the total arrival rate (i.e., sum of arrival rates of uncontrolled and controlled rates) at a link is less than 25%, very little congestion indication should be provided. The third property is again clear: the larger the arrival rate is, the greater is the fraction marked. Finally, the last condition is a technical condition, which says that the function is "smooth". As an example, a possible rate-based marking function is of the form

$$p(x) = \frac{(x - \tilde{c})^+}{x}$$

In a deterministic fluid model, this has the interpretation of the fraction of fluid lost when the arrival rate exceeds a certain level, called the "virtual" capacity, $N\tilde{c}$ [2]. In the absence of delays and noise (i.e., there is no fluctuation in the uncontrolled flows), the flow rate x_i is assumed to converge to $(c - a)$ for some $c > a$. Thus, $p(\cdot)$ satisfies

$$\Delta = \beta(c - a)p(c)$$

Our goal in this study is to determine when the above stochastic system can be approximated by the following deterministic system:

$$\dot{x} = \Delta - \beta x(t - d)p(x(t - d) + a) \tag{1}$$

with initial conditions given by $\theta(t), -T \le t \le 0$. The congestion control algorithm corresponds to a resource allocation problem where all flows have $\log(.)$ utility functions [15]. As an aside, we note that the local stability of this equation (i.e., whether $x(t) \to c - a$ for a linearized system) has been studied in [9], [10], while global boundedness has been studied in [6].

We have so far described the system model when a single flow accesses the link. We now describe how the model scales in n. First, there are n flows, and for every time step in the first model, we assume that there are n time-steps in the nth model. This represents the fact that we need to increase the time resolution to study n processes. In practice, we can view each time slot at a measurement interval over which rates are measured in the system and control actions by the routers and flows are updated. Typically, this measurement interval is measured in terms of the number of packets that can be processed by a typical router. For example, the time-step could be "100 packets long." By scaling both the time-step and the capacity, we maintain a constant time-step, as measured in packets. The flows are now represented by $\{y_i^k\}, k = 1, 2, \ldots, n$ where the subscript $i \in \{0, 1, \ldots, nT\}$ represents the time-index, and the superscript k represents the flow index. As the delay seen by a user does not change with n but the time resolution increases in n, it follows that the delay (measured in index steps) scales in n. Thus, each flow y_i^k evolves according to the following stochastic delay-difference equation

$$y_{i+1}^k = \left[\begin{array}{c} y_i^k + \frac{\Delta_k^n}{n} - \frac{\beta}{n} y_{i-nd}^k \\ p\left(\frac{1}{n} \sum_{j=1}^n (y_{i-nd}^j + \tilde{e}_{i-nd}^j + a) \right) \end{array} \right]^+ \tag{2}$$

with initial conditions given by $\theta(i) + \tilde{e}_0^k$ for $i \le 0$. Thus, the initial conditions of each user could be possibly different. As in the single flow case, the above equations represent n flows, each which has an additive increase factor of and the backoff for each flow is proportional to the delayed transmission rate, which the fraction marked being a function of the *total arrival rate* to the router. We remark that the above system can be interpreted as a decentralized means of achieving a proportionally fair allocation (see [3]) of bandwidth among the n users.

We note three features in the above equation. First, the gain constants Δ and β are scaled by n. This is because the time step in now $\frac{1}{n}$ of the first system. Thus, the gains are also scaled to maintain the same gain for each flow over the original time-step.

Second, there are now n uncontrolled flows (i.e., $\{e_i^k\}, k = 1, 2, \ldots, n, i \ge 0$) accessing the system. We assume that each uncontrolled flow has a bounded rate, i.e., there is a $K > 0$ such that $|\tilde{e}_i^k| < K$. We assume that these flows are i.i.d. (across flows, although these could be correlated in time), and are stationary and ergodic. For $i < 0$, we assume a random perturbation about a nominal initial condition $\theta(i)$, i.e., the initial condition is $\theta(i) + \tilde{e}_i^k$ with $\tilde{e}_i^k = \tilde{e}_0^k$. Then, we have for any $\epsilon > 0$,

$$\Pr\left(\sup_{-nT \le i \le nT} \left| \frac{1}{n} \sum_{j=1}^n \tilde{e}_i^j \right| > \epsilon \right)$$

$$\le (nT + 1)\Pr\left(\left| \frac{1}{n} \sum_{j=1}^n \tilde{e}_0^j \right| > \epsilon \right)$$

$$\le (nT + 1)e^{-na}$$

for some $a = a(\epsilon) > 0$. The first step follows from the union bound, and the fact that the flows are stationary. The second step (i.e., existence of $a > 0$) follows from Chernoff's bound (it trivially follows that the MGF exists, as the process is bounded). Thus, it follows that

$$\Pr\left(\sup_{-nT \le i \le nT} \left| \frac{1}{n} \sum_{j=1}^n \tilde{e}_i^j \right| > \epsilon \right) \stackrel{n \to \infty}{\longrightarrow} 0$$

exponentially fast. Now, applying Borel-Cantelli Lemma, it follows that

$$\sup_{-nT \le i \le nT} \left| \frac{1}{n} \sum_{j=1}^{n} \tilde{e}_i^j \right| \quad \to \quad 0 \quad \text{a.s.} \tag{3}$$

We remark that the above model allows each "noise" process to be long-range dependent. We only need the flows to be bounded and iid.

Third, the marking function is seen to scale its argument by n (the marking function acts on the *average* arrival rate as opposed to the *total* arrival rate), i.e., if the marking function for the nth system were to be represented by $p^n(x)$, then, we would have

$$p^n(nx) \quad = \quad p(x)$$

This is done so that for each n, if the centered error processes $\{\tilde{e}_i^k\}, k = 1, \ldots, n$ were identically zero, the delay-difference equations (if they converge) would have a steady state value of $\lim_{k \to \infty} y_i^k = c - a$. We remark that *adaptive* marking functions have been proposed (see [5]) which automatically scale in n as described above, without explicitly having knowledge of n.

Now, let x_i^n represent the average rate at time i, and e_i^n represent the average (centered) noise at time i, i.e., i.e.,

$$x_i^n \quad = \quad \frac{1}{n} \sum_{k=1}^{n} y_i^k$$

$$e_i^n \quad = \quad \frac{1}{n} \sum_{k=1}^{n} \tilde{e}_i^k$$

Then, (by adding the various equations for y_i^k and ignoring non-negativity constraints on y_i^k, which will be justified later), we have that x^n satisfies the following stochastic delay-difference equation:

$$x_{i+1}^n \quad = \quad x_i^n + \frac{\Delta}{n} - \frac{\beta}{n} x_{i-nd}^n \, p\left(x_{i-nd}^n + a + e_{i-nd}^n\right)$$

Now, we embed the above equation in "continuous-time", i.e., we study the above process over an interval of time $[0, T]$ (without loss of generality, assume that $T = \zeta d$, for some $\zeta > 1$). For $nt \in \mathcal{Z}$, we let

$$x^n(t) = x_{nt}^n, \qquad e^n(t) = e_{nt}^n,$$

and use a straight line approximation to interpolate between the times $t = \frac{i}{n}$. Thus, we see that the above equation can be represented by the equation

$$\dot{x}^n(t) \quad = \quad \Delta^n - \beta x^n(\frac{\lfloor n(t-d) \rfloor}{n}) \tag{4}$$
$$p\left(x^n(\frac{\lfloor n(t-d) \rfloor}{n}) + a + e^n(\frac{\lfloor n(t-d) \rfloor}{n})\right)$$

Further, assume that the initial condition is given by

$$x^n(t) \quad = \quad \theta(t) + e^n(0) \quad -T \le t \le 0, \ nt \in \mathcal{Z}$$

and a straight line interpolation is used for $nt \notin \mathcal{Z}$. This means that each flow $\{y^k\}, k = 1, 2, \ldots, n$ could have a different initial condition, but the nominal initial condition

is given by $\theta(t)$. Also, note that from (3), it follows that $e^n(t) \to 0$ uniformly in $[-T, T]$. We note that the above differential equation is to be interpreted as a shorthand notation to represent the process $x^n(t)$ given by the unique trajectory solving the integral equation

$$x^n(t) \quad = \quad x^n(0) + \int_{s=0}^{t} \Delta - \beta x^n(\frac{\lfloor n(s-d) \rfloor}{n})$$
$$p\left(x^n(\frac{\lfloor n(s-d) \rfloor}{n}) + a + e^n(\frac{\lfloor n(s-d) \rfloor}{n})\right) ds$$

In this paper, unless otherwise stated, any differential equation is to be interpreted as a representation of the corresponding integral equation.

Our objective is to show that the trajectory generated by (4) and that in (1) are "close" in some suitable sense.

B. Queue-Based Marking

We will now study systems where the marking function could be based on queue lengths. Examples of a such system are RED [18] which is a popular queue-based marking scheme, REM, a resource-pricing-based algorithm which can also be interpreted as a queue-based marking scheme [19] and a scheme closely-related to REM, based on feedback control ideas [20].

We assume n flows access the router. However, unlike before, where the marking depended on the instantaneous rates, here, we consider marking functions based on queue-lengths. As before let $x^n(t)$ be the aggregate average data rate at the router, with capacity nC. We have the uncontrolled "noise" process given by $a + e^n(t)$, where $a > 0$, $e^n(t)$ is a zero-mean process satisfying suitable assumptions as before. We let $c = C - a$. Then, the evolution of the aggregate rate can be described by the following SFDE:

$$\dot{x}^n(t) \quad = \quad \Delta - \beta x^n(\frac{\lfloor n(t-d) \rfloor}{n})$$
$$\left(p\left(q^n(\frac{\lfloor n(t-d) \rfloor}{n})\right) + \xi^n(\frac{\lfloor n(t-d) \rfloor}{n})\right) \tag{5}$$

where $q^n(t)$ is the scaled queue length. The marking function acts on the scaled (by $\frac{1}{n}$) queue length, whose evolution is given by

$$\dot{q}^n(t) \quad = \quad \begin{cases} x^n(t) + e^n(t) - c & \text{if } q^n(t) > 0 \\ (x^n(t) + e^n(t) - c)^+ & \text{if } q^n(t) = 0 \end{cases} \tag{6}$$

We have not explicitly shown the non-negativity constraints on $x(t)$. We will later show that for reasonable initial conditions and sufficiently large n, the trajectories will remain positive for all time. The marking function $p(\cdot)$ is assumed to satisfy conditions (i), (iii) and (iv) of Assumption 2.1. We assume that the initial for $x^n(t)$ is given by

$$x^n(t) \quad = \quad \theta(t) + e^n(0) \quad -T \le t \le 0, \ nt \in \mathcal{Z}$$

and a straight line interpolation is used for $nt \notin \mathcal{Z}$. The queue length is initially assumed to be zero over $[-T, 0]$, i.e.,

$$q^n(t) \quad = \quad 0 \quad -T \le t \le 0$$

Now, consider a deterministic system consisting of a single flow $x(t)$ driving a queue with capacity c, and whose evolution is given by

$$\dot{x}(t) = \Delta - \beta x(t-d) p\left(q(t-d)\right) \qquad (7)$$

$$\dot{q}(t) = \begin{cases} x(t) - c & \text{if } q(t) > 0 \\ (x(t) - c)^+ & \text{if } q(t) = 0 \end{cases} \qquad (8)$$

and $p(.)$ is a *queue-based* marking function. The equilibrium point is given by $q(t) = q_0$, $x(t) = c$, and they are related as $\Delta = \beta c p(q_0)$. The initial conditions are given by $\theta(t)$ for $t \in [0, -d]$ which is Lipschitz continuous and $0 \leq \theta(t) < c$, and $q(t) = 0$ for $t \leq 0$. Our objective is to show that the trajectories described by (5) and (6) are "close" to those described by (7)-(8) for large n.

In the next section, we will prove a general convergence result for functional differential equations. We will then show that the models we have described fit into this framework, and the desired results can be proved.

III. CONVERGENCE OF FUNCTIONAL DIFFERENTIAL EQUATIONS

Let $\mathcal{R}^2$ be endowed with the $\mathcal{L}_1$ norm, and $C([0,T], \mathcal{R}^2)$ be the space of continuous, $\mathcal{R}^2$ valued functions on $[0,T]$ with the supremum norm. We denote any element of $C([0,T], \mathcal{R}^2)$ by the tuple $(\bar{\phi}, \bar{\psi})$, and $\|(\bar{\phi}, \bar{\psi})\| = \sup_{t \in [0,T]}(|\phi(t)| + |\psi(t)|)$.

Let us consider a sequence of functionals b_n : $C([0,T], \mathcal{R}^2) \mapsto \mathcal{R}$, such that $\{b_n\}$ are Lipschitz continuous and bounded with parameters L and M respectively, i.e.,

Assumption 3.1:

$$|b_n(\bar{\phi}, \bar{\psi})| \leq M$$
$$|b_n(\bar{\phi}_1, \bar{\psi}_1) - b_n(\bar{\phi}_2, \bar{\psi}_2)| \leq L\|(\bar{\phi}_1, \bar{\psi}_1) - (\bar{\phi}_2, \bar{\psi}_2)\|$$
$$= L \sup_{t \in [0,T]} \begin{array}{l} |\phi_1(t) - \phi_2(t)| \\ + |\psi_1(t) - \psi_2(t)| \end{array}$$

∎

Let $C([0,T], \mathcal{R})$ be the space of continuous, real-valued functions on $[0,T]$, and endowed with the sup topology. Next, we define $\bar{x}_t$ to be the trajectory of $x(\cdot)$, in the time interval $[t-T, t]$. Note that a time-shifted version of $\bar{x}_t$ is an element of $C([0,T], \mathcal{R})$. However, to avoid notational complexity, we will avoid defining a shift operator and simply abuse the notation and refer to $\bar{x}_t$ as being an element of $C([0,T], \mathcal{R})$. Fix $a, b > 0$ and let $\widetilde{C}[0,T]_{a,b,c} = \{\bar{x}_T \in C([0,T], \mathcal{R}) : \bar{x}_T$ is Lipschitz continuous with parameter a, $x(t) \geq -c$ and $x(t) \leq b\}$. As an aside, it can be shown that $\widetilde{C}[0,T]_{a,b,c}$ is a compact subset of $C([0,T], \mathcal{R})$.

Now, we consider the following stochastic functional differential equations (SFDEs). Let $x^n(t)$ be the unique, continuous solution of the following FDE. By the conditions imposed on b_n, there exists such a solution [21] (the Caratheodory conditions). For $t \in [0,T]$, consider

$$\dot{x}^n(t) = b_n\left(\bar{x}^n_{\frac{\lfloor nt \rfloor}{n}}, \bar{0}\right) \qquad (9)$$

where $\bar{x}^n_t \triangleq [x^n(s), s \in [t-T, t]] \in C([0,T], \mathcal{R})$, and $\bar{0}$ is a process identically equal to zero in $[-T, T]$. In (9), we need

to truncate this zero trajectory to $[0,T]$ and use this truncated trajectory. To avoid unnecessary notation, we slightly abuse notation and use the same symbol. The initial condition is given by

$$x^n(t) = \theta(t) \quad -T \leq t \leq 0, \ nt \in \mathcal{Z}$$

and a straight line interpolation is used for $nt \notin \mathcal{Z}$. The above differential equation is to be interpreted as a shorthand notation to represent the process $x^n(t)$ given by the unique trajectory solving the integral equation

$$x^n(t) = \int_{s=0}^{t} b_n\left(\bar{x}^n_{\frac{\lfloor ns \rfloor}{n}}, \bar{0}\right) ds + x^n(0)$$

In this paper, unless otherwise stated, any differential equation is to be interpreted as a representation of the corresponding integral equation.

Assume $|\theta(.)| < A$, and that $\theta(t)$ is Lipschitz continuous with parameter M. Then, it follows $x(t)$ is Lipschitz continuous with Lipschitz constant M and bounded (with bounds $\pm(A + MT)$).

Next, consider for $t \in [0,T]$, the SFDE

$$\dot{\mathbf{x}}^n(t) = b_n\left(\bar{\mathbf{x}}^n_{\frac{\lfloor nt \rfloor}{n}}, \bar{\mathbf{e}}^n_{\frac{\lfloor nt \rfloor}{n}}\right)$$

here $\bar{\mathbf{x}}^n_t \in C([0,T], \mathcal{R})$, is given by $\bar{\mathbf{x}}^n_t = [\mathbf{x}^n(s), s \in [t-T, t]]$, $\bar{\mathbf{e}}^n_t \in C([0,T], \mathcal{R})$ is given by $\bar{\mathbf{e}}^n_t = [\mathbf{e}^n(s), s \in [t-T, t]]$, and the initial condition is given by

$$x^n(t) = \theta(t) + \mathbf{e}^n(t) \quad -T \leq t \leq 0, \ nt \in \mathcal{Z}$$

and a straight line interpolation is used for $nt \notin \mathcal{Z}$.

$[\mathbf{e}^n(t), -T \leq t \leq T]$ is a zero mean "error" process, satisfying the following condition.

Assumption 3.2: Assume that $\exists K > 0$ such that $|e^n(t)| < K$ uniformly in n, and satisfying the following condition:

$$\sup_{t \in [-T,T]} |\mathbf{e}^n(t)| \overset{n \to \infty}{\longrightarrow} 0 \quad \text{a.s.}$$

∎

Now, let $b : C([0,T], \mathcal{R}^2) \mapsto \mathcal{R}$ satisfy the following condition.

Assumption 3.3: Assume that $b(.,.)$ is a Lipschitz continuous and bounded function with parameters L, M respectively (without loss of generality, assume $L \geq 1$), satisfying

$$\sup_{\bar{x}_t \in \widetilde{C}[0,T]_{M,A+MT,-(A+MT)}} |b_n(\bar{x}_t, \bar{0}) - b(\bar{x}_t, \bar{0})| \overset{n \to \infty}{\longrightarrow} 0$$

∎

Finally, for $t \in [0,T]$, consider the FDE

$$\dot{x}(t) = b(\bar{x}_t, \bar{0})$$

where $\bar{x}_t \triangleq [x(s), s \in [t-T, t]] \in C([0,T], \mathcal{R})$, and the initial condition is given by

$$x(t) = \theta(t) \quad -T \leq t \leq 0$$

For n large, it seems reasonable to believe that the trajectories of $x(t)$, $x^n(t)$ and $\mathbf{x}^n(t)$ are "close". We will show that this is indeed true. Formally, we will show that

$$\sup_{t \in [-T,T]} |x(t) - \mathbf{x}^n(t)| \rightarrow 0 \quad \text{a.s.}$$

We prove the desired result in two steps. We first show that $||x - x^n||_{C([0,T],\mathcal{R})} \rightarrow 0$. Next, we show that $||x^n - \mathbf{x}^n||_{C([0,T],\mathcal{R})} \rightarrow 0$ almost surely, and the required result will follow.

Lemma 3.1: For the framework we have discussed, given any $\epsilon > 0$, $\exists N > 0$ such that $\forall n \geq N$,

$$\sup_{t \in [0,T]} |x(t) - x^n(t)| \leq LT \left(\frac{2M}{n} + \epsilon \right) e^{LT}$$

Thus,

$$\sup_{t \in [0,T]} |x(t) - x^n(t)| \stackrel{n \to \infty}{\longrightarrow} 0$$

Proof. Fix any $n > 1$. Let

$$
\begin{aligned}
y(t) &= b(\bar{x}_t, \mathbf{0}) \\
y^n(t) &= b(\bar{x}_t^n, \mathbf{0}) \\
y_n(t) &= b_n(\bar{x}_t^n, \mathbf{0})
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
\dot{x}^n(t) &= y_n \left(\frac{\lfloor nt \rfloor}{n} \right) \\
\dot{x}(t) &= y(t)
\end{aligned}
$$

Fix any $t > 0$. We have

$$|x(t) - x^n(t)| = \left| \begin{array}{l} \int_0^t (y(s) - y^n \left(\frac{\lfloor ns \rfloor}{n} \right)) ds \\ + x(0) - x^n(0) \end{array} \right|$$

$$\leq \int_0^t |y(s) - y_n \left(\frac{\lfloor ns \rfloor}{n} \right)| ds + |x(0) - x^n(0)|$$

$$\leq \int_0^t |y(s) - y_n(s)| ds$$
$$+ \int_0^t |y_n(s) - y_n \left(\frac{\lfloor ns \rfloor}{n} \right)| ds + |x(0) - x^n(0)|$$

$$\leq \int_0^t |y(s) - y^n(s)| ds$$
$$+ \int_0^t |y^n(s) - y_n(s)| ds$$
$$+ \int_0^t |y_n(s) - y_n \left(\frac{\lfloor ns \rfloor}{n} \right)| ds$$
$$+ \sup_{z \in [-T,0]} |x(z) - x^n(z)| \tag{10}$$

Let us define

$$
\begin{aligned}
m^n(t) &= ||\bar{x}_t - \bar{x}_t^n||_{C[0,T]} \\
&= \sup_{z \in [t-T,t]} |x(z) - x^n(z)|
\end{aligned}
$$

Consider a time $u \in [-T, 0]$. If $nu \in \mathcal{Z}$, we have $x(u) = x^n(u)$. For $nu \notin \mathcal{Z}$, as $x(u)$ and $x^n(u), n = 1, 2, \ldots$ are uniformly Lipschitz continuous (i.e., the same M works for all the trajectories), we have

$$
\begin{aligned}
m^n(0) &= \sup_{u \in [-T,0]} |x(u) - x^n(u)| \\
&\leq \frac{M}{n}
\end{aligned}
$$

We consider different terms in the (10). First, as $b(.,.)$ is Lipschitz with constant L, and $s \in [0, t]$,

$$
\begin{aligned}
|y(s) - y^n(s)| &= |b(\bar{x}_s, \mathbf{0}) - b(\bar{x}_s^n, \mathbf{0})| \\
&\leq L ||(\bar{x}_s, \mathbf{0}) - (\bar{x}_s^n, \mathbf{0})|| \\
&\leq L m^n(s)
\end{aligned}
$$

Next, as $x^n(t)$ and $x(t)$ are Lipschitz with constant M and bounded with $\pm(A + MT)$, for all n large enough, by the uniform convergence of $b_n(., \mathbf{0})$, we have

$$
\begin{aligned}
|y^n(s) - y_n(s)| &= |b(\bar{x}_s^n, \mathbf{0}) - b_n(\bar{x}_s^n, \mathbf{0})| \\
&\leq \frac{L\epsilon}{2}
\end{aligned}
$$

Finally, from Lipschitz continuity of $b(.,.)$ and $x^n(t)$,

$$\left| y^n(s) - y^n \left(\frac{\lfloor ns \rfloor}{n} \right) \right| = |b(\bar{x}_s^n, \mathbf{0}) - b(\bar{x}_{\frac{\lfloor ns \rfloor}{n}}^n, \mathbf{0})|$$

$$\leq L \sup_{z \in [s-T,s]} |x^n(z) - x^n \left(\frac{\lfloor nz \rfloor}{n} \right)|$$
$$\leq \frac{LM}{n}$$

Thus, we have

$$|x(t) - x^n(t)| \leq L \int_0^t m^n(s) ds$$
$$+ \left(\frac{LM}{n} + L\epsilon \right) t + m^n(0)$$

Next, for any $0 \leq \tau \leq t$, as $m^n(s) \geq 0$, we have

$$|x(\tau) - x^n(\tau)| \leq L \int_0^\tau m^n(s) ds$$
$$+ \left(\frac{LM}{n} + L\epsilon \right) \tau + m^n(0)$$
$$\leq L \int_0^t m^n(s) ds + \frac{LM}{n} t + m^n(0)$$

For any $\tau \leq 0$, we have

$$
\begin{aligned}
|x(\tau) - x^n(\tau)| &\leq m^n(0) \\
&\leq \frac{M}{n}
\end{aligned}
$$

Thus, it follows that

$$m^n(t) \leq L \int_0^t m^n(s) ds + \left(\frac{LM}{n} + L\epsilon \right) t + m^n(0)$$
$$\leq L \int_0^t m^n(s) ds + \left(\frac{LM}{n} + L\epsilon \right) t + \frac{M}{n}$$
$$\leq L \int_0^t m^n(s) ds + \left(\frac{LM}{n} + L\epsilon \right) T + \frac{M}{n}$$
$$\leq L \int_0^t m^n(s) ds + LT \left(\frac{2M}{n} + \epsilon \right)$$

where we used the fact that $LT \geq 1$. Using Gronwall's inequality, it now follows that

$$m^n(t) \leq LT\left(\frac{2M}{n} + \epsilon\right)e^{Lt}$$

Thus,

$$\limsup_{n \to \infty} \sup_{t \in [0,T]} |x(t) - x^n(t)| \leq LT\epsilon e^{LT}$$

As ϵ is arbitrary, we have the desired result. ∎

Lemma 3.2: For the framework we have discussed, we have

$$\sup_{t \in [0,T]} |\mathbf{x}^n(t) - x^n(t)| \leq 2LT\mathbf{E}^n(T)e^{LT}$$

where,

$$\mathbf{E}^n(T) = \sup_{z \in [-T,T]} |\mathbf{e}^n(z)|$$

Thus, we have

$$\sup_{t \in [0,T]} |\mathbf{x}^n(t) - x^n(t)| \overset{n \to \infty}{\longrightarrow} 0 \quad \text{a.s.}$$

Proof. The proof of this result follows a similar approach to Lemma 3.1. As before, let us fix any $n > 1$, and let

$$\begin{aligned} y^n(t) &= b_n(\bar{x}_t^n, \bar{\mathbf{0}}) \\ \mathbf{y}^n(t) &= b_n(\bar{\mathbf{x}}_t^n, \bar{\mathbf{e}}_t^n) \end{aligned}$$

Thus, we have

$$\begin{aligned} \dot{x}^n(t) &= y^n\left(\frac{\lfloor nt \rfloor}{n}\right) \\ \dot{\mathbf{x}}^n(t) &= \mathbf{y}^n\left(\frac{\lfloor nt \rfloor}{n}\right) \end{aligned}$$

Let[1]

$$\begin{aligned} \mathbf{E}^n(t) &= \sup_{z \in [t-2T,t]} |\mathbf{e}^n(z)| \\ \mathbf{m}^n(t) &= \sup_{z \in [t-T,t]} |\mathbf{x}^n(z) - x^n(z)| \end{aligned}$$

For $t \in [-T, 0]$,

$$|\mathbf{x}^n(t) - x^n(t)| \leq \mathbf{E}^n(0)$$

For $t \geq 0$, we have

$$|\mathbf{x}^n(t) - x^n(t)| =$$

$$= \left| \begin{array}{l} \int_0^t (\mathbf{y}^n\left(\frac{\lfloor ns \rfloor}{n}\right) - y^n\left(\frac{\lfloor ns \rfloor}{n}\right))ds \\ + \mathbf{x}^n(0) - x^n(0) \end{array} \right|$$

$$\leq \begin{array}{l} \int_0^t |(\mathbf{y}^n\left(\frac{\lfloor ns \rfloor}{n}\right) - y^n\left(\frac{\lfloor ns \rfloor}{n}\right))|ds \\ + \underbrace{|\mathbf{x}^n(0) - x^n(0)|}_{\leq \mathbf{E}^n(0)} \end{array}$$

As $b_n(.,.)$ is Lipschitz with constant L, and $s \in [0,t]$,

$$\left| \mathbf{y}^n\left(\frac{\lfloor ns \rfloor}{n}\right) - y^n\left(\frac{\lfloor ns \rfloor}{n}\right) \right|$$

[1]All processes are defined to be zero for $t \leq T$.

$$\leq L||(\bar{\mathbf{x}}_{\frac{\lfloor ns \rfloor}{n}}^n, \bar{\mathbf{e}}_{\frac{\lfloor ns \rfloor}{n}}^n) - (\bar{x}_{\frac{\lfloor ns \rfloor}{n}}^n, \bar{\mathbf{0}})||$$

$$\leq L \sup_{z \in [s-T,s]} \left(\begin{array}{l} |\mathbf{x}\left(\frac{\lfloor nz \rfloor}{n}\right) - x^n\left(\frac{\lfloor nz \rfloor}{n}\right)| \\ + |\mathbf{e}^n\left(\frac{\lfloor nz \rfloor}{n}\right)| \end{array} \right)$$

$$\leq L \sup_{z \in [s-T,s]} |\mathbf{x}\left(\frac{\lfloor nz \rfloor}{n}\right) - x^n\left(\frac{\lfloor nz \rfloor}{n}\right)|$$

$$+ L \sup_{z \in [s-T,s]} |\mathbf{e}^n\left(\frac{\lfloor nz \rfloor}{n}\right)|$$

$$\leq L\mathbf{m}^n(s) + L\mathbf{E}^n(s)$$

Thus, using the fact that for any $s \in [-T, T]$,

$$\mathbf{E}^n(s) \leq \mathbf{E}^n(T)$$

we have,

$$|\mathbf{x}^n(t) - x^n(t)| \leq L \int_0^t \mathbf{m}(s)ds + Lt\mathbf{E}^n(T) + \mathbf{E}^n(0)$$

$$\leq L \int_0^t \mathbf{m}(s)ds + (Lt+1)\mathbf{E}^n(T)$$

Next, for any $0 \leq \tau \leq t$, as $\mathbf{m}^n(s) \geq 0$, we have

$$|\mathbf{x}^n(\tau) - x^n(\tau)| \leq L \int_0^\tau m^n(s)ds + (L\tau+1)\mathbf{E}^n(T)$$

$$\leq L \int_0^t m^n(s)ds + (Lt+1)\mathbf{E}^n(T)$$

For any $\tau \leq 0$, we have

$$|\mathbf{x}^n(\tau) - x^n(\tau)| \leq \mathbf{E}^n(T)$$

Thus, it follows that for any $t \geq 0$,

$$\mathbf{m}^n(t) \leq L \int_0^t \mathbf{m}^n(s)ds + (Lt+1)\mathbf{E}^n(T)$$

From Gronwall's inequality, it follows that

$$\mathbf{m}^n(t) \leq 2LT\mathbf{E}^n(T)e^{Lt}$$

As $n \to \infty$, we have $\mathbf{E}^n(T) \to 0$ a.s., and the result follows. ∎

Theorem 3.1:

$$\sup_{t \in [0,T]} |x(t) - \mathbf{x}^n(t)| \overset{n \to \infty}{\longrightarrow} 0 \quad \text{a.s.}$$

Proof. We have from Lemmas 3.1 and 3.2,

$$\sup_{t \in [0,T]} |x(t) - \mathbf{x}^n(t)|$$

$$\leq \sup_{t \in [0,T]} |x(t) - x^n(t)| + \sup_{t \in [0,T]} |\mathbf{x}^n(t) - x^n(t)|$$

$$\overset{n \to \infty}{\longrightarrow} 0$$

Thus, we have the desired result. ∎

Remark 3.1: We note that this result can be easily extended to the case where $\mathbf{x}^n(t)$ and $\mathbf{e}^n(t)$ are finite-dimensional (say J) vectors. In such a case, we assume that b_n :

$C([0,T], \mathcal{R}^{2J}) \mapsto \mathcal{R}^J$ satisfies the following: for each $1 \leq i \leq J$ and $n \geq 1$, $b_{n,i}$, where $b_{n,i}$ is the ith component of b_n, is Lipschitz continuous and bounded with parameters L and M respectively. In other words,

$$
\begin{aligned}
|b_{n,i}(\bar{\phi}, \bar{\psi})| &\leq M \\
|b_{n,i}(\bar{\phi}_1, \bar{\psi}_1) - b_{n,i}(\bar{\phi}_2, \bar{\psi}_2)| &\leq L||(\bar{\phi}_1, \bar{\psi}_1) - (\bar{\phi}_2, \bar{\psi}_2)|| \\
&= L \sup_{t \in [0,T]} \max_{1 \leq j \leq J} \left(\begin{array}{c} |\phi_{1,j}(t) - \phi_{2,j}(t)| \\ +|\psi_{1,j}(t) - \psi_{2,j}(t)| \end{array} \right).
\end{aligned}
$$

Further, we assume that b_n converges to a functional $b(.,.)$ uniformly (analogous to Assumption 3.3), and that each component of the vector noise process goes to zero uniformly over $[0,T]$ as $n \to \infty$. Under these assumptions, we can show that as $n \to \infty$, almost surely, we have

$$
\sup_{t \in [0,T]} \max_{1 \leq i \leq J} |\mathbf{x}_i^n(t) - x_i(t)| \to 0,
$$

where $\mathbf{x}_i^n(t)$ and $x_i(t)$ are the ith component of $\mathbf{x}^n(t)$ and $x^n(t)$ respectively. ∎

We note that the above result is the functional differential equation analog of the result in [13] (see also [22, pages 219-222]). However, due to the Lipschitz and boundedness assumption on b_n, we are able to show almost sure convergence. In this sense, ours is a sample-path proof that extends the deterministic averaging results for differential equations in [14] to functional differential equations. Specifically, by defining the function $m^n(\cdot)$, we were able to apply the Gronwall inequality to functional differential equations.

A. Extensions to Infinite Time

In the previous section, we showed that under suitable conditions, over any bounded interval of time, the SFDE converges to the mean FDE. However, the rate of convergence depends on the length of the time interval. Thus, given any $T > 0, \epsilon > 0$, we could find a n large enough such that for all $t \in [0,T]$, $|x(t) - \mathbf{x}^n(t)| < \epsilon$. However, for a fixed $\epsilon > 0$, as T increases, n will correspondingly increase. This is not very surprising, as we made no assumptions on the stability of the mean FDE. In this section, under stricter conditions, and assuming that the mean FDE is stable, we will derive a uniform bound over arbitrarily large intervals of time.

We consider a sequence of systems as in the previous section. However, unlike the previous section, where we considered all systems over the time interval $[0,T]$, here, the nth system is defined over the interval of time $[0, nT]$. For $t \in [0, nT]$, we consider the the SFDE

$$
\dot{\mathbf{x}}^n(t) = b_n \left(\bar{\mathbf{x}}_{\frac{\lfloor nt \rfloor}{n}}^n, \bar{\mathbf{e}}_{\frac{\lfloor nt \rfloor}{n}}^n \right) \tag{11}
$$

here $\bar{\mathbf{x}}_t^n \in C([0,T], \mathcal{R})$, is given by $\bar{\mathbf{x}}_t^n = [\mathbf{x}^n(s), s \in [t - T, t]]$, $\bar{\mathbf{e}}_t^n \in C([0,T], \mathcal{R})$ is given by $\bar{\mathbf{e}}_t^n = [\mathbf{e}^n(s), s \in [t - T, t]]$, and the initial condition is given by

$$
x^n(t) = \theta(t) + \mathbf{e}^n(t) \quad -T \leq t \leq 0, \ nt \in \mathcal{Z}
$$

and a straight line interpolation is used for $nt \notin \mathcal{Z}$.

For $t \in [0, \infty)$, consider the FDE

$$
\dot{x}(t) = b(\bar{x}_t, \bar{\mathbf{0}}) \tag{12}
$$

where $\bar{x}_t \in C([0,T], \mathcal{R})$, i.e., $\bar{x}_t = [x(s), s \in [t - T, t]]$, and the initial condition is given by

$$
x(t) = \theta(t) \quad -T \leq t \leq 0
$$

To the framework described in the previous section, we make the following additional assumptions. The first assumption deals with the stability of the mean FDE. Suppose that (12) converges to 0 exponentially fast, i.e.,

Assumption 3.4: The mean FDE described by (12) is *exponentially stable*, i.e., there exists some $\kappa > 0$, $\alpha > 0$ such that for all $t \geq 0$,

$$
|x(t)| \leq \kappa ||\bar{x}_0|| e^{-\alpha t}
$$

∎

Secondly, we strengthen the noise process assumption.

Assumption 3.5: Let $[\mathbf{e}^n(t), -T \leq t \leq nT]$ be a zero mean "error" process. Assume that $\exists K > 0$ such that $|e^n(t)| < K$ uniformly in n, and satisfying the following condition:

$$
\sup_{t \in [-T, nT]} |\mathbf{e}^n(t)| \overset{n \to \infty}{\longrightarrow} 0 \quad \text{a.s.}
$$

∎

Finally, we strengthen the convergence assumption on the functionals $\{b_n(.)\}$.

Assumption 3.6: In the previous section, we assumed that these functionals converged uniformly over a compact set to a functional $b(.)$. We now assume that the convergence is uniform over the entire state-space, i.e.,

$$
\sup_{\bar{x}_t \in C([0,T], \mathcal{R})} |b_n(\bar{x}_t, \bar{\mathbf{0}}) - b(\bar{x}_t, \bar{\mathbf{0}})| \overset{n \to \infty}{\longrightarrow} 0
$$

∎

We remark that if for all n, $b_n = b$, then, the above assumption is trivially satisfied.

Next, we present the following result has been proved in [23, Section 7.4, vol. II & Section 3.8, vol. I], [24, Theorem II.12].

Theorem 3.2: Consider a *perturbation* of the functional differential equation described in (12). For $t \in [0, \infty)$, consider the FDE

$$
\dot{y}(t) = b(\bar{y}_t, \bar{\mathbf{0}}) + g(\bar{y}_t) \tag{13}
$$

where $\bar{y}_t \in C([0,T], \mathcal{R})$, i.e., $\bar{y}_t = [y(s), s \in [t - T, t]]$, and the initial condition is given by

$$
y(t) = \tilde{\theta}(t) \quad -T \leq t \leq 0
$$

Assume that (12) satisfies Assumption 3.4. Fix any $\epsilon > 0$. Then, there exists $\delta_1 = \delta_1(\epsilon) > 0, \delta_2 = \delta_2(\epsilon) > 0$ such that if

$$
\begin{aligned}
|\tilde{\theta}(t)| &< \delta_1 \quad t \in [-T, 0] \\
|g(\bar{y}_t)| &< \delta_2 \quad t \geq 0
\end{aligned}
$$

then, for any $t \geq 0$, we have

$$|y(t)| \quad < \quad \epsilon$$

∎

This result says that if the unperturbed system is exponentially stable, and the initial condition and perturbation terms are small enough, then the trajectory of (12) and (13) will be "close" for all $t \geq 0$. Using this result, along with the result in the previous section, we relax the requirement on the initial conditions and next show that $\sup_{t \in [-T, nT]} |x(t) - \mathbf{x}^n(t)| < \epsilon$.

Theorem 3.3: Consider the systems described in (11) and (12). Suppose that (12) satisfies Assumptions 3.4, 3.5 and 3.6. Fix any $R > 0$ and $\epsilon > 0$. Then, there exists $N > 0$ such that for all $n \geq N$, for all initial trajectories $\theta(t)$ satisfying $\|\theta_0\| < R$, we have

$$\sup_{t \in [-T, nT]} |x(t) - \mathbf{x}^n(t)| \quad < \quad \epsilon$$

Proof. We can rewrite (11) as the the SFDE

$$
\begin{aligned}
\dot{\mathbf{x}}^n(t) &= b_n\left(\bar{\mathbf{x}}^n_{\frac{\lfloor nt \rfloor}{n}}, \bar{\mathbf{e}}^n_{\frac{\lfloor nt \rfloor}{n}}\right) \\
&= b\left(\bar{\mathbf{x}}^n_t, \bar{\mathbf{0}}\right) + \left(b_n\left(\bar{\mathbf{x}}^n_{\frac{\lfloor nt \rfloor}{n}}, \bar{\mathbf{e}}^n_{\frac{\lfloor nt \rfloor}{n}}\right) - b\left(\bar{\mathbf{x}}^n_t, \bar{\mathbf{0}}\right)\right) \\
&= b\left(\bar{\mathbf{x}}^n_t, \bar{\mathbf{0}}\right) + g\left(\bar{\mathbf{x}}^n_t, \mathbf{e}^n_{\frac{\lfloor nt \rfloor}{n}}\right) \quad (14)
\end{aligned}
$$

where

$$g\left(\bar{\mathbf{x}}^n_t, \mathbf{e}^n_{\frac{\lfloor nt \rfloor}{n}}\right) \quad = \quad \left(b_n\left(\bar{\mathbf{x}}^n_{\frac{\lfloor nt \rfloor}{n}}, \bar{\mathbf{e}}^n_{\frac{\lfloor nt \rfloor}{n}}\right) - b\left(\bar{\mathbf{x}}^n_t, \bar{\mathbf{0}}\right)\right)$$

Note that $g(.,.)$ is a function of only $\bar{\mathbf{x}}^n_t$ and $\bar{\mathbf{x}}^n_{\frac{\lfloor nt \rfloor}{n}}$ need not be included as an argument because knowledge of $\bar{\mathbf{x}}^n_t$ implicitly gives the trajectory $\bar{\mathbf{x}}^n_{\frac{\lfloor nt \rfloor}{n}}$ (appropriate straight line extrapolation).

Now fix any $\epsilon > 0$. As (12) is assumed to be exponentially stable, from Theorem 3.2, it follows that there exists a $\delta_1 > 0$, $\delta_2 > 0$ such that for any $t_0 > 0$, if

$$
\begin{aligned}
\|\bar{\mathbf{x}}^n_{t_0}\| &< \delta_1 \\
\left|g\left(\bar{\mathbf{x}}^n_t, \mathbf{e}^n_{\frac{\lfloor nt \rfloor}{n}}\right)\right| &< \delta_2 \quad \forall t \geq t_0
\end{aligned}
$$

then, we have for all $t \geq t_0$,

$$|\mathbf{x}^n(t)| \quad < \quad \frac{\epsilon}{2}$$

Let us denote $\delta = \frac{\min(\delta_1, \delta_2, \epsilon)}{2}$. Now, consider the mean FDE described in (12). For any initial condition satisfying $\|\bar{x}_t\| < R$, from Assumption 3.4, it follows that there exists t_0 such that for all $t \geq t_0$, $\|\bar{x}_t\| < \delta$. From Theorem 3.1, it follows that there exists N_1 such that $\forall n \geq N_1$,

$$\sup_{t \in [-T, t_0]} |\mathbf{x}^n(t) - x(t)| \quad < \quad \delta$$

Thus, it follows from the triangle inequality that for $n \geq N_1$, $\|\bar{\mathbf{x}}^n_{t_0}\| < \delta_1$. Next, we show that for all $t \geq t_0$, $|g(\bar{\mathbf{x}}^n_t, \mathbf{e}^n_{\frac{\lfloor nt \rfloor}{n}}) <$

δ_2. We have

$$
\begin{aligned}
g\left(\bar{\mathbf{x}}^n_t, \mathbf{e}^n_{\frac{\lfloor nt \rfloor}{n}}\right) &= \left(b_n\left(\bar{\mathbf{x}}^n_{\frac{\lfloor nt \rfloor}{n}}, \bar{\mathbf{e}}^n_{\frac{\lfloor nt \rfloor}{n}}\right) - b\left(\bar{\mathbf{x}}^n_t, \bar{\mathbf{0}}\right)\right) \\
&= \left(b_n\left(\bar{\mathbf{x}}^n_{\frac{\lfloor nt \rfloor}{n}}, \bar{\mathbf{e}}^n_{\frac{\lfloor nt \rfloor}{n}}\right) - b_n\left(\bar{\mathbf{x}}^n_{\frac{\lfloor nt \rfloor}{n}}, \bar{\mathbf{0}}\right)\right) \\
&\quad + \left(b_n\left(\bar{\mathbf{x}}^n_{\frac{\lfloor nt \rfloor}{n}}, \bar{\mathbf{0}}\right) - b_n(\bar{\mathbf{x}}^n_t, \bar{\mathbf{0}})\right) \\
&\quad + \left(b_n(\bar{\mathbf{x}}^n_t, \bar{\mathbf{0}}) - b(\bar{\mathbf{x}}^n_t, \bar{\mathbf{0}})\right)
\end{aligned}
$$

Now, b_n is Lipschitz continuous and bounded (with parameters L, M respectively). Thus, $\mathbf{x}^n(t)$ is Lipschitz continuous with parameter M. Further, choosing N_2 large enough, from Assumption 3.6 and 3.5, it follows that for all $n \geq N_2$,

$$
\begin{aligned}
\left(b_n(\bar{\mathbf{x}}^n_t, \bar{\mathbf{0}}) - b(\bar{\mathbf{x}}^n_t, \bar{\mathbf{0}})\right) &< \frac{\delta}{3L} \\
\sup_{t \in [-T, nT]} |\mathbf{e}^n(t)| &< \frac{\delta}{3}
\end{aligned}
$$

Thus, $\forall n \geq N_2$,

$$
\begin{aligned}
g\left(\bar{\mathbf{x}}^n_t, \mathbf{e}^n_{\frac{\lfloor nt \rfloor}{n}}\right) &< L\|\mathbf{e}^n_t\| + L\|\bar{\mathbf{x}}^n_t - \bar{\mathbf{x}}^n_{\frac{\lfloor nt \rfloor}{n}}\| + \frac{\delta}{3} \\
&\leq \frac{2\delta}{3} + L\|\bar{\mathbf{x}}^n_t - \bar{\mathbf{x}}^n_{\frac{\lfloor nt \rfloor}{n}}\| \\
&\leq \frac{2\delta}{3} + \frac{LM}{n}
\end{aligned}
$$

Thus, choosing n large enough, for all $t \geq t_0$, $|g(.,.)| < \delta$, and thus, $|\mathbf{x}^n(t)| < \frac{\epsilon}{2}$. The result then follows by the triangle inequality. ∎

IV. RATE-BASED MARKING AND THE MANY-FLOWS LIMIT

In this section, we show that in the many-flows regime, the trajectories for the stochastic difference and deterministic differential equations introduced earlier are "close."

A. Convergence to the Fluid Limit: The Unconstrained Case

We recall from the model description in Section II-A that there is a non-negativity constraint on each of the flows. We ignore this constraint in this subsection. This issue will be handled in the next subsection, where we will show that for "reasonable" initial conditions, the trajectories remain non-negative.

Formally, we consider the delay-differential equations

$$
\begin{aligned}
\dot{x}^n(t) &= \Delta - \beta x^n\left(\frac{\lfloor n(t-d) \rfloor}{n}\right) \quad (15) \\
&\quad p\left(x^n\left(\frac{\lfloor n(t-d) \rfloor}{n}\right) + a + e^n\left(\frac{\lfloor n(t-d) \rfloor}{n}\right)\right)
\end{aligned}
$$

with initial conditions given by

$$x^n(t) \quad = \quad \theta(t) + e^n(0) \quad -T \leq t \leq 0, \ nt \in \mathcal{Z}$$

and a straight line interpolation is used for $nt \notin \mathcal{Z}$. The candidate limiting system is described by

$$\dot{x}(t) \quad = \quad \Delta - \beta x(t-d)p(x(t-d) + a)$$

with initial conditions given by $\theta(t)$. Using Theorem 3.1, we show the following result.

Theorem 4.1: For the systems described above, we have

$$\sup_{t\in[0,T]} |x^n(t) - x(t)| \quad \to \quad 0 \quad \text{a.s.}$$

as $n \to \infty$.

Proof. We prove some properties of (15). First, we note that as $|e_i^n| < K$, we have that if $a + x^n(\frac{\lfloor n(t-d)\rfloor}{n}) < -K$, $p\left(x^n(\frac{\lfloor n(t-d)\rfloor}{n}) + a + e^n(\frac{\lfloor n(t-d)\rfloor}{n})\right) = 0$ (as $p(z) = 0$ for $z < 0$). Thus, we have that

$$-(K + a) \quad \leq \quad x^n(\frac{\lfloor n(t-d)\rfloor}{n})$$
$$p\left(x^n(\frac{\lfloor n(t-d)\rfloor}{n}) + a + e^n(\frac{\lfloor n(t-d)\rfloor}{n})\right)$$

and

$$\dot{x}^n(t) \quad \leq \quad \Delta + \beta(K + a)$$

As we are studying the process over the interval $[0, T]$, and $|x^n(0)| < K + \theta(0) \triangleq A$, we can uniformly upper bound (in n) $x^n(t)$ by $A + (\Delta + \beta(a + K))T$. Next, we lower bound $x^n(t)$. As $p(z) \leq 1$, and using our upper bound, we have

$$\dot{x}^n(t) \quad \geq \quad -\beta[A + (\Delta + \beta(a + K))T]$$

Thus, we have

$$x^n(t) \quad \geq \quad \theta(0) - K - T\beta[A + (\Delta + \beta(a + K))T]$$

Next, we define

$$M_x \quad = \quad \max\left(\begin{array}{l} |\theta(0) - K \\ -T\beta[A + (\Delta + \beta(a + K))T]|, \\ A + (\Delta + \beta(a + K))T \end{array}\right)$$
$$L_x \quad = \quad \max(\beta[A + (\Delta + \beta(a + K))T], \Delta + \beta(K + a))$$

we observe that $|x^n(t)| \leq M_x$ and is Lipschitz continuous with parameter L_x. Let $M = \max(K, M_x)$. For any $z \in \mathcal{R}$, define $[z] = (-M) \bigvee (z \bigwedge M)$. Define

$$b(x_t, e_t) \quad = \quad \Delta - \beta[x(t-d)] \, p\left([x(t-d)] + [e(t-d)] + a\right)$$

As we note that in (4), $|x^n(t)| \leq M$ and $|e^n(t)| \leq M$, we can rewrite it as

$$\dot{x}^n(t) \quad = \quad \Delta - \beta[x^n(\frac{\lfloor n(t-d)\rfloor}{n})] \quad (16)$$
$$p\left([x^n(\frac{\lfloor n(t-d)\rfloor}{n})] + a + [e^n(\frac{\lfloor n(t-d)\rfloor}{n})]\right)$$
$$= \quad b(x^n_{\frac{\lfloor nt\rfloor}{n}}, e^n_{\frac{\lfloor nt\rfloor}{n}}) \quad (17)$$

Now, we show that $b(.,.)$ is bounded and Lipschitz continuous. First, as $0 \leq p(.) \leq 1$, we have

$$|b(x_t, e_t)| \quad = \quad \left| \begin{array}{l} \Delta - \beta[x(t-d)] \\ p\left([x(t-d)] + [e(t-d)] + a\right) \end{array} \right|$$
$$\leq \quad \Delta + \beta M$$

Next, we have

$$|b(x_t, e_t) - b(y_t, r_t)|$$

$$= \quad |\beta[y(t-d)] \, p\left([y(t-d)] + [r(t-d)] + a\right)$$
$$-\beta[x(t-d)] \, p\left([x(t-d)] + [e(t-d)] + a\right) |$$
$$\leq \quad \beta|[y(t-d)] \, p\left([y(t-d)] + [r(t-d)] + a\right)$$
$$-[x(t-d)] \, p\left([y(t-d)] + [r(t-d)] + a\right)$$
$$+[x(t-d)] \, p\left([y(t-d)] + [r(t-d)] + a\right)$$
$$-[x(t-d)] \, p\left([x(t-d)] + [e(t-d)] + a\right) |$$

As $p(.)$ is Lipschitz continuous with parameter P, and $[x] \leq M$, we have

$$|b(x_t, e_t) - b(y_t, r_t)| \quad \leq \quad \beta|[y(t-d)] - [x(t-d)]|$$
$$+\beta PM(|[y(t-d)] - [x(t-d)]|$$
$$+|[e(t-d)] - [r(t-d)]|)$$
$$\leq \quad \beta(1 + PM)|x(t-d) - y(t-d)|$$
$$+\beta PM|e(t-d) - r(t-d)|$$
$$\leq \quad \beta(1 + PM)||(x_t, e_t) - (y_t, r_t)||$$

Thus, the conditions for Theorem 3.1 are satisfied, and we have the desired result, i.e., as $n \to \infty$, we have

$$\sup_{t\in[0,T]} |x^n(t) - x(t)| \quad \to \quad 0 \quad \text{a.s.}$$

■

Next, we study a similar limit for *each flow* as opposed to the aggregate flow. Recall that the rate of flow i (when there are n flows in the system) adapts according to (2), which can be represented by

$$\dot{y}^{i,n}(t) \quad = \quad \Delta - \beta y^{i,n}(\frac{\lfloor n(t-d)\rfloor}{n}) \quad (18)$$
$$p\left(x^n(\frac{\lfloor n(t-d)\rfloor}{n}) + a + e^n(\frac{\lfloor n(t-d)\rfloor}{n})\right)$$

with the initial conditions given by $\theta(t) + \tilde{e}_0^i$, sampled appropriately and interpolated. It can be shown using an analysis similar to that carried out for the aggregate flow that, as $n \to \infty$, the flow trajectory approaches the trajectory of the following delay-differential equation

$$\dot{y}^i(t) \quad = \quad \Delta - \beta y^i(t-d)p(x(t-d) + a) \quad (19)$$

and with *random* initial conditions given by $\theta(t) + \tilde{e}_0^i$, with $|\tilde{e}_0^i| < K$.

In the following section, we will address the issue of non-negativity of the trajectories of $y^{i,n}(t)$, $x^n(t)$ and $x(t)$. We will show that under reasonable initial conditions, for n large enough, the trajectories will remain non-negative.

B. Behavior of the Trajectories: Non-negativity and Convergence

We now study the pair of delay differential equations given by

$$\dot{x}(t) \quad = \quad \Delta - \beta x(t-d)p(x(t-d) + a) \quad (20)$$
$$\dot{y}(t) \quad = \quad \Delta - \beta y(t-d)p(x(t-d) + a) \quad (21)$$

with initial conditions over $[-T, 0]$ given by $\theta(t)$ and $\theta(t) + r_0$, respectively, for some $|r_0| < K$; and the marking function

$p(x)$ is chosen such that in the absence of delays, $x(t) \to c-a$, i.e., $\Delta = \beta(c-a)p(c)$. Equation (20) corresponds to the "fluid-limit" of the aggregate flow derived in the previous section, and (21) corresponds to a sample path of the "fluid-limit" of an individual flow. We first show the following result.

Lemma 4.1: If $0 \le \theta(t) \le c - a$, $\forall t \le 0$, then

$$l \le x(t) \le M, \qquad \forall t \ge 0,$$

where

$$M = c-a+\bar{\Delta}d \quad \text{and} \quad l = c-a-d(\beta M p(M+a)-\Delta).$$

Proof: Let $\bar{\Delta} = \Delta + \delta$, where δ was defined in Assumption 2.1. We now observe that $\dot{x}(t) \le \bar{\Delta}$. To see this, suppose $x(t-d) \le 0$, then we have from Assumption 2.1 that $|\Delta - \beta x(t-d)p(x(t-d)+a)| \le \bar{\Delta}$. On the other-hand, if $x(t-d) > 0$, as $p(.) \ge 0$, we have $\dot{x}(t) = \Delta - \beta x(t-d) \ p(a+x(t-d)) < \bar{\Delta}$. Let $M_1 = c-a+\bar{\Delta}d+\epsilon$ for some $\epsilon > 0$. Suppose that $x(t) = M_1$ for some $t \ge 0$. Then, since $\dot{x}(t) \le \bar{\Delta}$, we have that $x(t-d) > c-a$. Thus,

$$\dot{x}(t) = \Delta - \beta x(t-d)p(c) < \Delta - \beta(c-a)p(c) = 0.$$

In other words, if $x(t)$ reaches M_1, it decreases. Thus, if the initial trajectory $\theta(t) \le c-a$, we see that $x(t) \le c-a+\bar{\Delta}d$ for all $t \ge 0$.

Next, let $l_1 = l - \epsilon$ for some $\epsilon > 0$ and suppose that for some $t \ge 0$, $x(t) = l_1$. Then, as $x(t) \le M$, it follows that

$$x(t-d) \le l_1 + d(\beta M p(M+a) - \Delta) = c-a-\epsilon$$

Thus,

$$\begin{aligned} \dot{x}(t) &= \Delta - \beta x(t-d)p(x(t-d)) \\ &\ge \Delta - \beta(c-a-\epsilon)p(c-a-\epsilon) \end{aligned}$$

Since $\Delta - \beta(c-a)p(c) = 0$, $\dot{x}(t) > 0$, and the result follows. ∎

We note that, since $c-a > 0$ by our assumptions, it follows that $M > 0$. From now, we will also assume that $l > 0$. We justify this through a numerical example at the end of this section. In fact, for most parameters of interest in the Internet $l > K$, where we recall that noise process is bounded by $a \pm K$. For more discussion on this, see [6].

For the rest of this section, we also assume

$$\alpha \overset{\Delta}{=} \beta p(M+a)d < 1. \qquad (22)$$

To understand what this means, we note that it has been shown in [6] that, for reasonable values of network parameters, $M \approx c-a$. Further, we note that $\Delta = \beta(c-a)p(c)$. Thus,

$$\beta p(M+a)d \approx \frac{\Delta d}{c-a}.$$

For self-clocked congestion controllers (see [2]) $\Delta d \approx 1$ and since the bandwidth-per-user, $c-a$, is typically large, $\alpha < 1$ is satisfied.

Theorem 4.2: Suppose M, β, d satisfy (22) and $l > K$.

(i) For all $t \ge 0$, $|x(t) - y(t)| \le |x(0) - y(0)|$.
(ii) If the initial condition $y(0)$ is non-negative, i.e., $\theta(0) + r_0 \ge 0$, then $y(t) \ge 0$.
(iii) For any $\epsilon > 0$, there is a $\tau < \infty$ such that for all $t \ge \tau$, $|x(t) - y(t)| < \epsilon$.

Proof. Define $r(t) = x(t) - y(t)$. Then, the evolution of this is described by

$$\dot{r}(t) = -\beta r(t-d)p(a+x(t-d))$$

with $r(t) = r_0$ for $t \le 0$. If $r_0 = 0$, clearly, $r(t) = 0$ for all $t \ge 0$. Thus, we can assume $r_0 \ne 0$.

We first observe that if for all $s \in [t-d, t)$, $r(s) \le 0$, then we have

$$\begin{aligned} \dot{r}(t) &= -\beta r(t-d)p(a+x(t-d)) \\ &= \beta(-r(t-d))p(a+x(t-d)) \\ &\ge 0 \end{aligned}$$

On the other-hand, if for all $s \in [t-d, t)$, we have $r(s) \ge 0$, then $\dot{r}(t) \le 0$.

Let us now assume that $r_0 < 0$, and let t_1 the first time such that $r(t_1) = 0$. Then, for all $s \in [-d, t_1)$, $r(s) < 0$. Thus, from the observation in the previous paragraph, it follows that over $s \in [-d, t_1)$, $\dot{r}(s) \ge 0$. Thus, from the facts that $r_0 < 0$ and $r(t)$ is increasing, it follows that $\forall t \le t_1$, $|r(t)| < |r_0|$. Now, for any $t \in [t_1, t_1 + d]$, as $-r(t-d) \ge 0$, we have

$$\begin{aligned} \dot{r}(t) &= \beta(-r(t-d))p(x(t-d)+a) \\ &\ge 0 \end{aligned}$$

Thus, it follows that $r(t)$ is increasing over $[t_1, t_1 + d]$. Next, as $p(.)$ is increasing in it's argument and $x(t) \le M$, we have

$$\begin{aligned} \dot{r}(t) &= \beta(-r(t-d))p(x(t-d)+a) \\ &\le \beta(-r(t-d))p(M+a) \end{aligned}$$

Now, for any $t \in [t_1, t_1 + d]$, we have $0 \le -r(t-d) \le |r_0|$. Thus,

$$\begin{aligned} \dot{r}(t) &\le \beta(-r(t-d))p(M+a) \\ &\le \beta p(M+a)|r_0| \end{aligned}$$

with $r(t_1) = 0$. Thus, integrating the above expression, we have for any $t \in [t_1, t_1 + d]$,

$$\begin{aligned} 0 \le r(t) &\le \beta p(M+a)|r_0|(t-t_1) \\ &\le \beta p(M+a)|r_0|d \\ &= \alpha|r_0| \\ &< |r_0| \end{aligned}$$

where the last step follows from (22), where we have assumed that $\alpha < 1$.

Next, let t_2 be the first time after $t_1 + d$ such that $r(t) = 0$. As we have $r(t) \ge 0$ for $t \in [t_1, t_2]$, it follows that $|r(t)| < \alpha|r_0|$ for all $t \in [t_1, t_2]$. This follows because, $\alpha|r_0| \ge r(t) \ge 0$ in $[t_1, t_1 + d]$, and for $t \in [t_1 + d, t_2]$, $\dot{r}(t) \le 0$, because $r(t-d) \ge 0$ over this interval. Thus, $r(t) < \alpha|r_0|$ for all $t \in [t_1, t_2]$.

Now, repeat the above argument for the negative cycle (i.e., for $t \in [t_2, t_3]$, where t_3 is the first time $r(t)$ hits zero after $t_2 + d$) and lower bound $r(t)$ by $-\alpha^2|r_0| > -|r_0|$. Continuing this argument proves (i) and (ii).

Now noting that $|r_0| < K$ and by choosing $w > 0$ such that $\alpha^w K < \epsilon$, eventually we can upper bound $|r(t)|$ by $\alpha^w K < \epsilon$ provided that w zero crossings of $r(t)$ occur in finite time.

However, if the derivative were to be equal to zero, it is possible that sufficient number of zero crossings will not occur in finite time. This will happen if $p(a + x(t)) = 0$ for all t. However, noting that $\dot{x}(t) > 0$ if $p(x(t - d) + a) = 0$, it can be easily shown that there exists $\gamma > 0, 0 < \rho < 1$ and a time $\hat{T} > 0$ such that over *any* interval of time $\hat{T}$, $p(x(t) + a) > \gamma$ for at-least an amount of time $\rho\hat{T}$. Using this, it can be shown that zero crossings will occur in finite time intervals, and $r(t)$ will eventually go to zero. ∎

We now briefly discuss the implications of the above theorem. Recall from the previous section that as $n \to \infty$, the rate process of an individual flow $y^{i,n}(t)$ converges uniformly over $[0, T]$ to a fluid-limit process $y(t)$ with random initial conditions which are bounded by K, and similarly for the aggregate flow. Thus, from the results above, for any finite $T > 0$, it follows that for n large enough, the trajectories of $y^{i,n}(t)$ and $x^n(t)$ will remain non-negative. Further, by choosing $T = \tau$ (defined in (iii)), and choosing n large enough, it follows that $|x^n(T) - y^{i,n}(T)| < 2\epsilon$.

Remark 4.1: The results of this section can be extended to the case where the arrival rate at a router is estimated based on averaging over a finite window. In other words, consider a system where the marking at the router is a function of the average arrival rate over some interval of time, and bases the marking decision on this average. Consider the congestion control equation

$$\dot{x}^n(t) = \Delta - \beta x^n(\frac{\lfloor n(t-d) \rfloor}{n}) \tag{23}$$
$$p\left(\frac{n}{W} \int_{s=0}^{\frac{W-1}{n}} \left(\begin{array}{c} x^n(\frac{\lfloor n(t-s-d) \rfloor}{n}) + a \\ + e^n(\frac{\lfloor n(t-s-d) \rfloor}{n}) \end{array} \right) ds \right)$$

We observe that in the above system, the rate is integrated over an interval of time $\frac{W}{n}$, not W. Recall that each time-step (for the nth system) is of length $\frac{1}{n}$. Thus, in an interval of time $\frac{W}{n}$, we are averaging over W steps. This would correspond to averaging over some *volume of data*, so that instantaneous effects from a small fraction of the data does not affect the system dynamics. As the number of flows increase, the volume of data over any interval of time linearly increases. We see that to have the same reliability in the estimates, we need to keep the volume fixed for all n. Hence, the interval of time shrinks inversely in n. We can see that a reasonable limit to expect as $n \to \infty$ is given by (1), i.e.,

$$\dot{x}(t) = \Delta - \beta x(t-d)p(x(t-d) + a) \tag{24}$$

with initial conditions given by $\theta(t), 0 \geq t \geq -T$. ∎

Example 4.1: We consider a network with n flows. The target arrival rate due to controlled flows is assumed to be $n * 120$ kbytes/sec, and the mean arrival rate of uncontrolled flows is assumed to be $n * 30$ kbytes/sec. We let the round trip delay be 50 msec. To translate these numbers into our model, we need to map time in msec into time units of the model. We normalize with respect to one-way propagation delay. Thus, we have $d = 2$, $c = 3750$ and $a = 750$. Further let the gain parameters be $\Delta = 0.25$, and $\beta = 0.67$. Now, we assume that

the marking function is of the form

$$p(x) = \frac{(x - \tilde{c})^+}{x}$$

Thus, as $p(c)$ satisfies $\Delta = \beta(c-a)p(c)$, we have $\tilde{c} = 3749.5$. Computing the bounds M, l from the last section, we have $M = 3000.5$ and $l = 2999.5$, while the nominal rate per source $c_1 = c - a = 3000$. Further, checking the condition in (22), we have $\beta p(M + a)d = \frac{1}{2900} \ll 1$. Thus, the condition $\alpha < 1$ is satisfied. Recall that K is defined such that the "noise" rate is $a \pm K$. If K satisfies $l - K > 0$, i.e., $K < 2999.5$, and the initial conditions are non-negative, from the results in the previous section, the rates remain non-negative, and from (iii) in Section IV-B, all users eventually get a fair share of the bandwidth. ∎

C. Rate Based Marking: Extensions to Infinite Time

So far, we have studied convergence of rate processes over a finite time interval $[0, T]$. We now extend this framework to infinite time. As in Section III-A, we consider fix some $T > 0$ and consider a sequence of systems over progressively larger intervals of time. We study (15), but over the time interval $[-T, nT]$, i.e., for $t \in [0, nT]$, consider

$$\dot{x}^n(t) = \Delta - \beta x^n(\frac{\lfloor n(t-d) \rfloor}{n}) \tag{25}$$
$$p\left(x^n(\frac{\lfloor n(t-d) \rfloor}{n}) + a + e^n(\frac{\lfloor n(t-d) \rfloor}{n}) \right)$$

with initial conditions given by

$$x^n(t) = \theta(t) + e^n(0) \quad -T \leq t \leq 0, \; nt \in \mathcal{Z}$$

and a straight line interpolation is used for $nt \notin \mathcal{Z}$. For $t \in [0, \infty)$, the candidate limiting system is described by

$$\dot{x}(t) = \Delta - \beta x(t-d)p(x(t-d) + a) \tag{26}$$

with initial conditions given by $\theta(t)$. As before, the equilibrium point of the above equation is $x^* = c - a$. We make the following stability assumption.

Assumption 4.1: Consider the fluid limit process described by (26). We assume that this process is exponentially stable, i.e., $\exists \kappa > 0, \alpha > 0$ such that for all $t \geq 0$,

$$|x(t) - x^*| < \kappa ||\tilde{\theta}_0||e^{-\alpha t}$$

where $\tilde{\theta}_0 = \sup_{t \in [-T, 0]} |\theta(t) - x^*|$. ∎

Similar to that in the finite time case, there are now n uncontrolled flows (i.e., $\{e_i^k\}, k = 1, 2, \ldots, n, \; i \geq 0$) accessing the system, with each flow having a bounded rate, i.e., $|\tilde{e}_i^k| < K$. We assume that these flows are iid, and are stationary and ergodic. For $i < 0$, we assume a constant (random) initial condition, i.e., $e_{-i}^k = e_0^k$. In the finite time case, we were interested in the interval of time $[-T, T]$. In this interval, we had $2nT + 1$ discrete-time steps, and thus, i ranged from $-nT$ to nT. Here, we are studying the system over the time-interval $[-T, nT]$. As each discrete time-step corresponds to a "continuous time" interval of length $\frac{1}{n}$, it

follows that that i ranges from $-nT$ to n^2T. Then, we have for any $\epsilon > 0$,

$$\Pr\left(\sup_{-nT \leq i \leq n^2 T}\left|\frac{1}{n}\sum_{j=1}^{n}\tilde{e}_i^j\right| > \epsilon\right)$$

$$\leq (n^2T+1)\Pr\left(\left|\frac{1}{n}\sum_{j=1}^{n}\tilde{e}_0^j\right| > \epsilon\right)$$

$$\leq (n^2T+1)e^{-na}$$

for some $a = a(\epsilon) > 0$. The first step follows from the union bound, and the fact that the flows are stationary. The second step (i.e., existence of $a > 0$) follows from Chernoff's bound (it trivially follows that the MGF exists, as the process is bounded). Thus, it follows that

$$\Pr\left(\sup_{-nT \leq i \leq n^2 T}\left|\frac{1}{n}\sum_{j=1}^{n}\tilde{e}_i^j\right| > \epsilon\right) \xrightarrow{n \to \infty} 0$$

exponentially fast. Now, applying Borel-Cantelli Lemma, it follows that

$$\sup_{-nT \leq i \leq n^2 T}\left|\frac{1}{n}\sum_{j=1}^{n}\tilde{e}_i^j\right| \to 0 \quad \text{a.s.} \tag{27}$$

As before, we remark that the above model allows each "noise" process to be long-range dependent. We only need the flows to be bounded and iid. Thus, we have that

$$\sup_{t \in [-T, nT]}|e^n(t)| \to 0 \quad \text{a.s.}$$

The main result in this section is the following.

Theorem 4.3: For the systems described above, we have

$$\sup_{t \in [0, nT]}|x^n(t) - x(t)| \to 0 \quad \text{a.s.}$$

as $n \to \infty$.

Proof: Define $z^n(t) = x^n(t) - x^*$ and $z(t) = x(t) - x^*$. These are the "centered" rate processes, and satisfy the following equations. For $t \in [0, nT]$, $z^n(t)$ satisfies

$$\dot{z}^n(t) = \Delta - \beta(z^n(\frac{\lfloor n(t-d)\rfloor}{n}) + x^*) \tag{28}$$
$$p\left(z^n(\frac{\lfloor n(t-d)\rfloor}{n}) + c + e^n(\frac{\lfloor n(t-d)\rfloor}{n})\right)$$

with initial conditions given by

$$z^n(t) = \theta(t) + e^n(0) - x^* \quad -T \leq t \leq 0, \ nt \in \mathcal{Z}$$

and a straight line interpolation is used for $nt \notin \mathcal{Z}$. For $t \in [0, \infty)$, $z9t)$ satisfies

$$\dot{z}(t) = \Delta - \beta z(t-d)p(z(t-d) + c) \tag{29}$$

with initial conditions given by $\theta(t) - x^*$. As before, assume $|\theta(t)| < K$ is a Lipschitz continuous initial trajectory. We can show as in Lemma 4.1 that there exists $M > 0$ such that for all $n \geq 0$,

$$\sup_{t \in [-T, nT]}|z^n(t)| < M \tag{30}$$
$$\sup_{t \in [-T, \infty)}|z(t)| < M \tag{31}$$

The rest of the proof consists of showing that the assumptions in Theorem 3.3 are satisfied for the delay-differential equations described in (28) and (29). Let us define $L = \Delta + \beta(x^* + M)$. We observe that

$$\left| \begin{array}{l} \Delta - \beta(z^n(\frac{\lfloor n(t-d)\rfloor}{n}) + x^*) \\ p\left(z^n(\frac{\lfloor n(t-d)\rfloor}{n}) + c + e^n(\frac{\lfloor n(t-d)\rfloor}{n})\right) \end{array} \right| \leq L$$

Thus, it follows that $|\dot{z}^n(t)| \leq L$, and similarly for $z(t)$, and hence, the trajectories of $z^n(t)$ are Lipschitz continuous with parameter L.

As in Lemma 4.1, for any $z \in \mathcal{R}$, define $[z] = (-M) \bigvee (z \bigwedge M)$. Define

$$b(z_t, e_t) = \Delta - \beta$$
$$([z(t-d)] + x^*) \ p([x(t-d)] + [e(t-d)] + c)$$

Then, we can rewrite (28) and (29) as

$$\dot{z}^n(t) = \Delta - \beta([z^n(\frac{\lfloor n(t-d)\rfloor}{n})] + x^*) \tag{32}$$
$$p\left([z^n(\frac{\lfloor n(t-d)\rfloor}{n})] + c + [e^n(\frac{\lfloor n(t-d)\rfloor}{n})]\right)$$
$$= b(z^n_{\frac{\lfloor nt \rfloor}{n}}, e^n_{\frac{\lfloor nt \rfloor}{n}}) \tag{33}$$
$$\dot{z}(t) = \Delta - \beta([z(t-d)] + x^*) \ p([z(t-d)] + c)$$
$$= b(z_t, \bar{0}) \tag{34}$$

Now, we show that $b(.,.)$ is bounded and Lipschitz continuous. First, as $0 \leq p(.) \leq 1$, we have

$$|b(z_t, e_t)| = \left| \begin{array}{l} \Delta - \beta([z(t-d)] + x^*) \\ p([z(t-d)] + [e(t-d)] + c) \end{array} \right|$$
$$\leq \Delta + \beta M$$

The proof of the Lipschitz property of $b(.,.)$ is identical to that in Lemma 4.1.

Next, we have, by Assumption 4.1, that the fluid-limit is exponentially stable. Finally, as discussed earlier, the "noise" process satisfies

$$\sup_{t \in [-T, nT]}|e^n(t)| \xrightarrow{n \to \infty} 0$$

Thus, the conditions for Theorem 3.3 are satisfied, and we have the desired result, i.e., as $n \to \infty$, we have

$$\sup_{t \in [0, nT]}|x^n(t) - x(t)| \to 0 \quad \text{a.s.}$$

∎

D. Weighted Proportional Fairness

In the previous sections, we considered the case where all users share the same utility function $\log(x)$. We now generalize this to the case where the utility function of Flow i is $\Delta_i \log(x_i)$ (leading to *weighted proportional fairness* [3]) where $\Delta_i > 0$ are possibly different for different flows. From a resource allocation point of view, the resource allocation achieved under any concave, increasing utility functions can be achieved by a weighted proportionally-fair allocation through appropriate choice of weights $\{\Delta_i\}$ [3]. To implement this in a truly decentralized fashion, the weights themselves have to be adapted by the individual users; however, we do not address this case here.

Suppose there are n flows, and user i uses the utility function $\Delta_i^n \log(y_i^n)$. Let us define

$$\Delta^n = \frac{1}{n}\sum_{i=1}^{n}\Delta_i^n$$

Continuing the same notation and framework as in the previous sections, the evolution of the user's rate follows the delay-differential equation

$$\dot{y}^{i,n}(t) = \Delta_i^n - \beta y^{i,n}\left(\frac{\lfloor n(t-d)\rfloor}{n}\right)$$
$$p\left(x^n\left(\frac{\lfloor n(t-d)\rfloor}{n}\right) + a + e^n\left(\frac{\lfloor n(t-d)\rfloor}{n}\right)\right). \quad (35)$$

with initial conditions given by

$$y^{i,n}(t) = \frac{\Delta_i^n}{\Delta}\theta(t) + \tilde{e}_n^i(0) \quad -T \le t \le 0,\ nt \in \mathcal{Z}$$

and a straight line interpolation is used for $nt \notin \mathcal{Z}$. We can then describe the evolution of the mean rate analogous to (15) by

$$\dot{x}^n(t) = \Delta^n - \beta x^n\left(\frac{\lfloor n(t-d)\rfloor}{n}\right)$$
$$p\left(x^n\left(\frac{\lfloor n(t-d)\rfloor}{n}\right) + a + e^n\left(\frac{\lfloor n(t-d)\rfloor}{n}\right)\right) \quad (36)$$

with initial conditions given by

$$x^n(t) = \frac{\Delta^n}{\Delta}\theta(t) + e^n(0) \quad -T \le t \le 0,\ nt \in \mathcal{Z}$$

and a straight line interpolation is used for $nt \notin \mathcal{Z}$. This means that each flow $\{y^k\},\ k = 1, 2, \ldots, n$ could have a different initial condition, but the nominal initial condition is given by $\frac{\Delta^n}{\Delta}\theta(t)$.

Suppose we assume that $\Delta^n \to \Delta$ as $n \to \infty$. Using methods identical to those in Section IV-A, we can show that the trajectory of (36) is close to that described by

$$\dot{x}(t) = \Delta - \beta x(t-d)p(x(t-d) + a) \quad (37)$$

with initial conditions given by $\theta(t)$. Formally,

Lemma 4.2: As $n \to \infty$, we have

$$\sup_{t \in [0,T]} |x^n(t) - x(t)| \to 0$$

∎

Analogous to the study in Section IV-A, we can study a limit for *each flow* as opposed to the aggregate flow. It can be shown that the trajectory of (35) approaches the trajectory of the following delay-differential equation

$$\dot{y}^i(t) = \Delta_i - \beta y^i(t-d)p(x(t-d) + a) \quad (38)$$

and with *random* initial conditions given by $\frac{\Delta_i}{\Delta}\theta(t) + \tilde{e}_0^i$, with $|\tilde{e}_0^i| < K$.

Non-negativity of the trajectories follows the analysis in Section IV-B, once we realize that we need to study a *scaled trajectory* of $y^i(t)$. Let us define

$$y(t) = \frac{\Delta y^i(t)}{\Delta_i}.$$

Then, we have

$$\dot{y}(t) = \Delta - \beta y(t-d)p(x(t-d) + a)$$

with initial conditions over $[-T, 0]$ given by $\theta(t) + r_0'$, respectively, for some $|r_0'| < K'$, where $K' = \frac{\Delta K}{\Delta_i}$. The results in Section IV-B then follow for these scaled system, and thus, non-negativity is ensured under suitable initial conditions.

V. RANDOM MARKING

In this section, we consider random marking functions, and show that in the many-flows limit, this random function behaves like a deterministic function operating on the average rate as in the previous sections.

We begin with motivating why we should study random marking. Consider a packet-based model operating in discrete-time. In a particular time-slot, suppose 10 packets arrive, and the marking function $p(.)$, takes the value 0.2. Then, every packet needs to marked with the value 0.2. However, in practice, a packet is either marked or not marked. Thus, a possible solution is to mark each packet independently with probability 0.2. Thus, we can see that the marking function is random with mean $p(.)$.

Now, we go back to our stochastic fluid model. Suppose there are n flows $y^{i,n}(t), i = 1, 2, \ldots, n$ with t such that $nt \in \mathcal{Z}$ and $0 \le t \le T$. Ideally, a fraction $y^{i,n}(t)p(.)$ of the ith flow should be marked. However, due to randomness in marking, the fraction marked will be $y^{i,n}(t)(p(.) + \xi^{i,n}(t))$, where $\xi^{i,n}(t), i = 1, 2, \ldots, n$ are iid random variables with zero mean, and $|\xi^{i,n}(t)| \le 1$. Further, we have that $y^{i,n}(t)$ is independent of $\xi^{i,n}(t)$. This follows because $y^{i,n}(t)$ depends on the noise and marking up to time $\frac{\lfloor n(t-d)\rfloor}{n}$, and the randomness in the marking process is assumed to be independent of past marking decisions (i.e., at each time, a flow is marked independently of the past, and with mean $p(.)$). Thus, analogous to (2), the congestion control model for each flow will be given by

$$\dot{y}^{i,n}(t) = \Delta - \beta y^{i,n}\left(\frac{\lfloor n(t-d)\rfloor}{n}\right)$$
$$\left(\begin{array}{c} p\left(x^n\left(\frac{\lfloor n(t-d)\rfloor}{n}\right) + a + e^n\left(\frac{\lfloor n(t-d)\rfloor}{n}\right)\right) \\ + \xi^{i,n}\left(\frac{\lfloor n(t-d)\rfloor}{n}\right) \end{array}\right)$$

with suitable non-negative initial conditions. Also, as before in (4), let

$$x^n(t) = \frac{1}{n}\sum_{i=1}^{n} y^i(t)$$

We first state the following lemma that the states that the individual sources rates and the average source rate are bounded. We omit the proof since it is similar to earlier proofs.

Lemma 5.1: There exists $M > 0$ such that for $t \in [0, T]$, for all n,

$$|y^{i,n}(t)| \leq M$$
$$|x^n(t)| \leq M$$

∎

Next, from the definition of $x^n(t)$, we can see that is satisfies the following delay-differential equation. For any $t \in [0, T]$, we have

$$\dot{x}^n(t) = \Delta - \beta x^n\left(\frac{\lfloor n(t-d) \rfloor}{n}\right)$$
$$p\left(x^n\left(\frac{\lfloor n(t-d) \rfloor}{n}\right) + a + e^n\left(\frac{\lfloor n(t-d) \rfloor}{n}\right)\right)$$
$$-\beta \xi^n\left(\frac{\lfloor n(t-d) \rfloor}{n}\right) \qquad (39)$$

where

$$\xi^n(t) = \frac{1}{n}\sum_{i=1}^{n} y^{i,n}(t)\xi^{i,n}(t) \qquad (40)$$

We now show that $\xi^n(t) \to 0$ almost surely, uniformly over $t \in [0, T]$ as $n \to \infty$.

Lemma 5.2: As $n \to \infty$,

$$\sup_{t \in [0,T]} |\xi^n(t)| \to 0 \quad \text{a.s.}$$

Proof. To show this, it is sufficient to show that

$$\Pr\left(\sup_{\substack{t \in [0,T] \\ nt \in \mathcal{Z}}} |\xi^n(t)| > \epsilon\right) \xrightarrow{n \to \infty} 0$$

fast enough, so that we can use the Borel-Cantelli lemma to show almost sure convergence. From the union bound, we have

$$\Pr\left(\sup_{\substack{t \in [0,T] \\ nt \in \mathcal{Z}}} |\xi^n(t)| > \epsilon\right) \leq \sum_{i=0}^{nT} \Pr\left(|\xi^n(\frac{i}{n})| > \epsilon\right)$$

We consider any term of this summation and show that it goes to zero at least as fast as $\frac{1}{n^3}$. Thus, as there are only n terms in the summation, we have convergence at rate $\frac{1}{n^2}$. From Borel-Cantelli Lemma, the result then follows. We have $\xi^n(1) = \frac{1}{n}\sum_{i=1}^{n} y^{i,n}(1)\xi^{i,n}(1)$. Thus, from Chebyshev's inequality, we have we have

$$\Pr\left(|\frac{1}{n}\sum_{i=1}^{n} y^{i,n}(1)\xi^{i,n}(1)| > \epsilon\right)$$

$$\leq \frac{1}{n^6\epsilon^6}E\left((\sum_{i=1}^{n} y^{i,n}(1)\xi^{i,n}(1))^6\right)$$

and $|y^{i,n}(1)| < M$, $|\xi^{i,n}(1)| \leq 1$. Thus, expanding and using the fact the $\xi^{i,n}(1)$ and $\xi_j(1)$ are independent, zero mean random variables (and thus, many product-terms cancel), we have

$$\Pr\left(|\frac{1}{n}\sum_{i=1}^{n} y^{i,n}(1)\xi^{i,n}(1)| > \epsilon\right) \leq$$

$$\frac{1}{n^6\epsilon^6}E\left(\begin{array}{c} \sum_{i=1}^{n}(y^{i,n}(1)\xi^{i,n}(1))^6 \\ +15\sum_{\substack{i,j \\ i \neq j}}(y^{i,n}(1)\xi^{i,n}(1))^4(y_j(1)\xi_j(1))^2 \\ +20\sum_{\substack{i,j \\ i \neq j}}(y^{i,n}(1)\xi^{i,n}(1))^3(y_j(1)\xi_j(1))^3 \\ +\sum_{\substack{i,j,k \\ i \neq j \neq k}} (y^{i,n}(1)\xi^{i,n}(1))^2 \\ (y_j(1)\xi_j(1))^2(y_k(1)\xi_k(1))^2 \end{array}\right)$$

Now, using the boundedness of $y^{i,n}(1)$ and $\xi^{i,n}(1)$, we have

$$\Pr\left(|\frac{1}{n}\sum_{i=1}^{n} y^{i,n}(1)\xi^{i,n}(1)| > \epsilon\right)$$

$$\leq \frac{1}{n^6\epsilon^6}(n + 15n^2 + 20n^2 + n^3)M^6$$
$$\leq \frac{37M^6}{\epsilon^6 n^3}$$

and the result follows. Thus, we have $\sup_{t \in [0,T]} |\xi^n(t)| \to 0$ almost surely as $n \to \infty$. ∎

Now, we can show that the appropriate limit to study even with random marking is the same as that of rate-based marking without noise in the marking process, i.e.,

$$\dot{x} = \Delta - \beta x(t-d)p(x(t-d) + a) \qquad (41)$$

with initial conditions given by $\theta(t), 0 \geq t \geq -T$. We state the following result without proof since the proof is similar to the proof of Theorem 4.1, with minor modifications that use Lemma 5.2.

Lemma 5.3: For the systems described in (39) and (41), we have

$$\sup_{t \in [0,T]} |x^n(t) - x(t)| \to 0 \quad \text{a.s.}$$

as $n \to \infty$. ∎

A. Non-negativity of the Trajectories

In the previous section, we studied a relaxed problem, where the user rates were not constrained to be non-negative. Similar to Section IV-B, we will now show that under suitable assumptions, the user rates will remain non-negative.

First, we observe that as the limit of the *aggregate flow* $x^n(t)$ is the same as that in Section IV-B, it follows that under reasonable initial conditions, the limiting process $x(t)$ is non-negative, and bounded away from 0 for all $t \in [0, T]$. As we have shown in the previous section that the aggregate rate $x^n(t)$ converges to $x(t)$ uniformly over $[0, T]$, it follows that for n large enough, the aggregate rate defined in (39) will remain non-negative.

We now study each individual user's rate, i.e., for n large enough, for $t \in [0, T]$, we study

$$
\dot{y}^{i,n}(t) = \Delta - \beta y^{i,n}(\frac{\lfloor n(t-d)\rfloor}{n})
$$
$$
\left(
\begin{array}{l}
p\left(x^n(\frac{\lfloor n(t-d)\rfloor}{n}) + a + e^n(\frac{\lfloor n(t-d)\rfloor}{n})\right) \\
+\xi^{i,n}(\frac{\lfloor n(t-d)\rfloor}{n})
\end{array}
\right) \quad (42)
$$

Consider the randomness in marking for each flow, i.e., $\xi^{i,n}(t)$. Suppose that the variability in the marking is small, i.e., $|\xi^{i,n}(t)| < \gamma$ for some small $\gamma > 0$. This assumption can be justified by the following reasoning. We consider for the moment, the packet model we had used to motivate random marking. Suppose that at some time, each is to be marked with probability p. For a user with only one packet in the queue, the variance in the number of marks received will be $p(1-p)$. However, it the user has a large number of packets (N packets) in the queue, then the variance in the average number of marks received will be $O(\frac{1}{N})$. The analog in the rate based model is the large rate regime. If each user has a large enough data rate, then a law of large numbers result will ensure that the variability in marking per user at any time is small. In the following Lemma, we will assume that the user data rates are large enough for this regime to operate.

Lemma 5.4: There exists some $\gamma > 0$ such that $\sup_{t \in [0,T]} |\xi^{i,n}(t)| < \gamma$, then, for all n large enough, for all $t \in [0, T]$, $y^{i,n}(t) \geq 0$.

Proof. Fix any i and define $z^n(t) = x^n(t) - y^{i,n}(t)$, where $x^n(t)$ is defined in (39). Then, $z^n(t)$ satisfies

$$
\dot{z}^n(t) = -\beta z^n(\frac{\lfloor n(t-d)\rfloor}{n})
$$
$$
p\left(x^n(\frac{\lfloor n(t-d)\rfloor}{n}) + a + e^n(\frac{\lfloor n(t-d)\rfloor}{n})\right)
$$
$$
+\beta \xi^{i,n}(\frac{\lfloor n(t-d)\rfloor}{n})y^n(\frac{\lfloor n(t-d)\rfloor}{n})
$$
$$
-\beta \xi^n(\frac{\lfloor n(t-d)\rfloor}{n})x^n(\frac{\lfloor n(t-d)\rfloor}{n}) \quad (43)
$$

Now, recall from Section IV-B that under reasonable initial conditions, there is some $M > c - a$ and $l > 0$ such that (41) satisfies $l < \sup_{t \in [0,T]} x(t) < M$ (in fact, the bound is valid for all $t \geq 0$). Assume that (22) is satisfied, i.e., we assume that M, β, d satisfy

$$
\alpha \triangleq \beta p(M + a)d < 1
$$

A discussion of this condition is provided in Section IV-B. Next, as the above inequality is strict, and $p(.)$ is continuous and increasing, there exists some $\epsilon > 0$ such that for all $0 \leq \delta \leq \epsilon$, $\beta p(M + a + \delta)d < 1$. Now, we choose n large enough such that

$$
\sup_{t \in [-T, T]} |e^n(t)| < \frac{\epsilon}{2}
$$
$$
\sup_{t \in [-T, T]} |x^n(t) - x(t)| < \frac{\epsilon}{2}
$$

Thus, it follows that for all $t \in [0, T]$,

$$
\beta p\left(x^n(\frac{\lfloor n(t-d)\rfloor}{n}) + a + e^n(\frac{\lfloor n(t-d)\rfloor}{n})\right)d < 1 \quad (44)
$$

We now consider (43), but without any random perturbation of the marking process, i.e., for $t \in [0, T]$, we consider the delay-differential equation

$$
\dot{r}^n(t) = -\beta r^n(\frac{\lfloor n(t-d)\rfloor}{n})
$$
$$
p\left(x^n(\frac{\lfloor n(t-d)\rfloor}{n}) + a + e^n(\frac{\lfloor n(t-d)\rfloor}{n})\right)
$$

Using (44), an analysis of this delay-differential equation in a manner identical to that in Lemma 4.2 yields that $|r^n(t)| \leq |r^n(0)|$. In fact, even though we will not use this here, the analysis will yield that $r^n(t)$ converges to zero exponentially fast (recall that we have a geometrically decreasing bound on the trajectory).

Now, consider the "marking perturbation" terms in (43), i.e.,

$$
\beta\left(
\begin{array}{l}
\xi^{i,n}(\frac{\lfloor n(t-d)\rfloor}{n})y^n(\frac{\lfloor n(t-d)\rfloor}{n}) \\
-\xi^n(\frac{\lfloor n(t-d)\rfloor}{n})x^n(\frac{\lfloor n(t-d)\rfloor}{n})
\end{array}
\right)
$$

From the discussion above, we have that $x^n(\frac{\lfloor n(t-d)\rfloor}{n})$ is bounded by $M + \frac{\epsilon}{2}$. Further, from Lemma 5.2, we know that as $n \to \infty$, $\xi^n(\frac{\lfloor n(t-d)\rfloor}{n})$ goes to zero uniformly over $[0, T]$. Thus, it follows that

$$
\sup_{t \in [0,T]} |\xi^n(\frac{\lfloor n(t-d)\rfloor}{n})x^n(\frac{\lfloor n(t-d)\rfloor}{n})| \xrightarrow{n\to\infty} 0 \quad \text{a.s.}
$$

Using ideas similar to that in Lemmas 3.1 and 3.2, where Gronwall's inequality on the sup-norm is used to give a bound on differences between the trajectories of systems with small perturbations, we can show that $\exists \gamma > 0$ small enough such that if $|\xi^{i,n}(t)| < \gamma$, then, $\sup_{t \in [0,T]} |r^n(t) - z^n(t)|$ is small. Thus, it will follow that over $[0, T]$, by choosing γ small enough, we can ensure that $y^{i,n}(t)$ remains non-negative. ∎

For the marking schemes described in this section, the difficulty in extending the convergence result to arbitrarily large intervals of time lies in proving uniform boundedness of the trajectories with random perturbations, i.e., the analog of (30)-(31). If the existence of a uniform finite bound can be proved, the extension to infinite time will easily follow. This is an interesting problem for future research.

VI. QUEUE-BASED MARKING

We will now study systems where the marking function could be based on queue lengths as described in Section II-B. As in th previous section, we will first consider the case where the trajectories can be negative or positive, and prove convergence in the many flows regime over a finite interval of time. Then, we show that the trajectories are non-negative for suitable initial conditions.

A. Convergence to the Fluid Limit: The Unconstrained Case

Like systems considered earlier, we assume n flows access the router. However, unlike before, where the marking depended on the instantaneous rates (or average rate over a time-window), here, we consider marking functions based on queue-lengths.

As described in Section II-B, let $x^n(t)$ be the aggregate average data rate at the router, with capacity nC. We have the uncontrolled "noise" process given by $a + e^n(t)$, where $a > 0$, $e^n(t)$ is zero-mean process satisfying suitable assumptions as before. We let $c = C - a$. Then, for $t \in [0, T]$, the evolution of the aggregate rate can be described by the following SFDE:

$$\dot{x}^n(t) \quad = \quad \Delta - \beta x^n(\frac{\lfloor n(t-d) \rfloor}{n}) \left(p\left(q^n(\frac{\lfloor n(t-d) \rfloor}{n}) \right) \right) \quad (45)$$

where $q^n(t)$ is the scaled queue-length. We assume as before that the marking function $p(.)$ is chosen so that for each n, the fixed point is $(x(t), q(t)) = (c, q_0)$. Thus, the marking function acts on the scaled (by $\frac{1}{n}$) queue-length, and whose evolution is given by

$$\dot{q}^n(t) \quad = \quad \begin{cases} x^n(t) + e^n(t) - c & \text{if } q^n(t) > 0 \\ (x^n(t) + e^n(t) - c)^+ & \text{if } q^n(t) = 0 \end{cases} \quad (46)$$

As before, assume that the initial condition is given by

$$x^n(t) \quad = \quad \theta(t) + e^n(0) \quad -T \le t \le 0, \, nt \in \mathcal{Z}$$

and a straight line interpolation is used for $nt \notin \mathcal{Z}$, and the $q^n(t)$ satisfies

$$q^n(t) \quad = \quad 0 \quad t \in [-T, 0]$$

Also consider a deterministic system consisting of a single flow $x(t)$ driving a queue with capacity c, and whose evolution is given by

$$\dot{x}(t) \quad = \quad \Delta - \beta x(t - d) p\left(q(t - d) \right) \quad (47)$$

$$\dot{q}(t) \quad = \quad \begin{cases} x(t) - c & \text{if } q(t) > 0 \\ (x(t) - c)^+ & \text{if } q(t) = 0 \end{cases} \quad (48)$$

and $q(t) = 0$ for $t \in [-T, 0]$. We assume $p(.)$ is a *queue-based* marking function. The marking function is chosen such that in equilibrium, $\dot{x} = 0$, $\dot{q} = 0$, $q(t) = q_0$, $x(t) = c$, and

$$\Delta \quad = \quad \beta c p(q_0)$$

Lemma 6.1: For the system described above, as $n \to \infty$, we have

$$\sup_{t \in [0, T]} |x^n(t) - x(t)| \quad \to \quad 0 \quad \text{a.s.}$$

Proof. We first show boundedness of the trajectories, i.e., $\exists M > 0$ such that for all n,

$$\sup_{t \in [0, T]} |x^n(t)| \quad < \quad M$$

To see this, from the fact that $p(,) \le 1$, we have

$$\dot{x}^n(t) \quad \le \quad \Delta + \beta |x^n(\frac{\lfloor n(t-d) \rfloor}{n})|$$

The trajectory of $x^n(t)$ can thus seen to be upper bounded by the trajectory of $z(t)$ (with non-negative initial conditions) satisfying

$$\dot{z}(t) \quad = \quad \Delta + \beta z(\frac{\lfloor n(t-d) \rfloor}{n})$$
$$\le \quad \Delta + \beta z(t)$$

The second step follows as $z(t)$ is increasing in t. Thus, clearly, over $t \in [0, T]$, the trajectory of $x^n(t)$ is uniformly

bounded. Similarly, a lower bound can be derived, and we can show that there exists $M_1 > 0$ such that $|x^n(t)| < M_1$ for $t \in [0, T]$.

Recall that $q^n(t) = 0$ for all $t \le 0$. Thus, for any $t \in [0, T]$, the queue-length $q^n(t)$ depends only on the trajectory of $x^n(t)$ in the interval $[0, t] \subset [t-T, t]$. Thus, we can interpret $q^n(t)$ as a functional mapping of the trajectory of $x^n(t)$ to a non-negative number. We will now show that this mapping is Lipschitz continuous, so that we can apply Theorem 3.1.

Formally, we define $M_2 = K + M_1$ and let $[z] = (-M_2) \bigvee (z \bigwedge M_2)$. We let $Q : C([0, T], \mathcal{R}^2) \times [0, T] \mapsto \mathcal{R}$ to be defined as follows. With $\bar{y} = \{y(v), 0 \le v \le T\}$, a continuous trajectory of length T, we define

$$Q(\bar{y}, s) \quad = \quad \begin{cases} \tilde{q}(s - d) & s \in [d, T] \\ 0 & s \in [0, d) \end{cases}$$

where $\tilde{q}(s - d)$ is the queue-length at time $s - d$ (with $s \ge d$) of a queueing system with zero initial conditions, and driven by the trajectory $\{[y(T - s + v)], 0 \le v \le s - d\}$. This is a "sliced" and bounded segment of the trajectory of $\bar{y}$ over the interval $[T - s, T - d]$, which is of length $s - d$. Formally, we define $\tilde{q}(t)$ by the following differential equation:

$$\dot{\tilde{q}}(r) \quad = \quad \begin{cases} [y(T - s + r)] - c & \text{if } \tilde{q}(r) > 0 \\ ([y(T - s + r)] - c)^+ & \text{if } \tilde{q}(r) = 0 \end{cases}$$

and $\tilde{q}(r) = 0$ for $r \le 0$. From the above definition, the following properties of $Q(\bar{y}, s)$ follow:

$$|Q(\bar{y}, t) - Q(\bar{y}, s)| \quad \le \quad (M_2 + c)|t - s|$$
$$|Q(\bar{y}, t) - Q(\bar{x}, t)| \quad \le \quad T \| \bar{y} - \bar{x} \|$$

Finally, we note that for the rate trajectories $x^n(t)$ and $x(t)$, we have

$$Q(x^n_{\frac{\lfloor nt \rfloor}{n}}, \frac{\lfloor nt \rfloor}{n}) \quad = \quad q(\frac{\lfloor n(t-d) \rfloor}{n})$$
$$Q(x_t, t) \quad = \quad q(t - d)$$

It can also be easily shown that $Q(\bar{y}, s)$ is bounded over $\bar{y} \in C[0, T]$ and $s \in [0, T]$. Let us denote the bound by M_3. Defining $M = \max(M_2, M_3)$ and redefining $[z] = (-M) \bigvee (z \bigwedge M)$, we define

$$b(x_t, e_t, t) \quad = \quad \Delta - \beta([x(t - d)]) \, p([Q(x_t, t)])$$

It can easily be shown that $b(., ., .)$ is a Lipschitz continuous, bounded functional using techniques identical to that in Lemma 4.1. Equations (45) and (47) can be written in terms of $b(., .)$ as

$$\dot{x}^n(t) \quad = \quad b(x^n_{\frac{\lfloor nt \rfloor}{n}}, e^n_{\frac{\lfloor nt \rfloor}{n}}, \frac{\lfloor nt \rfloor}{n})$$
$$\dot{x}(t) \quad = \quad b(x_t, \bar{0}, t)$$

We can easily generalize Theorem 3.1 to include convergence of finite-dimensional vector processes, i.e., we can let $\mathbf{x}^n(t), \mathbf{e}^n(t)$ to be vector processes of dimension J, and let $b_n(., .), b(., .)$ map vector processes to $\mathcal{R}^J$. With similar assumptions as before (i.e., the functions are bounded and Lipschitz continuous), we can derive Theorem 3.1 in this more general setting.
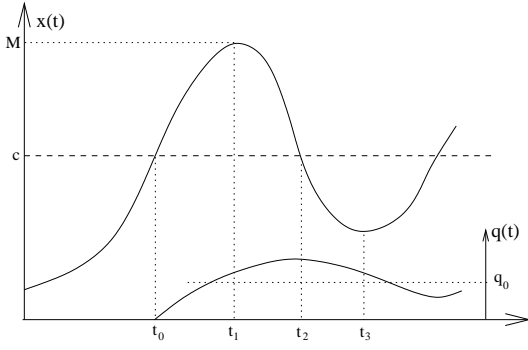
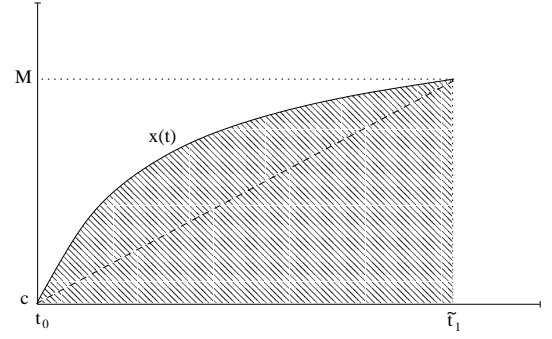Fig. 1. Typical trajectories for the rate process $x(t)$ and the queue length process $q(t)$.

Fig. 2. Trajectory of the rate process $x(t)$. The shaded area corresponds to the queue length at time $\tilde{t}_1$.

In this case, we have $J = 2$. We treat $(x(t), t)$ to be the vector process. Using the fact the t is Lipschitz continuous (with parameter 1) and bounded (by T), from a vector version of Theorem 3.1, it follows that

$$\sup_{t \in [0,T]} |x^n(t) - x(t)| \quad \to \quad 0 \quad \text{a.s.}$$

as $n \to \infty$. ∎

Next, as in Section IV-A, we study a similar limit for *each flow* as opposed to the aggregate flow. Recall that the rate of flow i (when there are n flows in the system) adapts according to (2), which can be represented by

$$\dot{y}^{i,n}(t) \quad = \quad \Delta - \beta y^{i,n}(\frac{\lfloor n(t-d) \rfloor}{n}) \, p\left(q^n(\frac{\lfloor n(t-d) \rfloor}{n})\right)$$

with the initial conditions given by $\theta(t) + \tilde{e}_0^i$, sampled appropriately and interpolated. It can be shown using an analysis similar to that carried out for the aggregate flow that as $n \to \infty$, the flow trajectory approaches the trajectory of the following delay-differential equation

$$\dot{y}^i(t) \quad = \quad \Delta - \beta y^i(t-d)p(q(t-d)) \tag{49}$$

and with *random* initial conditions given by $\theta(t) + \tilde{e}_0^i$, with $|\tilde{e}_0^i| < K$.

In the following section, we will address the issue of non-negativity of the trajectories of $y^{i,n}(t)$, $x^n(t)$ and $x(t)$. We will show that under reasonable initial conditions, for n large enough, the trajectories will remain non-negative.

B. The Fluid Limit: Bounds for Queue Based Marking

We study the pair of delay differential equations given by (47) and (49), with the queue-length process driven by (48). We first show the following result.

Lemma 6.2: Define M, q_{max}, l to be

$$\Delta \quad = \quad \beta(M - \Delta d)p\left(\frac{(M - \Delta d - c)^2}{2\Delta}\right)$$

$$q_{max} \quad = \quad \inf_{q_b > 0}\left(\begin{array}{c} q_0 + q_b + (M - c)d \\ + \frac{(M-c)^2}{2(\beta c p(q_0 + q_b) - \Delta)} \end{array}\right)$$

$$l \quad = \quad \frac{\Delta(1 + \beta d p(q_{max})) - (\beta p(q_{max}))^2 M d}{\beta p(q_{max})}$$

If $0 \leq \theta(t) < c$ for all $t \leq 0$, and $l \geq 0$, then,

$$0 \leq l \leq x(t) \leq M, \qquad \forall t \geq 0,$$

$$0 \leq q(t) \leq q_{max}, \qquad \forall t \geq 0.$$

Proof. Let us first define the following epochs. Let t_0 be the first time after 0 such that $x(t)$ crosses c, t_1 be the first time after t_0 such that $x(t)$ is decreasing, t_2 be the first time after t_1 for $x(t)$ to cross c and t_3 be the first time after t_2 for $x(t)$ to increase. These times are illustrated in Figure 1. Formally, we can define these epochs by

$$t_0 \quad = \quad \inf\{t > 0 : x(t) > c\}$$
$$t_1 \quad = \quad \inf\{t > t_0 : \dot{x}(t) < 0\}$$
$$t_2 \quad = \quad \inf\{t > t_1 : x(t) < c\}$$
$$t_3 \quad = \quad \inf\{t > t_2 : \dot{x}(t) > 0\}$$
$$t_4 \quad = \quad \inf\{t > t_3 : x(t) > c\}$$

Finally, let $\tilde{t}_1 = t_1 - d$. We observe that as $q(t) = 0$ up to t_0, $x(t)$ is increasing over $[0, t_0]$ and will continue to increase up to t_1. Further, $x(t_1)$ is the maximum $x(t)$ will grow to in $[0, t_4]$. We will hereafter refer to the interval $[0, t_4]$ as a cycle. In this cycle, the minimum will be $x(t_3)$. Further, over this cycle, the queue-length process will be increasing on $[t_0, t_2]$ (i.e., when ever $x(t) > c$) will achieve its maximum at t_2. We will derive upper bounds for $x(t_1)$ and $q(t_2)$ and we will derive a lower bound for $x(t_3)$. Then, using induction, we can "propagate" the bounds for all times.

We first derive a bound on $x(t_1)$, i..e, we want to find a $M > c$ such that if $x(t_1) = M$, we must have $\dot{x}(t_1) \leq 0$. Now, as $x(t) \geq 0$ on $[-d, t_1]$, we have over this interval $\dot{x}(t) \leq \Delta$. Thus, at $\tilde{t}_1 = t_1 - d$, we have

$$M - \Delta d \leq x(\tilde{t}_1) \leq M \tag{50}$$

As $\dot{x}(t) \leq \Delta$, $x(t_0) = c$, and $x(\tilde{t}_1) \geq M - \Delta d$, we have

$$\tilde{t}_1 - t_0 \quad \geq \quad \frac{M - \Delta d - c}{\Delta} \tag{51}$$

Finally, from Figure 2, we see that $q(\tilde{t}_1)$ (given by the area of the shaded region) can be lower-bounded by the area of the triangle given by the dotted line, i.e.,

$$q(\tilde{t}_1) \quad \geq \quad \frac{1}{2}(\tilde{t}_1 - t_0)(M - \Delta d - c)$$

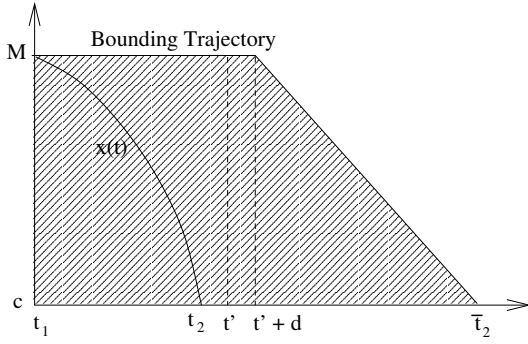$$\geq \quad \frac{(M - \Delta d - c)^2}{2\Delta} \tag{52}$$

Fig. 3. Trajectory of the rate process $x(t)$ and the upper-bounding process. The shaded area, along with the growth in the interval $[t_0, t_1]$ corresponds to q_{max}.

where the last step follows from (51). Thus, from (50) and (52), it follows that $\dot{x}(t_1) \leq 0$ if M is chosen to satisfy

$$\Delta \leq \beta(M - \Delta d)p\left(\frac{(M - \Delta d - c)^2}{2\Delta}\right) \quad (53)$$

Thus, we have over this cycle, $x(t) \leq M$. We next derive an upper-bound on $q(t_2)$ (and will implicitly derive an upper bound on $t_2 - t_1$).

As a trivial bound (see Figure 3), we assume that $x(t)$ *does not* decrease for some time (say up to t') after t_1. As a result, $q(t)$ will continue to increase (at rate $M - c$). Say t' is chosen such that $q(t') = q_b + q_0$ for some $q_b > 0$. Then, from $t' + d$, $x(t)$ will decrease till it hits c, *at least* as fast as $\beta c p(q_0 + q_b) - \Delta$, and will hit c eventually. An upper bound to this hitting time, denoted by $\bar{t}_2$ is given by $\bar{t}_2 - t' - d = \frac{M-c}{\beta c p(q_0 + q_b) - \Delta}$. During this interval, an upper bound to the additional amount the queue will grow is given by $(M - c)d + 0.5(\bar{t}_2 - t' - d)(M - c)$ (the area of a triangle with side $M - c$ and base $\bar{t}_2 - t' - d$), i.e., $(M-c)d + \frac{(M-c)^2}{2(\beta c p(q_0 + q_b) - \Delta)}$. As q_b is arbitrary, we have the upper-bound q_{max} given by

$$q_{max} \leq \inf_{q_b > 0} \left(\begin{array}{c} q_0 + q_b + (M-c)d \\ + \frac{(M-c)^2}{2(\beta c p(q_0 + q_b) - \Delta)} \end{array}\right) \quad (54)$$

This procedure is illustrated in Figure 3.

Finally, we derive a lower-bound on $x(t_3)$. Suppose $x(t_3) = l$. We want to choose l such that $\dot{x}(t_3) \geq 0$. Now, from (53) and (54), we have that

$$x(t_3 - d) \leq l + d(\beta M p(q_{max}) - \Delta)$$
$$q(t_3 - d) \leq q_{max}$$

Thus, we have

$$\dot{x}(t_3) \geq \Delta - \beta(l + d(\beta M p(q_{max}) - \Delta))p(q_{max})$$

Thus, choosing l such that

$$l = \frac{\Delta(1 + \beta d p(q_{max})) - (\beta p(q_{max}))^2 M d}{\beta p(q_{max})} \quad (55)$$

we have the required lower bound. Now, by assumption, $l > 0$ (this will be the case for typical marking functions). Then, over this cycle (i.e., $[0, t_4]$), we have $x(t) \geq 0$. Thus, a similar analysis can be carried out over the next cycle for deriving

the upper and lower bounds. Note that M, l derived over this cycle will work over the next cycle too as we assumed a "worst-case" behavior for the first cycle analysis. Formally, we can now setup an induction (over cycles) and prove that M, l are suitable upper and lower bounds for $x(t)$. As this step is straightforward, we shall skip the details in this paper. ■

Analogous to Section IV-B, we now assume that

$$\beta p(q_{max})d < 1 \quad (56)$$

We now state the following result without proof since it is identical to that of Theorem 4.2.

Theorem 6.1: Suppose M, β, d satisfy (56 and $l > K$. Then,
 (i) For all $t \geq 0$, $|x(t) - y^i(t)| \leq |x(0) - y(0)|$.
 (ii) If the initial condition $y^i(0)$ is non-negative, then $y(t) \geq 0$.
 (iii) For any $\epsilon > 0$, there is a $\tau < \infty$ such that for all $t \geq \tau$, $|x(t) - y^i(t)| < \epsilon$.
 ■

As in Section IV-B, the implication of this theorem is that for n large enough, the trajectories of $y^{i,n}(t)$ and $x^n(t)$ will remain non-negative. Further, by choosing $T = \tau$ (defined in (iii)), and choosing n large enough, it follows that $|x^n(T) - y^{i,n}(T)| < 2\epsilon$.

We comment that in the queue-based marking scheme, an extension to infinite time seems difficult. This difficulty arises from the fact that the queue-length process is not a Lipschitz continuous functional (in the sup topology) of the arrival process over an unbounded interval of time. Under weaker topologies on the space of trajectories, where the queue-length process is a Lipschitz continuous map, it is not clear that we can prove convergence results for functional differential equations.

VII. ADAPTIVE MARKING, MULTIPLE DELAYS AND GENERAL NETWORK TOPOLOGY

So far, we have considered a single link case where all the flows share the same round trip delay. We now generalize the above framework to include the more than one round trip delay, and more general network topologies.

A. Adaptive Marking

Using the vector framework from the previous section, we now consider adaptive marking functions. Marking functions we considered so far were designed so that for each n, the equilibrium point was $c - a$ per flow. Recall that for this to happen, we assumed that for the nth system (see Section IV-A), the marking function $p^n(.)$ satisfied

$$p^n(nx) = p(x) \quad (57)$$

For example, if we consider a marking function of the form

$$p(x) = \frac{(x - \tilde{c})^+}{x}$$

Then, for the nth system, we need to choose $p^n(\lambda)$ to be

$$p^n(\lambda) = \frac{(\lambda - \tilde{C}^n)^+}{\lambda}$$

with $\tilde{C}^n = n\tilde{c}$, which we denote as the *virtual capacity* of the link. Thus, we explicitly need to know n to design the marking function, which is hard to implement in practice. In [5], the authors have proposed *adaptive marking functions* where they have shown that even without the knowledge of n, marking functions can be designed to satisfy (57). We will illustrate this with the example considered above. Suppose we let $\tilde{C}^n$ adapt based on the total arrival rate $X^n(t) = nx^n(t) + na + ne^n(t)$, which consists of the sum of the controlled and the uncontrolled arrival rates over all flows, and the total link capacity $C^n = nc$ in the following manner. Fix some $\alpha > 0$, and let

$$
\begin{aligned}
\tilde{C}^n(t) &= \alpha\,(C^n - X^n(t)) \\
p^n(\lambda) &= \frac{(\lambda - \tilde{C}^n)^+}{\lambda}
\end{aligned}
$$

The equilibrium state of this is then $X^n(t) \to C^n$. Thus, in the absence of noise variations (i.e., $e^n(t) = 0$), we have $x^n(t) \to c - a$ which is the desired equilibrium point. As in the previous sections, we consider the aggregate flow $x^n(t)$ and the scaled virtual capacity $\tilde{c}^n(t)$. With delays and "noise", the evolution of this pair can be given by

$$
\begin{aligned}
\dot{x}^n(t) &= \Delta - \beta x^n\left(\frac{\lfloor n(t-d)\rfloor}{n}\right) \\
&\quad p\left(\begin{array}{l} x^n(\frac{\lfloor n(t-d)\rfloor}{n}) + a + e^n(\frac{\lfloor n(t-d)\rfloor}{n}), \\ \tilde{c}^n(\frac{\lfloor n(t-d)\rfloor}{n}) \end{array} \right) \\
\dot{\tilde{c}}^n(t) &= \alpha\left(c - \left(x^n(\frac{\lfloor nt\rfloor}{n}) + a + e^n(\frac{\lfloor nt\rfloor}{n}) \right) \right)
\end{aligned}
$$

where $p(x, \tilde{c}) = \frac{(x-\tilde{c})^+}{x}$. We can show using the vector framework used in the past section (treating $(x^n(t), \tilde{c}^n(t))$ as a vector process) that the above pair of equations converge as $n \to \infty$ to

$$
\begin{aligned}
\dot{x}(t) &= \Delta - \beta x^n(t-d) \\
&\quad p\,(x(t-d) + a, \tilde{c}(t-d)) \\
\dot{\tilde{c}}(t) &= \alpha\,(c - a - x(t)) \quad\quad\quad (58)
\end{aligned}
$$

As before, we have studied a relaxed problem where the non-negativity constraints on the trajectories were ignored. To prove that the trajectories are indeed non-negative, we need to study the limiting process (58) and show that the trajectories of this system is non-negative. A study of this system is hard even without delays (see [5]). With delays, we do not know if the trajectories will remain non-negative. We conjecture that for initial conditions close enough to the equilibrium point, and small enough delays, the trajectories will remain non-negative. Thus, the results for adaptive marking hold only for the relaxed problem, where we do not constrain the rate processes to be non-negative.

B. Multiple Round-Trip Delays

Let $\mathcal{D} = \{d_1, d_2, \dots, d_J\}$ be a set non-negative integers, and $\mathcal{D}$ be the set of possible *one-way* propagation delays that any flow can have. We say that a collection of flows are of the same *delay-class* if they all share the same one-way propagation delay. In this section, we will study the case

where there are n flows of each delay class, and derive the appropriate limit as $n \to \infty$.

Consider $\{y^n_{i,j}(t), 1 \le i \le n, 1 \le j \le J\}$ with $t \in [-T, T]$, where $y^n_{i,j}(t)$ be the transmission rate of the ith flow of the jth class at time t. We assume that associated with each data flow $y^n_{i,j}(t)$, there is a bounded, stationary ergodic "noise" process $a + e^n_{i,j}(t)$, with $E(e^n_{i,j}(t)) = 0$. Analogous to the previous sections, the system evolves in discrete time steps of size $\frac{1}{n}$, and a straight line interpolation is used in-between these time steps to embed the process in continuous time. The source rate at time t (with $nt \in calZ$) depends on the amount of flow marked half a round-trip time back. This amount in-turn, depends on the the amount of data sent by each flow, a further half a round-trip time back. Thus, the evolution of the rates $\{y^n_{i,j}(t), 1 \le i \le n, 1 \le j \le J\}$ can be described by

$$
\begin{aligned}
\dot{y}^n_{i,j}(t) &= \Delta_j - \beta\, y^n_{i,j}\left(\frac{\lfloor n(t-2d_j)\rfloor}{n}\right) \quad\quad (59) \\
&\quad p\left(\frac{1}{n}\sum_{k=1}^{J}\sum_{l=1}^{n} \left(\begin{array}{l} y^n_{l,k}(\frac{\lfloor n(t-d_j-d_k)\rfloor}{n}) \\ +e^n_{l,k}(\frac{\lfloor n(t-d_j-d_k)\rfloor}{n}) \end{array} \right) + Ja \right)
\end{aligned}
$$

with suitable initial conditions. Here, Δ_j is the additive increase factor, which could be different for each delay-class. Now, let

$$
\begin{aligned}
x^n_j(t) &= \frac{1}{n}\sum_{i=1}^{n} y^n_{i,j}(t) \\
e^n_j(t) &= \frac{1}{n}\sum_{i=1}^{n} e^n_{i,j}(t)
\end{aligned}
$$

Then, it follows from (59) that the evolution of the aggregate rate of each *delay-class* can be described by

$$
\begin{aligned}
\dot{x}^n_j(t) &= \Delta_j - \beta\, x^n_j\left(\frac{\lfloor n(t-2d_j)\rfloor}{n}\right) \quad\quad (60) \\
&\quad p\left(\sum_{k=1}^{J} \left(\begin{array}{l} x^n_k(\frac{\lfloor n(t-d_j-d_k)\rfloor}{n}) \\ +e^n_k(\frac{\lfloor n(t-d_j-d_k)\rfloor}{n}) \end{array} \right) + Ja \right)
\end{aligned}
$$

for each $j = 1, 2, \dots, J$, and with initial conditions given by $\theta_j(t) + e^n_j(0)$, for $nt \in \mathcal{Z}$ and linearly interpolated in-between. Note that in this framework, each class has a possibly different mean initial trajectory.

We can easily generalize Theorem 3.1 to include convergence of finite-dimensional vector processes, i.e., we can let $\mathbf{x}^n(t), \mathbf{e}^n(t)$ to be vector processes of dimension J, and let $b_n(., .), b(., .)$ map vector processes to $\mathcal{R}^J$. With similar assumptions as before (i.e., the functions are bounded and Lipschitz continuous), we can derive Theorem 3.1 in this more general setting. We define

$$
\begin{aligned}
x^n(t) &= [x^n_1(t), \dots, x^n_J(t)]' \\
e^n(t) &= [e^n_1(t), \dots, e^n_J(t)]'
\end{aligned}
$$

As in Section IV-A, we can show that the trajectories of (60) are bounded, and we can rewrite (60) as

$$
x^n(t) = b(x^n(t), e^n(t))
$$

where $b(., .) : C([0, T], \mathcal{R}^{2J}) \mapsto \mathcal{R}^J$, is a bounded, Lipschitz continuous functional. The details of it's construction is

analogous to that in Section IV-A, where the scalar case was considered. Thus, it can be shown that the appropriate limit to consider in the multiple delay case is given by the following coupled delay-differential equations

$$\dot{x}_j(t) = \Delta_j - \beta x_j(t - 2d_j) p\left(\sum_{k=1}^{J} x_k(t - d_j - d_k) + Ja\right)$$

with initial conditions $\{\theta_j(t), -T \leq t \leq 0, j = 1, 2, \ldots, J\}$. Analogous to Sections V and VI, we can generalize the results of this section to include random marking and queue-based marking.

We finally comment that non-negativity of the trajectories can be established when the number of delay classes are small. To show this, as a worst case, we can assume that the congestion is caused due to a single class. Then we can derive an upper bound on how large each class can grow using methods identical to that in Section IV-B. Using these upper bounds, we can construct lower bounds on the trajectories. This procedure is very similar to the $O(N^2)$ bounds constructed in [6]. In this case, we can show that bounds are $O(|\mathcal{J}|)$. Thus, as we increase $|\mathcal{J}|$, the lower bound will become negative.

C. Arbitrary Network Topology

So far, we have studied the congestion control problem with a single router. We now study the case where we have a more general network topology. We assume that there are $|\mathcal{R}|$ routers (resources) in the network, and they are indexed by $\mathcal{R} = \{1, 2, \ldots, R\}$. We assume that there are $|\mathcal{L}|$ classes of traffic flows, and they are indexed by $\mathcal{L} = \{1, 2, \ldots, L\}$. Each class consists of a collection of flows sharing the same path through the network.

For example, consider the UIUC domain. We will model the interior of the domain as a network with unlimited resources, and thus, no congestion occurs there. The only place where congestion could occur is at the edge router, which connects this domain to the Internet. Similarly, we consider various domains, and assume that their interiors are well-provisioned, and cause no congestion. The model we study would correspond to such a scenario, with many flows between each domain.

Let us denote $d_1(l, r)$ to be the forward-path delay from a source of class l to the resource (router) r. Let $d_2(l, r)$ be the reverse-path delay from resource r to the source of class l, with $d_1(l, r), d_2(l, r)$ being integers, and $D_l = d_1(l, r) + d_2(l, r)$ being the round-trip delay of any flow of class l. Further, let $\mathcal{R}_l$ be the collection of resources along the path of any flow of class l, and $\mathcal{L}_r$ be the collection of flow-classes passing through resource r.

We assume that there are n flows in each class, and the system evolves in discrete time-steps of $\frac{1}{n}$. As before, we have a collection of uncontrolled flows, iid, stationary and ergodic. We denote the source transmission rate for the ith flow of class l, when there are n flows in each class by $y_{i,l}^n(t)$ for $t \in [-T, T]$, with $nt \in \mathcal{Z}$. The source additively increases its rate and backs off proportional to the aggregate amount of marks it receives due to congestion at resources on its path.

We assume router (resource) r uses a congestion cost function $p_r(\cdot)$. If $p_r(\cdot)$ is interpreted as a marking function, then it is often assumed that the network operates in the rare negative feedback regime (see [7], [9]) and thus, the end-to-end fraction of marked packets is approximated by the sum of the fraction marked at each link along the path of the flow. However, this approximation is not necessary if we interpret $p_r(\cdot)$ as a congestion cost. For notational ease, we will assume that the marking functions are rate based, deterministic functions. We could easily generalize to random marking or/and queue-based marking also, but for ease of exposition, we will avoid that setup in this paper. Then, we can describe the evolution of $y_{i,l}^n(t)$ by

$$\dot{y}_{i,l}^n(t) = \Delta_l - \beta \, y_{i,l}^n\left(\frac{\lfloor n(t - D_l)\rfloor}{n}\right)$$

$$\sum_{k \in \mathcal{R}_l} p_k \left(\begin{array}{c} \frac{1}{n}\sum_{\substack{j \in \mathcal{L}_k \\ 1 \leq v \leq n}} y_{v,j}^n\left(\frac{\lfloor n(t - d_1(j,k) - d_2(l,k))\rfloor}{n}\right) \\ + e_{v,j}^n\left(\frac{\lfloor n(t - d_1(j,k) - d_2(l,k))\rfloor}{n}\right) \\ + |\mathcal{L}_k|a \end{array} \right)$$

with suitable initial conditions. As in the previous section, we can study the evolution of the aggregate rate of each class, and this process is described by

$$\dot{x}_l^n(t) = \Delta_l - \beta \, x_l^n\left(\frac{\lfloor n(t - D_l)\rfloor}{n}\right)$$

$$\sum_{k \in \mathcal{R}_l} p_k \left(\begin{array}{c} \sum_{j \in \mathcal{L}_k}\left(\begin{array}{c} x_j^n\left(\frac{\lfloor n(t - d_1(j,k) - d_2(l,k))\rfloor}{n}\right) \\ + e_j^n\left(\frac{\lfloor n(t - d_1(j,k) - d_2(l,k))\rfloor}{n}\right) \end{array} \right) \\ + |\mathcal{L}_k|a \end{array} \right)$$

for each $l = 1, 2, \ldots, L$, and with initial conditions given by $\theta_l(t) + e_l^n(0)$, for $nt \in \mathcal{Z}$ and linearly interpolated in-between. We can show that the appropriate fluid limit to study in this framework is described by

$$\dot{x}_l(t) = \Delta_l - \beta \, x_l(t - D_l)a \qquad (61)$$
$$\sum_{k \in \mathcal{R}_l} p_k \left(\begin{array}{c} \sum_{j \in \mathcal{L}_k} x_j(t - d_1(j,k) - d_2(l,k)) \\ + |\mathcal{L}_k|a \end{array} \right)$$

for each $l = 1, 2, \ldots, L$, and with initial conditions given by $\theta_l(t)$. The proof follows a vector formulation of Theorem 3.1 (see Section VII-B and Remark 3.1 for a description of this). The assumptions of Lipschitz continuity and boundedness of the trajectories over any finite interval of time follows from the fact that for any $\mathcal{R}_l \subset \mathcal{R}$, we have

$$\sum_{k \in \mathcal{R}_l} p_k(.) \leq R$$

Thus, a proof analogous to that in Theorem 4.1 can be used to establish the required result.

As in the previous section, we would like to establish the non-negativity of the trajectories. However, the bounds used to establish non-negativity become less tight as the number of classes and the number of links increase. Thus, establishing non-negativity of the trajectories for a general topology network is still an open issue.

VIII. Dual Algorithms: Convergence with Delays

So far, we have have studied primal algorithms based on the framework developed in [3]. We next study differential equation models derived from a dual approach to the congestion control problem [25]. As in the primal case, we restrict our study to the log utility function.

Analogous to the primal problem with an arbitrary network topology (Section VII-C), we study a system consisting of $|\mathcal{R}|$ routers (resources) in the network, which are indexed by $\mathcal{R} = \{1, 2, \ldots, R\}$. We assume that there are $|\mathcal{L}|$ *classes* of traffic flows, and they are indexed by $\mathcal{L} = \{1, 2, \ldots, L\}$. Each class consists of a collection of flows sharing the same path through the network.

As before, let us denote $d_1(l, r)$ to be the forward-path delay from a source of class l to the resource (router) r. Let $d_2(l, r)$ be the reverse-path delay from resource r to the source of class l, with $d_1(l, r), d_2(l, r)$ being integers, and $D_{l,r} = d_1(l, r) + d_2(l, r)$ being the round-trip delay of any flow of class l to resource r. Further, let $\mathcal{R}_l$ be the collection of resources along the path of any flow of class l, and $\mathcal{L}_r$ be the collection of flow-classes passing through resource r.

We assume that there are n flows in each class, and the system evolves in discrete time. As before, we have a collection of uncontrolled flows, iid, stationary and ergodic and with mean $n|\mathcal{L}_r|a$. We denote this collection by $\{a + \tilde{e}_{i,l}^n(k)\}, 1 \le i \le n, l \in \mathcal{L}$, where $\tilde{e}_{i,l}^n(k)$ are zero mean. We assume that the resource capacity is scaled as $n|\mathcal{L}_r|(c_r + a)$. We denote the source transmission rate for the ith flow of class l, when there are n flows in each class. Now, from the dual formulation, we have that the each resource updates its link cost (marking function) based on the total arrival rate to it, and each user is fed back the sum of the link costs along its path. We study the system in discrete time, the time index being $k \in \{0, 1 \ldots, nT\}$. We denote the link cost (marking function) at discrete time k at resource r by $p_r(k)$, and the date rate of user i of class l at time k by $y_{i,l}^n(k)$. Then, for any $r \in \mathcal{R}$, the update algorithms can be written as

$$p_r^n(k+1) = p_r^n(k) - \gamma^{(n)} n|\mathcal{L}_r| c_r$$

$$+ \gamma^{(n)} \left(\min \left(\sum_{l \in \mathcal{L}_r} \sum_{i=1}^{n} (y_{i,l}^n(k) + a + \tilde{e}_{i,l}^n(k)), Mn|\mathcal{L}_r| \right) \right)^+$$

$$y_{i,l}^n(k) = \frac{\kappa_l}{\sum_{r \in \mathcal{R}_l} p_r^n(k - nD_{l,r})} \quad 1 \le i \le n, \ l \in \mathcal{L}$$

where $\gamma^{(n)}$ is the update "gain" for the nth system. We will later see that this is of the order $\frac{1}{n}$ after embedding in continuous time. For technical reasons, we assume that there is a large enough M such that if the total input rate exceeds $Mn|\mathcal{L}_r|$, then the link will react as though the rate is $Mn|\mathcal{L}_r|$. Further, if the total rate goes below 0, then the link will react as though the rate is 0. As before, we will denote $[z] = 0 \bigvee (z \bigwedge M)$.

Now, let $\tilde{p}_r^n(k)$ be the *price per user* at link r, and $x_l^n(k)$ be the average rate at time k over flows of class l, i.e.,

$$\tilde{p}_r^n(k) = \frac{1}{n|\mathcal{L}_r|} p_r^n(k) \quad r \in \mathcal{R}$$

$$x_l^n(k) = \frac{1}{n} \sum_{i=1}^{n} y_{i,l}^n(k) \quad l \in \mathcal{L}$$

$$e_l^n(k) = \frac{1}{n} \tilde{e}_{i,l}^n(k) \quad l \in \mathcal{L}$$

Thus, for $k \in \{0, \ldots, nT\}$ and each $r \in \mathcal{R}$, the evolution of the average cost at each resource follows

$$
\begin{aligned}
\tilde{p}_r^n(k+1) &= \tilde{p}_r^n(k) - \gamma^{(n)}(c_r + a) \\
&\quad + \gamma^{(n)} \left[\frac{1}{n\mathcal{L}_r} \left(\sum_{\substack{l \in \mathcal{L}_r \\ i \le i \le n}} y_{i,l}^n(k) + a + e_{i,l}^n(k) \right) \right] \\
&= \tilde{p}_r^n(k) - \gamma^{(n)}(c_r + a) \\
&\quad + \gamma^{(n)} \left[\frac{1}{|\mathcal{L}_r|} \left(\sum_{l \in \mathcal{L}_r} x_l^n(k) + \sum_{l \in \mathcal{L}_r} e_l^n(t) \right) + a \right]
\end{aligned}
$$

and

$$x_l^n(k) = \frac{\kappa_l}{\sum_{r \in \mathcal{R}_l} \tilde{p}_r^n(k - nD_{l,r})} \quad l \in \mathcal{L}.$$

Thus, we have for any $r \in \mathcal{R}$,

$$\tilde{p}_r^n(k+1) = \tilde{p}_r^n(k) - \gamma^{(n)}(c_r + a)$$

$$+ \gamma^{(n)} \left[\frac{1}{|\mathcal{L}_r|} \left(\sum_{l \in \mathcal{L}_r} \frac{\kappa_l}{\sum_{r \in \mathcal{R}_l} \tilde{p}_r^n(k - nD_{l,r})} + \sum_{l \in \mathcal{L}_r} e_l^n(k) \right) + a \right]$$

Embedding this equation in in continuous time (each discrete time-step corresponds to an interval to time of length of $\frac{1}{n}$), we see that as the step size is order $\frac{1}{n}$, the gain $\gamma^{(n)}$ should be scaled as well, i.e., for some $\gamma > 0$, $\gamma^{(n)} = \gamma/n$. Thus, in a differential equation form (as in the primal algorithms considered earlier), we have for $t \in [0, T]$, for any $r \in \mathcal{R}$,

$$\dot{\tilde{p}}_r^n(t) = -\gamma(c_r + a) \tag{62}$$

$$+ \gamma \left[\frac{1}{|\mathcal{L}_r|} \left(\sum_{l \in \mathcal{L}_r} \frac{\kappa_l}{\sum_{r \in \mathcal{R}_l} \tilde{p}_r^n(t - D_{l,r})} + \sum_{l \in \mathcal{L}_r} e_l^n(t) \right) + a \right]$$

Finally, as $\tilde{p}(t)$ is Lipschitz continuous (as the derivative is bounded by $L = \gamma(M + c + a)$, where $c = \max_{r \in \mathcal{R}} c_r$) and $\tilde{p}(t)$ is bounded over $[0, T]$ (the bound being LT), we have that as $n \to \infty$, from a vector formulation of Theorem 3.1, $\sup_{t \in [0,T]} \max_{r \in \mathcal{R}} |\tilde{p}_r^n(t) - \tilde{p}_r(t)| \to 0$, where $\tilde{p}_r(t)$ is defined by

$$\dot{\tilde{p}}_r(t) = \gamma \left(\begin{array}{c} \left[\frac{1}{|\mathcal{L}_r|} \left(\sum_{l \in \mathcal{L}_r} \frac{\kappa_l}{\sum_{r \in \mathcal{R}_l} \tilde{p}_r(t - D_{l,r})} \right) + a \right] \\ -(c_r + a) \end{array} \right)$$

A. Non-negativity of the Trajectory: A Fluid Limit Analysis

So far, we have neglected the constraint that $p_r(.)$ should be non-negative. We will now consider a special case of the above system, where we have only one router and a single delay-class. We will show that for γ small enough, the trajectory will remain non-negative.

We consider the delay-differential equation

$$\dot{p}(t) = \gamma \left(\left[\frac{\kappa}{\tilde{p}(t-d)} + a \right] - (c+a) \right) \qquad (63)$$

with suitable non-negative initial conditions.

Lemma 8.1: There exists some $\delta > 0$ such that $\forall 0 < \gamma \leq \delta$, the trajectory of (63) remains non-negative for all $t \in [0, \infty)$.

Proof. It is sufficient to show that $\exists l > 0$ such that if at any time $t > d$, $\tilde{p}(t) < l$, $\dot{p}(t) > 0$. Thus, if the initial condition is "reasonable", the trajectory will remain non-negative over $[0, d]$ and we will be done.

Suppose for some $l > 0$, $\tilde{p}(t) < l$. From (63), it follows that $\tilde{p}(t-d) < l + \gamma(c+a)d$. Thus, we have

$$\dot{p}(t) = \gamma \left(\left[\frac{\kappa}{\tilde{p}(t-d)} + a \right] - (c+a) \right)$$
$$> \gamma \left(\left[\frac{\kappa}{l + \gamma(c+a)d} + a \right] - (c+a) \right)$$

Now, recall that $[z] = 0 \bigvee (z \bigwedge M)$. By choosing M arbitrarily large (but finite), we can ensure that

$$\left[\frac{\kappa}{l + \gamma(c+a)d} + a \right] = \frac{\kappa}{l + \gamma(c+a)d} + a$$

Thus, we have that

$$\dot{p}(t) > \frac{\kappa}{l + \gamma(c+a)d} + a - (c+a)$$
$$= \frac{\kappa}{l + \gamma(c+a)d} - c$$

Thus, to ensure $\dot{p}(t) > 0$, we must have

$$\frac{\kappa}{l + \gamma(c+a)d} > c \qquad (64)$$

Also, note that the equilibrium point p^* of (63) satisfies (again M is chosen large enough so that the ceiling does not matter)

$$p^* = \frac{\kappa}{c}$$

Thus, from (64), it follows that choosing any l satisfying

$$l < p^* - \gamma(c+a)d$$

ensures that $\dot{p}(t) > 0$. Now, choosing $\gamma > 0$ small enough, we can ensure that we can choose $l > 0$. Thus, we are done. ∎

Finally, for n, the number of flows large enough, it now follows from the uniform convergence result in the last section that the trajectory of the stochastic system (for a single class case) will be non-negative as well.

REFERENCES

[1] K. K. Ramakrishnan and S. Floyd, "A proposal to add explicit congestion notification (ECN) to IP," RFC 2481, January 1999.

[2] S. Kunniyur and R. Srikant, "End-to-end congestion control: utility functions, random losses and ECN marks," in *Proceedings of IEEE Infocom*, Tel Aviv, Israel, March 2000.

[3] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1998.

[4] ——, "Mathematical modeling of the Internet," in *Mathematics Unlimited - 2001 and Beyond (Editors B. Engquist and W. Schmid).* Berlin: Springer-Verlag, 2001, pp. 685–702.

[5] S. Kunniyur and R. Srikant, "A time-scale decomposition approach to adaptive ECN marking," *IEEE Transactions on Automatic Control*, vol. 47, no. 6, pp. 882–894, June 2002.

[6] S. Shakkottai, R. Srikant, and S. Meyn, "Bounds on the throughput of congestion controllers in the presence of feedback delay," *To appear in the IEEE/ACM Transactions on Networking*, December 2003, a shorter version appeared in the Proceedings of the IEEE Conference on Decision and Control, Orlando, FL, December 2001.

[7] M. Vojnovic, J. Le Boudec, and C. Boutremans, "Global fairness of additive-increase and multiplicative-decrease with heterogeneous round-trip times," in *Proceedings of IEEE Infocom 2000*, Tel-Aviv, Israel, March 2000.

[8] H. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications.* Springer-Verlag, 1997.

[9] R. Johari and D. Tan, "End-to-end congestion control for the Internet: Delays and stability," *IEEE/ACM Transactions on Networking*, vol. 9, no. 6, pp. 818–832, December 2001.

[10] L. Massoulie, "Stability of distributed congestion control with heterogeneous feedback delays," *IEEE Transactions on Automatic Control*, June 2002.

[11] G. Vinnicombe, "On the stability of end-to-end congestion control for the Internet," 2001, university of Cambridge Technical Report.

[12] S. Kunniyur and R. Srikant, "Analysis and design of an adaptive virtual queue algorithm for active queue management," in *Proceedings of ACM Sigcomm*, San Diego, CA, August 2001, pp. 123–134.

[13] T. Kurtz, "Solutions of ordinary differential equations as limits of pure jump Markov processes," *Journal of Applied Probability*, vol. 7, pp. 49–58, 1970.

[14] H. Khalil, *Nonlinear Systems.* Upper Saddle River, NJ: 2nd edition, Prentice Hall, 1996.

[15] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[16] S. H. Low and D. E. Lapsley, "Optimization flow control, I: Basic algorithm and convergence," *IEEE/ACM Transactions on Networking*, pp. 861–875, December 1999.

[17] H. Yaiche, R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 667–678, October 2000.

[18] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 397–413, August 1993.

[19] S. Athuraliya, D. E. Lapsley, and S. H. Low, "Random early marking for Internet congestion control," in *Proceedings of IEEE Globecom*, 1999.

[20] C. Hollot, V. Misra, D. Towsley, and W. Gong, "On designing improved controllers for AQM routers supporting TCP flows," in *Proceedings of IEEE Infocom*, Anchorage, AK, April 2001.

[21] J. Hale, *Theory of Functional Differential Equations.* Springer-Verlag, 1977.

[22] J. A. Bucklew, *Large Deviation Techniques In Decision, Simulation And Estimation.* New York: Wiley, 1990.

[23] V. Lakshmikantham and S. Leela, *Differential and Integral Inequalities: Theory and Applications, Vol. I and II.* New York, NY: Academic Press, 1969.

[24] J. Hale, "Asymptotic behavior of the solutions of differential-difference equations," in *Proceedings of Int. Symp. Nonlinear Vibrations*, vol. 2. IUTAM, 1963, pp. 409–426.

[25] F. Paganini, J. Doyle, and S. Low, "Scalable laws for stable network congestion control," in *Proceedings of the IEEE Conference on Decision and Control*, Orlando, FL, December 2001.

Sanjay Shakkottai (M'02) received his Ph.D. from the University of Illinois at Urbana-Champaign in 2002. He is currently with the University of Texas at Austin, where he is an Assistant Professor in the Department of Electrical and Computer Engineering. He was the finance chair of the 2002 IEEE Computer Communications Workshop in Santa Fe, NM. His current research interests include wireless and sensor networks, stochastic processes and queueing theory. His email address is shakkott@ece.utexas.edu.

R. Srikant (M '91, SM '01) received his B.Tech. from the Indian Institute of Technology, Madras in 1985, his M.S. and Ph.D. from the University of Illinois in 1988 and 1991, respectively, all in Electrical Engineering. He was a Member of Technical Staff at AT&T Bell Laboratories from 1991 to 1995. He is currently with the University of Illinois, where he is an Associate Professor in the Department of Electrical and Computer Engineering and a Research Associate Professor in the Coordinated Science Lab. He was an associate editor of Automatica, and is currently on the editorial boards of the IEEE/ACM Transactions on Networking and IEEE Transactions on Automatic Control. He was the chair of the 2002 IEEE Computer Communications Workshop in Santa Fe, NM. His research interests include communication networks, stochastic processes, queueing theory, information theory, and game theory. His email address is rsrikant@uiuc.edu.