Discrete-Time Analysis of Adaptive Rate Control Mechanisms

Eitan Altman[1], François Baccelli[1], and Jean-Chrysostome Bolot
INRIA, B.P. 93, 06902 Sophia Antipolis Cedex, France

Abstract

We analyze the performance of a generic feedback flow control mechanism which captures the properties of several such mechanisms recently proposed in the literature. These mechanisms dynamically regulate the rate of data flow into a network based on feedback information about the network state. They are used in a variety of networks and they have been advocated for upcoming high-speed networks. However, they are difficult to model realistically. In this paper, we present a stochastic discrete-time approach that yields models which are realistic and yet tractable and computationally easy to solve.

For our generic mechanism, the feedback consists of an exponentially averaged estimate of the bottleneck service rate and queue size. We obtain a model described by non-linear stochastic difference equations. We find the conditions under which these equations converge to a steady-state and we characterize the speed of convergence to steady-state. We then consider a linearized version of the model for which we can derive closed-form solutions. These solutions bring out a tradeoff between efficient steady-state behavior and rapid adaptability to varying network conditions. We indicate how a dynamic averaging technique of the feedback information can overcome this disadvantage. We also identify the cases when the linear model is a good approximation to the non-linear model.

Keyword Codes: C.2.0; C.2.1; C.4
Keywords: Computer-Communication Networks, General; Network Architecture and Design; Performance of Systems

1 Introduction

In a packet-switched network, packets generated by source nodes are routed via a sequence of intermediate nodes to destination nodes. A flow control mechanism limits source rates in order to avoid congestion in the intermediate nodes. The goal of the flow control mechanism is to match the source rate on a connection to the capacity available on this

[1]Supported by the European Grant BRA-QMIPS of CEC DG XIII.

connection, i.e. to the rate of the bottleneck node on the connection. The bottleneck rate changes with time because the number of connections routed through each intermediate node varies and because sources do not send packets at a constant rate. The flow control mechanism must be somehow informed of such changes. One way is for the source to receive feedback about the state of intermediate nodes, and to control the rate at which it sends data into the network accordingly.

In the Internet TCP, the feedback is implicit. There, the source uses timeouts and duplicate acknowledgements to detect buffer overflows in the intermediate nodes [11]. Other schemes rely on feedback information that includes bottleneck service rate [13], bottleneck queue size [8], round trip delays [1], etc.

It is natural to model feedback flow control mechanisms by queueing systems with delayed feedback where the service times and job arrivals are time-varying and state-dependent. Unfortunately, the analysis of such systems is difficult, and only simple models have been considered (e.g. [12, 25, 2]). These difficulties have led to the development of approximate techniques. One such technique is the fluid or flow approximation, where the basic idea is to replace stochastic processes by corresponding deterministic processes which are related by means of differential or difference equations [9, 4].

In this paper, we present a discrete-time stochastic approach to modeling flow control mechanisms that dynamically regulate the source based on network feedback. Our approach yields models that are realistic and yet tractable. The models are expressed in terms of stochastic difference equations. They appear more realistic than deterministic models [7, 9], and more tractable than continuous-time stochastic models [20].

The paper is organized as follows. In Section 2, we formulate a discrete-time stochastic model of a generic flow control mechanism. This mechanism captures the properties of several recently proposed flow control mechanisms such as the packet-pair mechanism [13] or the HBH mechanism [19]. In Section 3, we show that the flow control mechanism is unstable if the source sending rate is adjusted to match a smoothed estimate of the bottleneck service rate. Interestingly, a similar result was observed by others using simulation [15]. In Section 4, we show that the flow control mechanism is stable if the source sending rate is adjusted to match a smoothed quantity combining estimates of the bottleneck service rate and of the bottleneck queue size. This indicates that it is important to use both estimates in a real flow control mechanism. And indeed, we observe that the flow control mechanisms proposed in [13, 19] do use both these estimates. In Section 5, we linearize the model considered in Section 4 to obtain a model for which we can derive closed-form solutions in both steady-state and transient regimes. The solutions show that expected values of the source sending rate and the bottleneck queue size converge to their target values. However, the variances indicate a tradeoff between good steady-state behavior and rapid adaptability to changing network conditions. We indicate how a dynamic exponential averaging of the feedback information can overcome the limitations of this tradeoff. Section 6 concludes the paper.

2 A Discrete-Time Stochastic Model

In this section, we present a discrete-time stochastic model of a connection with a generic rate-based flow control mechanism. We assume that there is one bottleneck node in the sequence of nodes from the source to the destination. We divide time into steps. During each step, the source maintains estimates for the bottleneck service rate and the bottleneck queue size. At the end of each step, the source updates its sending rate using these estimates.

Both estimates can be obtained in a number of ways. In the HBH scheme, adjacent intermediate nodes periodically exchange control packets which include these estimates [19]. In the packet-pair scheme, sources send pairs of packets close together [14]. The bottleneck rate is estimated by taking an inverse measure of the time between receiving the acknowledgement packets for such pairs. The bottleneck queue size is estimated by taking an average of the roundtrip times measured during a step, subtracting out the propagation delay (this delay can in turn be estimated by taking the smallest roundtrip time measured since the beginning of the connection), and multiplying by the bottleneck rate estimate.

In this paper, we consider a control mechanism similar to the packet pair mechanism [13]. Then, each step represents one roundtrip delay. At the beginning of a step, the source sends a pair of specially-marked packets. The end of the step occurs when the source receives the acknowledgement packets for this pair. Let x_n denote the source rate during step n, and μ_n denote the bottleneck service rate during step n. Let q_{n+1} denote the average bottleneck queue size during the interval between the arrival at the bottleneck of the nth and the $(n+1)$th specially-marked pairs. Let d_n denote the duration of this interval. We have

$$q_{n+1} = \max(q_n + (x_n - \mu_n)d_n, 0)$$

Clearly, we expect d_n to be approximately equal to the duration of the nth step, i.e. the roundtrip delay of the nth specially-marked pair. For our analysis, we assume that d_n is constant and independent of n. Without loss of generality, we choose $d_n = 1$ for all n. Then the above equation can be rewritten as

$$q_{n+1} = \max(q_n + x_n - \mu_n, 0) \tag{1}$$

The bottleneck service rate μ_n changes between successive steps because the number of sources sharing the bottleneck queue varies as connections are set up and terminated, and because sources do not send at a constant rate. Previous models of feedback flow control mechanisms have typically considered two models for μ_n. In the first model, μ_n is represented as fluctuating around a nominal value. However, this model does not capture important features such as the positive correlation between μ_n and μ_{n+1} which exists when the number of connections sharing the bottleneck queue is large. This has led to the introduction of a model in which μ_n is a random walk [13]. However, the random walk model has its own drawbacks too. For example, the successive differences $\mu_{n+1} - \mu_n$ are uncorrelated. Furthermore, μ_n can become negative or arbitrarily large. In this paper, we model μ_n as a truncated ARMA process. Specifically, we let

$$\mu_n = \mu + \xi_n \tag{2}$$

$$\xi_n = \min\left\{\max\left(\sum_{i=1}^{d} l_i \xi_{n-i} + k\gamma_n, \underline{\mu} - \mu\right), \overline{\mu} - \mu\right\} \tag{3}$$

for $n \geq 0$, where γ_n are independent and identically distributed (i.i.d.) standard Gaussian random variables, k and l_i, $i = 1 .. d$, are constant, and $\xi_{-d}, ..., \xi_{-1}$ are given initial conditions. The truncation ensures that μ_n is bounded below by $\underline{\mu}$ and bounded above by $\overline{\mu}$. A similar model (although not truncated) has been used in [6] to design feedback flow control mechanisms. ARMA models have several advantages. For example, they can capture an extremely wide range of possible behavior [16]. They are analytically tractable (see Section 5). Furthermore, efficient methods have been developed to identify ARMA models, i.e. to find the values of k and l_i from a time series of measured values of μ_n [16].

We assume that, at the end of step n, the source obtains a noisy estimate of the value of μ_n, and a noisy estimate of the value of q_n. We use $\hat{\mu}_n$ to denote the estimate for the bottleneck rate, and $\hat{q}_n$ to denote the estimate for the bottleneck queue size. For our analysis, we assume that these estimates can be written as

$$\hat{\mu}_n = \mu_n + \lambda_{n+1} \tag{4}$$

$$\hat{q}_n = q_n + \omega_{n+1} \tag{5}$$

where λ_{n+1} and ω_{n+1} are i.i.d. Gaussian random variables. We assume that λ_{n+1} and ω_{n+1} are zero-mean processes, i.e. the estimates $\hat{\mu}_n$ and $\hat{q}_n$ are unbiased.

At the end of step n, the source updates its sending rate as follows

$$x_{n+1} = \max(x_n + \alpha(\hat{\mu}_n - x_n) - \beta(\hat{q}_n - Q), 0) \tag{6}$$

where $0 < \alpha$, $0 < \beta$, and $0 \leq Q$. The max operation ensures that $x_n \geq 0$ for all $n \geq 0$. Note that two terms are added to x_n in the right-hand side of equation (6). The first term is the usual exponential forgetting mechanism with a forgetting factor equal to α. If there were no second term (namely $\beta = 0$), then we refer to this flow control mechanism as a *pure rate-matching mechanism* [15]. It turns out that in this cases, q_n converges in law to $+\infty$ (i.e. for any $K > 0$ the probability $P[q_n < K] \to 0$ as $n \to \infty$) even though x_n converges to some stationary distribution. We prove this in Section 3 for the special case when $\xi_n = 0$, i.e. $\mu_n = \mu$ is constant, and under statistical assumptions specified there. The term $-\beta(\hat{q}_n - Q)$ in equation (6) ensures that the send rate stays below the estimated value of μ_n as long as the bottleneck queue size is above a target value Q.

Let δ_{n+1} be the i.i.d. Gaussian random variables defined by

$$\alpha\delta_{n+1} = \alpha\lambda_{n+1} - \beta\omega_{n+1} \tag{7}$$

Then equation (6) becomes

$$x_{n+1} = \max((1-\alpha)x_n + \alpha(\mu_n + \delta_{n+1}) - \beta(q_n - Q), 0) \tag{8}$$

Our objective is to analyze the stochastic dynamical system defined by equations (1, 2, 3, 8). The max operations in the equations make the system non linear and hence closed form solutions unlikely. Nevertheless, using recently developed techniques, it is possible to analyze the stability and convergence of the system. Specifically, we find in Section 4 that if $\alpha > 0$ and $\beta > 0$, then all the moments of x_n and q_n exist in steady-state. Furthermore, the transient moments of x_n and q_n converge geometrically fast to the steady-state moments. However, we cannot obtain closed-form expressions for the transient or steady-state moments. In Section 5, we consider a linearized version of the model defined by equations (1, 3, 8) for which we can obtain closed-form solutions.

3 Instability of the Pure Rate Matching Algorithm

In this section, we show that the pure rate matching flow control mechanism is unstable. By instability, we mean here that for all initial queue size conditions and for all finite constants C, $P[q_n < C]$ tends to 0 as n tends to infinity, namely the queue size tends to infinity in probability. We consider the special case in which the bottleneck service rate is constant $\mu_n = \mu$ and where the variable q_k given by

$$\begin{aligned} q_{k+1} &= \max(q_k + x_k - \mu, 0) \\ x_{k+1} &= (1-\alpha)x_k + \alpha(\mu + e_k) \end{aligned} \tag{9}$$

We assume that $\{e_k\}$ is centered, i.i.d. and admits an absolutely continuous component with a positive density. This assumption on $\{x_k\}$ is more specific than what was considered above, but the technique that we develop can be extended to more general frameworks. For instance, our result implies also the instability of the same system where (9) is replaced by $x_{k+1} = \max((1-\alpha)x_k + \alpha(\mu + e_k), 0)$; the latter modification ensures that the input rate cannot become negative. This conclusion follows from the fact that with this dynamics, the couple (q_k, x_k) dominates (i.e. it is greater than or equal to) the corresponding values obtained by the dynamics of (9) pathwise.

Throughout the rest of this section, we let $y_k = x_k - \mu$. We can rewrite the above equations under the following equivalent form:

$$q_{k+1} = \max(q_k + y_k, 0) \tag{10}$$

$$y_{k+1} = (1-\alpha)y_k + \alpha e_k \tag{11}$$

For sake of simplicity, we will assume that $q_0 = 0$ (if $q_0 > 0$, the system dominates the one with $q_0 = 0$; this dominance allows us to prove instability for all initial conditions if we prove it for $q_0 = 0$).

Note that under the above statistical assumptions, $\{y_k\}$ is a Harris recurrent ergodic Markov chain (see the following section for this terminology). Therefore, it has a unique stationary regime, which is reached with coupling (see [17], Chapter III). So, we shall assume that this chain is in its steady state (if it is not, the result below extends immediately as the chain couples with its unique stationary regime in finite time).

Our analysis will be based on the following relation which is obtained by an immediate induction argument:

$$q_k = \max(0, y_{k-1}, y_{k-1} + y_{k-2}, \ldots, y_{k-1} + \ldots + y_0), \quad \forall k \geq 0 \tag{12}$$

It can be shown from this relation that q_k converges in distribution to the random variable

$$Q = \sup(0, y_{-1}, y_{-1} + y_{-2}, \ldots, y_{-1} + \ldots + y_{-k}, \ldots) \tag{13}$$

(see [3] for a proof). So, it is enough to prove that the random variable Q is almost surely infinite to show that $P[q_k < C] \to 0$ as k goes to ∞ for all finite constant C, namely that the system is unstable.

Let Z_n, $n \geq 1$ be the random variable

$$Z_n = y_{-1} + \ldots + y_{-n} \tag{14}$$

We will need the following computational result:

Lemma 1 *For all $n \geq 2$,*

$$Z_n = y_{-n} \frac{1-(1-\alpha)^n}{\alpha} + \sum_{k=1}^{n-1} e_{-n+k-1} b_{n-k} \tag{15}$$

where

$$b_l = 1 - (1-\alpha)^l \tag{16}$$

Proof: From the very definition of $\{y_k\}$, we have $y_{-n+k+1} = (1-\alpha)y_{-n+k} + \alpha e_{-n+k}$. If (15) holds true for n, then

$$\begin{aligned} Z_{n+1} &= Z_n + y_{-n-1} \\ &= y_{-n} \frac{1-(1-\alpha)^n}{\alpha} + \sum_{k=1}^{n-1} e_{-n+k-1} b_{n-k}) + y_{-n-1} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1-(1-\alpha)^n}{\alpha}((1-\alpha)y_{-n-1}+\alpha e_{-n-1}) + \sum_{k=1}^{n-1} e_{-n+k-1}b_{n-k} + y_{-n-1} \\
&= y_{-n-1}\frac{1-(1-\alpha)^{n+1}}{\alpha} + \sum_{k=0}^{n-1} e_{-n+k-1}b_{n-k} \\
&= y_{-n-1}\frac{1-(1-\alpha)^{n+1}}{\alpha} + \sum_{k=1}^{n} e_{-n-1+k-1}b_{n+1-k} \qquad (17)
\end{aligned}
$$

and the proof is completed by induction. ■

Lemma 2 *Under the foregoing assumptions, if variance*$(e_1) > 0$*, then* $Q = \infty$ *a.s.*

Proof Let Q_n be the random variable

$$Q_n = \sup(0, Z_1, Z_2, \ldots, Z_n) \qquad (18)$$

Note that we have $Q = \lim_{n\to\infty} Q_n$ in the sense of the almost sure convergence (the sequence Q_n is nondecreasing). Clearly, $Q = \infty$ a.s. is equivalent to $Z = \limsup_{n\to\infty} Z_n = \infty$ a.s. It is easily checked that since the sequence $\{e_k\}$ is ergodic, then Z is constant (because it is shift-invariant). Assume that

$$P[Z \le C] = 1 \qquad (19)$$

where C is a finite constant. This can be rewritten as

$$P[\limsup_{n\to\infty} A_n] = \lim_{n\to\infty} P[\cup_{m\ge n} A_m] = 0 \qquad (20)$$

where A_n is the event $\{Z_n > C\}$. But

$$P[\bigcup_{m\ge n} A_m] \ge P[A_n] \qquad (21)$$

When using the independence between z_{-n} and $\{e_{-n+k}\}_{k\ge 0}$, we get from Lemma 1

$$
\begin{aligned}
P[A_n] &= P[y_{-n}\frac{1-(1-\alpha)^n}{\alpha} + \sum_{k=1}^{n-1} e_{-n+k-1}b_{n-k} > C] \\
&\ge P[\{y_{-n}\frac{1-(1-\alpha)^n}{\alpha} \ge 0\} \bigcap \{\sum_{k=1}^{n-1} e_{-n+k-1}b_{n-k} > C\}] \\
&= P[y_{-n}\frac{1-(1-\alpha)^n}{\alpha} \ge 0] P[\sum_{k=1}^{n-1} e_{-n+k-1}b_{n-k} > C]
\end{aligned}
$$

In view of the stationarity assumption on y_n, $P[y_{-n}\frac{1-(1-\alpha)^n}{\alpha} \geq 0] = P[y_n \geq 0] = q$, where q is a positive constant that does not depend on n. Similarly, we prove below that under the assumption on the variance of e_1,

$$\liminf_n P[\sum_{k=1}^{n-1} e_{-n+k-1} b_{n-k} > C] \geq \frac{1}{2} \tag{22}$$

Thus, $\liminf_n P[A_n] \geq q/2$, which precludes the property $\lim_n P[\cup_{m \geq n} A_m] = 0$ in view of (21).

So the proof will be concluded if we prove (22). We have

$$\sum_{k=1}^{n-1} e_{-n+k-1} b_{n-k} = \sum_{k=1}^{n-1} e_{-n+k-1} - v_{n-1}$$

where

$$v_n = \sum_{l=1}^{n} e_{-l-1}(1-\alpha)^l$$

It is easily checked that the random variable v_∞ is a.s. finite (the positive random variable $\sum_{l=1}^{\infty} |e_{-l-1}|(1-\alpha)^l$ is a.s. finite because its expectation is finite). Therefore

$$\lim_n \frac{1}{\sqrt{n}} v_n = 0 \quad \text{a.s.}$$

Thus

$$\lim_n P[\sum_{k=1}^{n} e_{-n+k-2} b_{n+1-k} > C] \quad = \quad \lim_n P[\frac{\sum_{k=1}^{n} e_{-n+k-2}}{\sqrt{n}} > \frac{C}{\sqrt{n}} + \frac{v_n}{\sqrt{n}}]$$

Since

$$\lim_n \frac{C}{\sqrt{n}} + \frac{v_n}{\sqrt{n}} = 0 \quad \text{a.s.}$$

if Ω_n denotes the event

$$\{\left|\frac{C}{\sqrt{n}} - \frac{v_n}{\sqrt{n}}\right| < a\}$$

we have $P[\Omega_n] \to 1$, for all positive constant a. Therefore, by using this and the central limit theorem, we get

$$\begin{aligned}
\lim_n P[\frac{\sum_{k=1}^{n} e_{-n+k-2}}{\sqrt{n}} > \frac{C}{\sqrt{n}} + \frac{v_n}{\sqrt{n}}] \quad &\geq \quad \lim_n P[\left\{\frac{\sum_{k=1}^{n} e_{-n+k-2}}{\sqrt{n}} > \frac{C}{\sqrt{n}} + \frac{v_n}{\sqrt{n}}\right\} \cap \Omega_n] \\
&\geq \quad \lim_n P[\left\{\frac{\sum_{k=1}^{n} e_{-n+k-2}}{\sqrt{n}} > a\right\} \cap \Omega_n] \\
&= \quad 1 - F_G(a)
\end{aligned}$$

where $F_G(a)$ is the c.d.f. of the appropriate centered Gaussian distribution. Since this is true for all $a > 0$, the proof of (22) is concluded. ■

4 Stability and Convergence of the Modified Rate Matching Algorithm

In this section, we analyze the non linear stochastic dynamical system given by equations

$$
\begin{aligned}
q_{n+1} &= \max(q_n + x_n - \mu_n, 0) && (23)\\
\mu_n &= \mu + \xi_n && (24)\\
\xi_n &= \min\left\{\max\left(\sum_{i=1}^{d} l_i \xi_{n-i} + k\gamma_n, \underline{\mu} - \mu\right), \overline{\mu} - \mu\right\} && (25)\\
x_{n+1} &= \max((1-\alpha)x_n + \alpha(\mu_n + \delta_{n+1}) - \beta(q_n - Q), 0) && (26)
\end{aligned}
$$

for $n \geq 0$, where γ_n are independent and identically distributed (i.i.d.) standard Gaussian random variables, k and l_i, $i = 1 .. d$, are constant, $\xi_{-d}, ..., \xi_{-1}$ are given initial conditions, and δ_n are i.i.d. Gaussian random variables defined by (7). We assume below that $\underline{\mu} > 0$.

Note that $\vec{\xi_n} := (\xi_{n-d+1}, \xi_{n-d+2}, ..., \xi_n)$ is a Markov chain with a state space $\mathcal{X} \subset I\!R^d$, for some $d > 0$. We denote by $\mathcal{B}(\mathcal{X})$ the Borel σ-field associated with $\mathcal{X}$. Let $P : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to [0,1]$ be the transition probabilities associated with ξ_n. It is easily checked that $\vec{\xi_n}$ satisfies the following property for any values of l_i, provided that $k \neq 0$: There exists a non-zero subprobability measure ψ on $\mathcal{B}(\mathcal{X})$ and an integer j such that, for all $B \in \mathcal{B}(\mathcal{X})$,

$$\inf_{x \in \mathcal{X}} \sum_{n=1}^{j} P^n(x, B) > \psi(B) \tag{27}$$

Let S_n denote the Markov chain $S_n = (q_n, x_n, \xi_n)$. We denote by $\mathcal{S}$ the state space of the Markov chain S_n, and by $\mathcal{B}(\mathcal{S})$ the associated Borel σ-field. Our main result is the following:

Theorem 1 *If $\alpha > 0$ and $\beta > 0$, then*

(i) S_n is Harris ergodic (see definition in [23]).

(ii) S_n is geometric recurrent. All moments of x_n and q_n exist in steady state (as $n \to \infty$), and the moments of x_n and q_n converge geometrically fast to the steady state moments.

Remark: It turns out that Theorem 1 holds for more general models for μ_n than is used above. Specifically, ξ_n may be any Markov chain with a state space $\mathcal{X} \subset I\!R^d$ provided that it satisfies (27) (instead of (25)). We would then define the bottleneck service rate μ_n to be given by $\mu_n = v(\xi_n)$ where $v : I\!R^d \to [\underline{\mu}, \overline{\mu}]$ (instead of (24)).

We prove Theorem 1 using the following generalization of Foster's criterion due to [18], [23] based on drift analysis:

Theorem 2 *Consider a strongly aperiodic Markov chain Y_n on a state space $\mathcal{Y}$ with transition probabilities $P : \mathcal{Y} \times \mathcal{B}(\mathcal{Y}) \to [0,1]$. Assume that there exists a set $\mathcal{H} \in \mathcal{Y}$ and a function $g : \mathcal{Y} \to I\!R$, $g(\cdot) \geq 1$, such that (i) $E[g(Y_{n+1})|Y_n] < \infty$ for $Y_n \in \mathcal{H}$; (ii) $\mathcal{H}$ is a small set. Then the following hold*

> *If (iiia) there exists $\epsilon > 0$ such that $E[g(Y_{n+1}) - g(Y_n)|Y_n] \leq -\epsilon$ for $Y_n \in \mathcal{H}^c$, then Y_n is Harris ergodic.*
>
> *If (iiib) there exists some $\epsilon > 0$ such that $E[g(Y_{n+1}) - g(Y_n)|Y_n] \leq -\epsilon g(Y_n)$ for $Y_n \in \mathcal{H}^c$, then*
>
> *(a) Y_n is geometrically ergodic (see definition in [18]).*
>
> *(b) Let π denote the steady state distribution of Y_n and let $\hat{Y}$ denote the random variable with distribution π. For any function $f : \mathcal{Y} \to I\!R$ such that $0 \leq f \leq ag$ for some constant $a > 0$, we have*
>
> *(b.1) $E[f(\hat{Y})] < \infty$*
>
> *(b.2) There exist $\zeta < 1$ and $R \geq 0$ such that*

$$\left| \int_{y \in \mathcal{Y}} f(y) P^n(y', dy) - \int_{y \in \mathcal{Y}} f(y) \pi(dy) \right| \leq R \zeta^n g(y') \qquad forally' \in \mathcal{Y}$$

In Theorem 2, the set $\mathcal{H}$ is said to be small if there exists some positive measure ϕ on $\mathcal{Y}$ such that for any $B \subset \mathcal{Y}$ with $\phi(B) > 0$, there exists j such that $\inf_{y \in \mathcal{H}} \sum_{n=1}^{j} P^n(y, B) > 0$. The Markov chain Y_n is said to be strongly aperiodic if there exists a set $\mathcal{C}$, a probability ν with $\nu(\mathcal{C}) = 1$, and $\eta > 0$ such that $P(Y_n$ enters $\mathcal{C}$ for some $n \geq 1 | Y_0 = y) > 0$ for all $y \in \mathcal{Y}$ and $P(y, B) \geq \eta \nu(B)$ for $y \in \mathcal{C}$ and $B \in \mathcal{B}(\mathcal{Y})$.

Proof of Theorem 2: The first claim as well as (a) are proved in [23], (b.1) is proved in [23] and (b.2) follows from [18] (corollary to Theorem 6.2). In [18], it is required that the set $\mathcal{H}$ be a "petite" set. However, any small set is also a petite set. ■

Proof of Theorem 1: Follows from Theorem 2. We show that the conditions of that Theorem indeed hold. We first prove (i), i.e. the Harris ergodicity. Let $z_n := 1 + aq_n + x_n$ where a is some constant satisfying $0 < a < \min(1, \alpha)$. Thus we choose here $g(q_n, x_n, \xi_n) =$

z_n as the function appearing in Theorem 2. Condition (i) of Theorem 2 is straightforward to verify. Next we establish condition (iiia). Suppose

$$x_n + q_n - \mu_n \geq 0 \tag{28}$$

Then

$$z_{n+1} \;=\; 1 + \max\left(q_n[a-\beta] + x_n[a+1-\alpha] + \alpha\delta_{n+1} + \alpha\mu_n + \beta Q, a(x_n+q_n)\right) - a\mu_n$$

and

$$\Delta := z_{n+1} - z_n = \max\left(-\{\beta q_n + [\alpha - a]x_n\} + \alpha\delta_{n+1} + \alpha\mu_n + \beta Q, -(1-a)x_n\right) - a\mu_n$$

Assume that

$$\beta q_n + [\alpha - a]x_n \geq \alpha(M + \overline{\mu}) + \beta Q \quad \text{ for some } M. \tag{29}$$

Then since $0 < a < 1$,

$$\begin{aligned} \Delta &\leq \max(-\alpha(M+\overline{\mu}) + \alpha\delta_{n+1} + \alpha\mu_n, 0) - a\underline{\mu} \\ &\leq -a\underline{\mu} - \alpha M + \max(\alpha\delta_{n+1}, \alpha M) \\ &\leq -a\underline{\mu} + \alpha\delta_{n+1} 1(\delta_{n+1} \geq M) \end{aligned} \tag{30}$$

where $1(\delta_{n+1} \geq M)$ is defined as 1 if $\delta_{n+1} \geq M$ and 0 otherwise. It follows that

$$E[\Delta | S_n] \leq -a\underline{\mu} + \alpha E[\delta_{n+1} 1(\delta_{n+1} \geq M)] \tag{31}$$

whenever $S_n \in \mathcal{H}_M^c$, where

$$\mathcal{H}_M = \left\{(q, x, \xi) : q \leq L_q^M, x \leq L_x^M\right\},$$

$$L_q^M = \max\left(\frac{\alpha(M+\overline{\mu}) + \beta Q}{\beta}, \overline{\mu}\right), \; L_x^M = \frac{\alpha(M+\overline{\mu}) + \beta Q}{\alpha - a}$$

Note that $L_x^M \geq \overline{\mu}$. Hence $S_n \in \mathcal{H}_M^c$ implies (28) and (29). Indeed, if $x_n > L_x^M$ then (28) and (29) hold; if $q_n > L_q^M$ then since $a < \alpha$ and $x_n \geq 0$, $\beta q_n \geq \alpha(M+\overline{\mu}) + \beta Q \geq \alpha(M+\overline{\mu}) + \beta Q - (\alpha - a)x_n$ which yields (29), and (28) follows since $x_n \geq 0$.

Since δ is integrable, by choosing M large enough the right hand side of (31) is strictly negative, from which we obtain condition (iiia).

Next we show that $\mathcal{H}_M$ is a small set for any $M > 0$. Define the measure ϕ as follows. For any $B \in \mathcal{B}(\mathcal{S})$

$$\phi(B) = \int_{(q,x,\xi)\in B} 1\{q = 0, x = 0\}\psi(d\xi)$$

where ψ is given in (27). Assume $S_n \in \mathcal{H}_M$. Let $L'_M = \max[0, (1-\alpha)L_x^M] + \beta Q$. If $\delta_{n+1} < -(\overline{\mu} + L'_M/\alpha)$ then it follows that $x_{n+1} = 0$. If for some integer $N > 0$, $\delta_{n+k} < -(\overline{\mu} + L'_M/\alpha)$ for all $k = 1, ..., N$, then it follows that $x_{n+k} = 0$ for all $k = 1, ..., N$.

In particular, let $N(M)$ be the smallest integer greater than $j + (L_q^M + L_x^M)/\underline{\mu}$ (where j is given in (27)). Note that since $S_n \in \mathcal{H}_M$ then $q_{n+1} < L_q^M + L_x^M$. Then $q_{n+N(M)} = 0$. Moreover, for any $B \in \mathcal{B}(\mathcal{S})$ for which $\phi(B) > 0$ we have

$$\begin{aligned} \inf_{s \in \mathcal{H}_M} \sum_{i=1}^{N(M)} P^i(s, B) &\geq \inf_{s \in \mathcal{H}_M} \sum_{i=N(M)-j}^{N(M)} P^i(s, B) \\ &\geq \phi(B)(Prob(\delta_n < -(\mu + L'_M/\alpha)))^{N(M)} < 0 \end{aligned}$$

This establishes the smallness of $\mathcal{H}_M$.

Next we show that S_n is strongly aperiodic. The recurrent set $\mathcal{C}$ can be chosen as $\mathcal{C} = \{0\} \times \{0\} \times \mathcal{X}$, and the measure ν can be chosen to be $\nu(\bullet) = \phi(\bullet)/\psi(\mathcal{X})$ ($\mathcal{C}$ and ν thus do not depend on M). η is then given by $\eta = Prob(\delta_n < -(\overline{\mu} + L'_M/\alpha))$ where we may take $M = 0$. (Note that if $\delta_n < -(\overline{\mu} + L'_0/\alpha)$ then $x_{n+1} = -\mu$. Moreover, if $q_n = 0$ and $x_n = 0$ then $q_{n+1} = 0$).

Next we establish the second part of the Theorem. Denote $r_n := \exp\{z_n\}$. (We choose here $g(q_n, x_n, \xi_n) = r_n$ as the function appearing in Theorem 2). We show that (iiib) holds in Theorem 2. We have

$$r_{n+1} - r_n = r_n(\exp(\Delta - 1))$$

Choose some ρ, $0 < \rho < 1$. It thus suffices to show that a proper choice of M yields $E(\exp\Delta | S_n) < \rho$. Indeed, it follows from (30) that

$$E(\exp\Delta | S_n) \leq \exp(-a\underline{\mu}) E[\exp(\alpha\delta_{n+1}) 1(\delta_{n+1} \geq M)]. \qquad (32)$$

Since $\exp(\alpha\delta_{n+1})$ is integrable for any α, by choosing M large enough, the right hand side of (32) can be made smaller than any ρ. The integrability of $\exp(\alpha\delta_{n+1})$ also implies condition (i). ∎

The Gaussian assumption for δ_n is not essential in the proof of Theorem 1. Specifically, the proof of the Harris ergodicity of S_n only requires that δ_n have finite expectation, and that δ_n have a positive probability of being smaller than $-(\mu + L'_M/\alpha)$ (the latter condition can be further relaxed). The proof of the geometric recurrence of S_n only requires that $\exp(\alpha\delta_{n+1})$ be integrable, and again that δ_n have a positive probability of being smaller than $-(\mu + L'_M/\alpha)$).

5 A Linear Approximation

In Section 4, we analyzed the model given by equations (23-26). However, we did not obtain closed-form expressions for the steady-state or transient moments of q_n or x_n. In this section, we analyze an approximate model which we obtain by eliminating the max operation in equations (23-26), specifically

$$\begin{aligned} q_{n+1} &= q_n + x_n - \mu_n && (33) \\ \mu_n &= \mu + \xi_n && (34) \\ \xi_n &= \sum_{i=1}^{d} l_i \xi_{n-i} + k\gamma_n && (35) \\ x_{n+1} &= (1-\alpha)x_n + \alpha(\mu_n + \delta_{n+1}) - \beta(q_n - Q) && (36) \end{aligned}$$

The system (33-36) is a linear system, and hence it is easily tractable. However, we note that q_n and x_n can now take negative values. This is clearly unrealistic, but simulations indicate that in many cases the behavior of the linear and the non-linear systems are similar (example simulations are shown below). This indicates that the analysis of the linear system can provide insight into the behavior of the non-linear system.

For the analysis, it is convenient to rewrite the system in matrix form. Specifically, let

$$Y_n = \begin{bmatrix} x_n - \mu \\ q_n - Q \\ \xi_n \\ \vdots \\ \xi_{n-d+1} \end{bmatrix} \qquad A = \begin{bmatrix} 1-\alpha & -\beta & \alpha & 0 & \dots & 0 & 0 \\ 1 & 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 0 & l_1 & l_2 & \dots & l_{d-1} & l_d \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \qquad W_n = \begin{bmatrix} \alpha\delta_{n+1} \\ 0 \\ k\gamma_{n+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then it is easy to verify that the system (33-36) can be rewritten as

$$Y_{n+1} = AY_n + W_n \qquad (37)$$

For convenience, we decompose the matrix A into block matrices. Throughout the rest of the paper, we denote by B the upper left 2×2 block matrix of A, and by C the lower left $d \times d$ block matrix of A, i.e.

$$B = \begin{bmatrix} 1-\alpha & -\beta \\ 1 & 1 \end{bmatrix} \qquad C = \begin{bmatrix} l_1 & l_2 & \dots & l_{d-1} & l_d \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

Equation (37) describes the evolutions of a discrete-time linear stochastic system of order $d+2$. The general solution of (37) is (e.g. [16])

$$Y_n = A^n Y_0 + \sum_{i=0}^{n-1} A^{n-i-1} W_i \tag{38}$$

If Y_0 and W_n are normally distributed, then Y_n is also normally distributed for all $n \geq 1$ [16]. Hence its distribution is completely characterized by its mean and its covariance. We use Σ_n to denote the covariance of Y_n i.e. $\Sigma_n = Cov(Y_n) = E[Y_n Y_n^T]$. Taking the expected value on both sides of equation (38), we obtain

$$E[Y_n] \;=\; A^n E[Y_0] + \sum_{i=0}^{n-1} A^{n-i-1} E[W_i]$$

Since δ_n and γ_n are zero-mean variables, the above equation becomes

$$E[Y_n] = A^n E[Y_0] \tag{39}$$

Let V denote the covariance matrix of W_n, i.e. $V = E[W_n W_n^T]$. The covariance Σ_n is computed recursively as follows:

$$\begin{aligned}
\Sigma_0 &= Cov(Y_0) \\
\Sigma_n &= Cov(AY_{n-1}) + Cov(W_{n-1}) \\
&= AE[Y_{n-1}Y_{n-1}^T]A^T + V \\
&= A\Sigma_{n-1}A^T + V \qquad \text{for} \quad n \geq 1
\end{aligned}$$

If Y_0 and W_n are not Gaussian, then Y_n is not Gaussian. However, its mean and covariance are given by the same expressions as above.

We next analyze the behavior of $E[Y_n]$. We first consider the steady-state behavior of $E[Y_n]$. Standard results from linear algebra indicate that the limit $\lim_{n\to\infty} A^n$ exists and is equal to 0 if A is stable, i.e. if all eigenvalues of A have modulus less than 1 [16]. In this case, we have $\lim_{n\to\infty} E[Y_n] = 0$ and hence $\lim_{n\to\infty} E[x_n] = \mu$ and $\lim_{n\to\infty} E[q_n] = Q$. Specifically, we obtain:

Theorem 3 *If (i) $0 < \beta < \alpha < 2 + \beta/2 < 4$ and (ii) $\sum_{i=1}^{d} |l_i| < 1$, then*

$$\lim_{n\to\infty} E[x_n] = \mu \quad \textit{and} \quad \lim_{n\to\infty} E[q_n] = Q$$

If (i) does not hold, then $\lim_{n\to\infty} |E[x_n]| = \infty$ *and* $\lim_{n\to\infty} |E[q_n]| = \infty$

Proof of Theorem 3: The matrix A has $d+2$ eigenvalues, namely the 2 eigenvalues of B and the d eigenvalues of C. Let λ_1 and λ_2 denote the eigenvalues of B. λ_1 and λ_2 are the roots of the equation $\lambda^2 - (2-\alpha)\lambda + 1 - \alpha + \beta = 0$. It is easy to verify that these roots have modulus less than 1 if $0 < \beta \leq \alpha \leq 2 + \beta/2 < 4$. The eigenvalues of C are the roots of a complicated polynomial of degree d. Using Gerschgorin's theorem [21], we find that the modulus of any root λ of this polynomial satisfies the inequalities $|\lambda| \leq 1$ or $|\lambda - l_1| \leq \sum_{i=2}^{d} |l_i|$. Thus a sufficient condition for stability of C is $\sum_{i=1}^{d} |l_i| < 1$. ■

We note that the region of stability of the linear model, i.e. the values of α, β, and l_i for which the model is stable, is much smaller than the region of stability of the non-linear model (refer to Theorem 1). This brings out the limitations of the linear model. The shaded area in Figure 1 shows the region of stability of the matrix B, i.e. the region defined by $0 < \beta < \alpha < 2 + \beta/2 < 4$ in the (α, β) plane (the curves $\alpha = 2\sqrt{\beta}$ and $\alpha = 1 + \beta$ will be useful shortly).

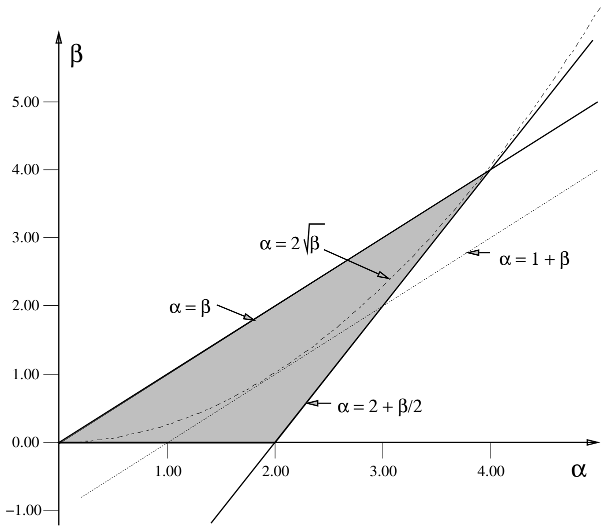

Figure 1: The domain $0 < \beta < \alpha < 2 + \beta/2 < 4$.

We next consider the transient behavior of $E[Y_n]$. Because of equation (39), it suffices to obtain an analytic expression of A^n. This is done by diagonalizing A. For reasons of space, we omit the analysis and present the results only for $E[x_n]$ assuming $E[q_0] = Q$ and $E[\xi_{-i}] = 0$ for $i = 0, ..., d-1$. Recall that λ_1 and λ_2 denote the eigenvalues of the matrix B.

Theorem 4 *Assume that $0 < \beta \leq \alpha \leq 2 + \beta/2 < 4$ and $E[q_0] = Q$. Then the following holds*

- *If $\alpha = 2\sqrt{\beta}$, then $\lambda_1 = \lambda_2 = 1 - \alpha/2$ and*

$$E[x_n] - \mu = (E[x_0] - \mu)(1 - \alpha/2)^n$$

- *If* $\alpha < 2\sqrt{\beta}$, *then* λ_1 *and* λ_2 *are complex conjugate eigenvalues and*

$$E[x_n] - \mu = -\frac{E[x_0] - \mu}{\lambda_1 - \lambda_2}\rho^{n-1}[\rho \sin(n+1)\theta - \sin n\theta]$$

where $\rho = \sqrt{1 - \alpha + \beta}$ *and* $\cos\theta = (2 - \alpha)/2\rho$

- *If* $\alpha > 2\sqrt{\beta}$, *then* λ_1 *and* λ_2 *are distinct real eigenvalues and*

$$E[x_n] - \mu = \frac{E[x_0] - \mu}{\lambda_1 - \lambda_2}[(\lambda_1 - 1)\lambda_1^n - (\lambda_2 - 1)\lambda_2^n]$$

Thus we see that the behavior of $E[x_n]$ follows one of three possible patterns. It turns out that the behavior of $E[q_n]$ follows the same patterns. These patterns correspond to the standard patterns of second-order linear systems [16], with critical damping corresponding to the case $\alpha = 2\sqrt{\beta}$, overcontrol corresponding to the case $\alpha < 2\sqrt{\beta}$, and undercontrol corresponding to the case $\alpha > 2\sqrt{\beta}$. In all three cases, the following holds:

Theorem 5 *The average source rate* $E[x_n]$ *converges rapidly to* μ *if* α *is close to 2 and* β *is close to 1.*

Proof of Theorem 5: If $\alpha = 2\sqrt{\beta}$, then the difference $E[x_n] - \mu$ decreases monotonically as $(1 - \alpha/2)^n$. Therefore, $E[x_n]$ stabilizes rapidly to μ if α is close to 2. Since $\beta = \alpha^2/4$, it follows that β is then close to 1. If $\alpha < 2\sqrt{\beta}$, then it is easy to see that $E[x_n] - \mu$ oscillates with a period equal to $2\pi/\theta$ within an envelope which decreases as ρ^{n-1}. Thus, $E[x_n]$ converges rapidly to μ if ρ is close to 0, i.e. $\alpha \approx 1 + \beta$. With the constraint $\alpha < 2\sqrt{\beta}$, this implies that α is close to 2 and β is close to 1 (refer to Figure 1). If $\alpha > 2\sqrt{\beta}$, then the behavior of $E[x_n]$ is more complicated, but it is clear that $E[x_n]$ converges rapidly to μ if both λ_1 and λ_2 are small. It turns out that $\lambda_1\lambda_2 = 1 - \alpha + \beta$ and $\lambda_1 + \lambda_2 = 2 - \alpha$. Thus, λ_1 and λ_2 are small if α is close to $1 + \beta$ and α is close to 2, i.e. α is close to 2 and β is close to 1. ∎

Similarly, it can be shown that the expected queue size $E[q_n]$ converges rapidly to the target queue size Q if α is close to 2 and β is close to 1. We note that the value $\alpha = 2$ is much higher than typical values of forgetting factors, which range between 0.01 and 0.05 [16]. Next, we show that such a large value of α implies a very large covariance, and hence an inefficient behavior, for x_n and q_n in steady-state. Specifically, we have the following:

Theorem 6 *If* $0 < \beta < \alpha < 2 + \beta/2 < 4$ *and* $\sum_{i=1}^{d} |l_i| < 1$, *then the covariance* Σ_n *converges to a semidefinite matrix* Σ *which is the unique solution of the linear equation* $\Sigma = A\Sigma A^T + V$.

Proof of Theorem 6: The theorem holds if A is stable (e.g. [16, p.29]). We have shown earlier that A is stable if $0 < \beta < \alpha < 2 + \beta/2 < 4$ and $\sum_{i=1}^{d} |l_i| < 1$, which concludes the proof. ■

We obtain closed-form expressions for the limiting variance of x_n and q_n by solving the matrix equation above. However, these expressions are complicated when the order of the ARMA model for μ_n is high (i.e. the value of d in equation (36) is large). In the simplest case, i.e. for the $0th$ order ARMA model when $\mu_n = \mu + \gamma_n$, we obtain for example

$$\lim_{n\to\infty} E[(x_n - \mu)^2] = \frac{2\alpha^2\sigma_\delta^2}{(\alpha - \beta)(-2\alpha + \beta + 4)}$$

where σ_δ^2 denotes the variance of δ_n.

Thus, the steady-state variance of x_n increases as α increases. This means that the flow of data sent by the source into the network becomes more bursty as α increases. It can be shown that the steady-state variance of q_n also increases as α increases, implying large buffer requirements at the bottleneck if we are to avoid frequent buffer overflow. Therefore, a tradeoff has to be found between a rapid convergence of $E[x_n]$ to μ and $E[q_n]$ to Q, i.e. a rapid relaxation of the system to steady-state, and the efficiency of the flow control mechanism in steady-state. This is consistent with the results reported in the literature [8, 9, 19].

Validation by simulation

Recall that our objective in Section 2 was to analyze the stochastic dynamical system defined by equations (23-26). In Section 4, we obtained stability and convergence results for this system. In this section, we obtained closed-form expressions for the behavior of an approximate model of the system, defined by equations (33-36). We now compare, by simulation, the behaviors of the two systems. We use q_n^+ and x_n^+ to denote the solutions of equations (23-26), and q_n and x_n to denote the solutions of equations (33-36).

In all simulations, we assume that the target queue size is $Q = 50$. We assume that the noise process δ_n is normally distributed with mean 0 and variance σ_δ^2. We choose the parameters of the truncated ARMA model for μ_n as follows: $\mu = 10$, $\underline{\mu} = 0$, $\overline{\mu} = 20$, $d = 2$, $l_1 = 0.7$, $l_2 = -0.3$, and $k = 1$. We assume that γ_n is normally distributed with mean 0 and variance 5.

Figure 2 shows the evolutions of μ_n, x_n^+, x_n, q_n^+, and q_n in the range $0 \le n \le 100$ when $\alpha = 0.15$, $\beta = 0.05$, and $\sigma_\delta^2 = 2$. We observe that the behavior of x_n is similar to the behavior of x_n^+, and the behavior of q_n is similar to the behavior of q_n^+ when q_n is positive. Thus, it seems reasonable to approximate x_n^+ by x_n and q_n^+ by $\max(q_n, 0)$. However, it is clear that these approximations are not always valid. For example, the stability regions for x_n and x_n^+ are different. Nevertheless, we have found the approximations above to be reasonable for the values of α and β that ensure convergence of $E[x_n]$, i.e. the values that fall in the shaded area of Figure 1. In the rest of this section, we only present simulations of the behavior of x_n^+ and q_n^+.

We examine the behavior of x_n^+ and q_n^+ for different values of α and β. Figure 3 shows evolutions of μ_n, x_n^+, and q_n^+ in the range $0 \leq n \leq 100$ for $\alpha = 0.15$, $\beta = 0.01$, and $\sigma_\delta^2 = 2$. Figure 4 shows the same evolutions when $\alpha = 1.5$ and $\beta = 0.3$. As predicted by Theorems 4 and 6, we observe that as α increases, there is rapid convergence of the source rate but large queue size variance (leading to inefficient steady state operation).

We could choose various criteria to determine optimal values for α and β (e.g., [5]). However, in all cases, the performance of the control mechanism is limited by the above tradeoff. Recent work in control theory has shown that a control mechanism with a dynamic α, i.e. where α varies dynamically with n, can overcome the limitations of the tradeoff (e.g. [10]).

If α and β are large, then the average send rate and the average bottleneck queue size converge very quickly to their steady-state values. However, when the steady-state is reached, then the flow of data sent by the source into the network is very bursty and the buffer requirements at the bottleneck node are high. This suggests that the source should choose small values for α and β if the network remains in steady-state for long periods of time (i.e. if μ_n is either constant or varies slowly with time), and large values for α and β if the state of the network changes frequently and abruptly (i.e. if μ_n varies rapidly with time). The obvious solution then is for the source to adjust the values of α and β dynamically as the behavior of μ_n changes with time.

There are many possible ways to adjust α and β as described above (e.g. [10, 13]). In this paper, we assume that β is static and we let

$$\begin{aligned} \alpha_n &= D \, \frac{(x_n - \hat{\mu}_n)^2}{\sigma_n^2} \\ \sigma_n^2 &= (1 - D')\sigma_{n-1}^2 + D'(x_n - \hat{\mu}_n)^2 \end{aligned} \tag{40}$$

where D and D' are constant. We also require that $0 \leq \alpha_n \leq 1$ for all n.

Figure 5 shows the evolutions of μ_n and of x_n^+ in the range $0 \leq n \leq 100$ when $\beta = 0.05$, $\sigma_\delta^2 = 2$, and $D = D' = 0.1$, for the three cases of static $\alpha = 0.5$, static $\alpha = 0.05$, and dynamic α_n given by equation (40). We observe that for dynamic α_n, x_n^+ increases to steady-state at the same rate as with static $\alpha = 0.5$, but its steady-state variance is much smaller. Thus, it appears that a dynamic forgetting factor can provide both rapid relaxation to steady-state and efficient behavior in steady-state.

6 Conclusion

We have used a discrete-time stochastic approach to analyze a generic flow control mechanism which captures the properties of several recently proposed such mechanisms. We have found that our generic mechanism is stable and efficient only if the control actions taken by the source are based on feedback information that includes not only an estimate of the bottleneck rate, but also an estimate of the bottleneck queue size. We note that the control mechanisms proposed in [13, 19] use both these estimates.

We are extending our work in several ways. One is to extend the analysis of the non linear system (23-26) to consider a dynamic forgetting factor. Another one is to model and analyze control mechanisms in which several control actions are taken per roundtrip time.

The discrete-time stochastic approach used in this paper appears useful to analyze other network flow and congestion control problems. One such problem we are considering is the feedback control of video codecs. There, the goal is to control the output rate of the codec (this is done by adjusting parameters of the codec such as the movement detection threshold or the quantizer value) based on feedback information about the available capacity in the network. We have implemented one such control scheme in IVS, a videoconference system for the Internet developed at INRIA [22], and we are currently evaluating its performance.

References

[1] A. K. Agrawala, D. Sanghi, "Design and evaluation of an adaptive flow control mechanism", *Proc. IEEE INFOCOM '92*, Florence, Italy, pp. 2391-2397, May 1992.

[2] E. Altman, P. Nain, "Closed loop control with delayed information", *Proc. ACM SIGMETRICS '92*, Newport, RI, pp. 193-204, May 1992.

[3] F. Baccelli, P. Brémaud, *Palm Probabilities and Stationary queues.* Lecture Notes in Statistics No. 41, Springer Verlag, Berlin 1987.

[4] L. BenMohamed, S. Meerkov, "Feedback control of congestion in store-and-forward datagram networks: the case of a single congested node", *Proc. 31st Conf. on Decision and Control*, Tucson, AZ, pp. 991-996, Dec. 1992.

[5] A. Benveniste, G. Ruget, "A measure of the tracking capability of recursive stochastic algorithms with constant gains", *IEEE Trans. Aut. Cont.*, vol. 27, pp. 639-649, June 1982.

[6] J-C. Bolot, "A self-tuning regulator for overload control in communication networks", *Proc. 31st Conf. on Decision and Control*, Tucson, AZ, pp. 1022-1023, Dec. 1992.

[7] J-C. Bolot, A. U. Shankar, "Analysis of a fluid approximation to flow control dynamics", *Proc. IEEE INFOCOM '92*, Florence, Italy, pp. 2398-2407, May 1992.

[8] T. Faber, L. Landweber, A. Mukherjee, "Dynamic time windows: Packet admission control with feedback", *Proc. ACM SIGCOMM '92*, Baltimore MD, pp. 124-135, Aug. 1992.

[9] K. Fendick, M. A. Rodrigues, A. Weiss, "Analysis of a rate-based control strategy with delayed feedback", *Proc. ACM SIGCOMM '92*, Baltimore, MD, pp. 136-148, August 1992.

[10] T. R. Fortescue, L. Kershenbaum, B. Ydstie, "Implementation of self-tuning regulators with variable forgetting factors", *Automatica*, vol. 17, no. 6, pp. 831-835, 1981.

[11] V. Jacobson, "Congestion avoidance and control", *Proc. ACM SIGCOMM '88*, Stanford, CA, pp. 314-329, August 1988.

[12] P. Kermani, L. Kleinrock, "Dynamic flow control in store-and-forward computer networks", *IEEE Trans. Comm.*, vol. COM-28, no. 2, pp. 263-271,

[13] S. Keshav, "A control-theoretic approach to flow control", *Proc. ACM SIGCOMM '91*, Zurich, Switzerland, September 1991.

[14] S. Keshav, "Flow control in high-speed networks with long delays", *Proc. INET '92*, Kobe, Japan, pp. 461-470, May 1992.

[15] K-T. Ko, P. Mishra, "Interactions among virtual circuits using predictive congestion control", *Computer Networks and ISDN Systems*, vol. 25, no. 6, Jan. 1993.

[16] P. R. Kumar, P. Varaiya, *Stochastic Systems: Estimation, Identification and Control*, Prentice-Hall, 1986.

[17] T. Lindvall, "Lectures on the Coupling Method", *Wiley*, 1992

[18] S. P. Meyn and R. L. Tweedie, "Stability of Markovian processes I: criteria for discrete time chains", *Advances in Applied Probability*, vol. 24, no. 3, pp. 542-574, 1992.

[19] P. P. Mishra, H. Kanakia, "A hop-by-hop rate-based congestion control scheme", *Proc. ACM SIGCOMM '92*, Baltimore, MD, Aug. 1992, pp. 112-123.

[20] A. Mukherjee, J. C. Strikwerda, "Analysis of dynamic congestion control protocols: A Fokker-Plank approach", *Proc. ACM SIGCOMM '91*, Zurich, Switzerland, Sept. 1991.

[21] B. Noble, *Applied Linear Algebra*, Prentice-Hall, Englewood Cliffs, NJ, 1969.

[22] T. Turletti, "A H.261 software codec for Internet videoconference", INRIA Research Report 1834, Feb. 1993.

[23] R. L. Tweedie, "Criteria for rates of convergence of Markov chains, with application to queuing and storage theory", *Probability, Statistics and Analysis*, ed. J.F.C. Kingman and G.E.H. Renter, London Math Soc. Lecture Notes Series 79, Cambridge Univ. Press, pp. 260-276, 1983.

[24] R. L. Tweedie, "The existence of moments for stationary Markov chains", *J. Applied Probability*, vol. 20, pp. 191-196, 1983.

[25] Y-T. Wang, B. Sengupta, "Performance analysis of a feedback congestion control policy under non-negligible propagation delay", *Proc. ACM SIGCOMM '91*, Zurich, Switzerland, pp. 149-158, Sept. 1991.

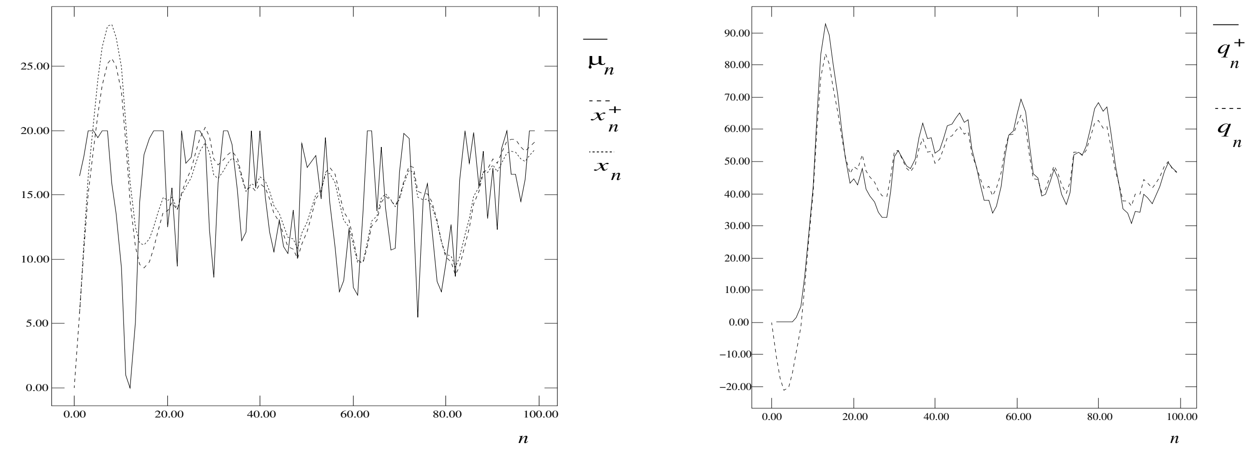

Figure 2: Evolutions of μ_n, x_n^+, x_n, q_n^+, and q_n when $\alpha = 0.15$, $\beta = 0.05$, and $\sigma_\delta^2 = 2$.

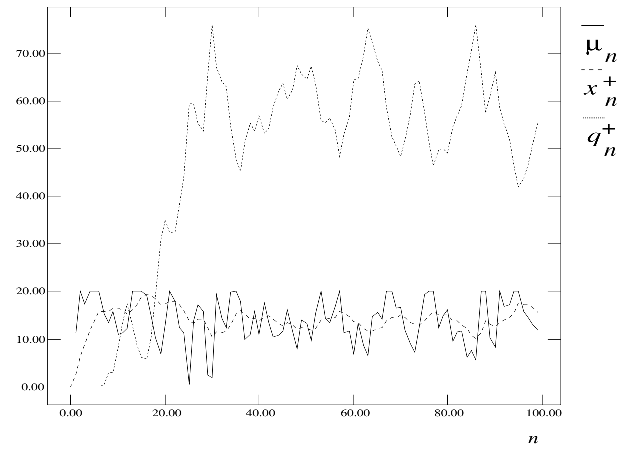

Figure 3: Evolutions of μ_n, x_n^+, and q_n^+ when $\alpha = 0.15$, $\beta = 0.01$, and $\sigma_\delta^2 = 2$.

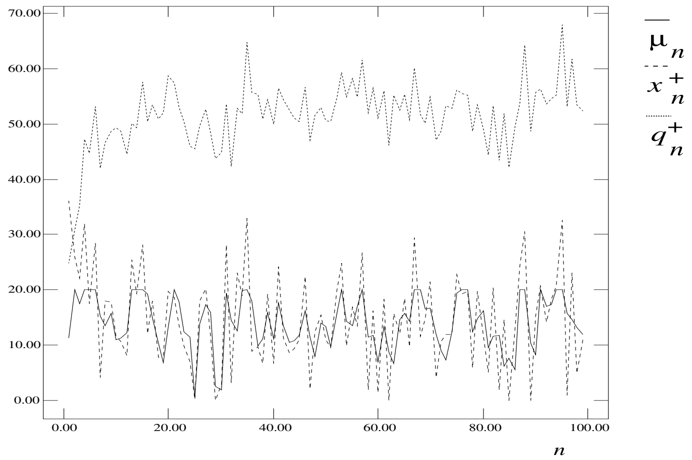

Figure 4: Evolutions of μ_n, x_n^+, and q_n^+ when $\alpha = 1.5$, $\beta = 0.3$, and $\sigma_\delta^2 = 2$.

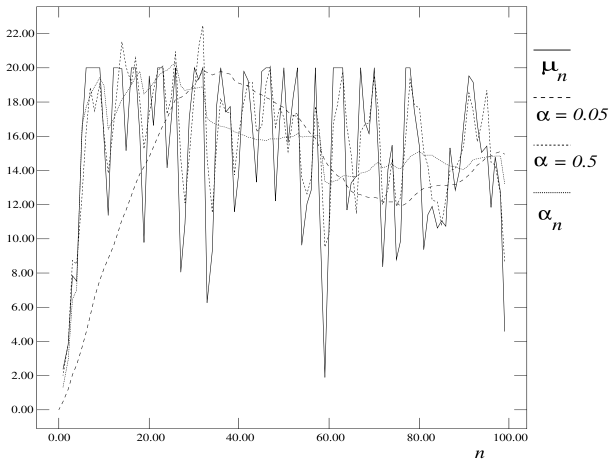

Figure 5: Evolutions of μ_n and x_n^+ for static and dynamic α's.