

TECHNICAL RESEARCH REPORT

Optimal Multilevel Feedback Policies for ABR Flow Control
using Two Timescale SPSA

by Shalabh Bhatnagar, Michael C. Fu, Steven I. Marcus

TR 99-86



ISR develops, applies and teaches advanced methodologies of design and analysis to solve complex, hierarchical, heterogeneous and dynamic problems of engineering technology and systems for industry and government.

ISR is a permanent institute of the University of Maryland, within the Glenn L. Martin Institute of Technology/A. James Clark School of Engineering. It is a National Science Foundation Engineering Research Center.

Web site <http://www.isr.umd.edu>

Optimal Multilevel Feedback Policies for ABR Flow Control using Two Timescale SPSA *

Shalabh Bhatnagar
Institute for Systems Research
University of Maryland
College Park, MD 20742
shalabh@isr.umd.edu
fax: (301) 314-9920

Michael C. Fu
The Robert H. Smith School of Business
and Institute for Systems Research
University of Maryland
College Park, MD 20742
mfu@umd5.umd.edu
fax: (301) 314-9157

Steven I. Marcus
Department of Electrical Engineering
and Institute for Systems Research
University of Maryland
College Park, MD 20742
marcus@isr.umd.edu
fax: (301) 314-9920

July 12, 1999

Abstract

Optimal multilevel feedback control policies for rate based flow control in available bit rate (ABR) service in asynchronous transfer mode (ATM) networks are obtained in the presence of information and propagation delays, using a numerically efficient two timescale simultaneous perturbation stochastic approximation (SPSA) algorithm. Convergence analysis of the algorithm is presented. Numerical experiments demonstrate fast convergence even in the presence of significant delays and large number of parametrized policy levels.

Key Words Optimal multilevel feedback policy, rate based ABR flow control, two timescale SPSA.

*This research was supported by the NSF under Grant DMI-9713720, by the Semiconductor Research Corporation under Grant 97-FJ-491 and by DoD contract Number MDA90497C3015.

1 Introduction

The available bit rate (ABR) service in asynchronous transfer mode (ATM) networks is used primarily for data traffic. As the name suggests, bandwidth allocation for ABR service is done only after the higher priority services such as constant bit rate (CBR) and variable bit rate (VBR) have been allocated bandwidth. The available bandwidth is a time-varying quantity, and for proper utilization, the network requires the ABR sources to control their own traffic flow. Two main proposals discussed at the ATM forum for flow control in ABR service [32] were the rate based and the credit based schemes. The rate based scheme [7] was recently accepted by the ATM forum primarily because of the higher hardware complexity and costs involved in the latter scheme [31] (see [32] for a general survey of the various proposals). Several heuristic algorithms for computation of explicit ABR rates have since been proposed at the ATM forum [28], [17]. Most other approaches ([3], [22], [19], [24], [25], [12], [35]) use fluid models of the network. Even though of obvious practical interest, very little is available in terms of stochastic control approaches in the queueing framework largely because they lead to analytical intractability when several realistic features are incorporated into the models. For instance, in [1] the problem is formulated as a team problem in the discrete time queueing framework but with linearized queue dynamics where the queue length is in fact allowed to become negative. In [30], a continuous time queueing model is studied and stability conditions for delayed feedback policies are obtained. However, performance analysis there is done only under the assumptions of no delays and continuous observations. In this paper, we adopt a model that is similar to [30]. We use an efficient simulation based stochastic approximation scheme for computing optimal ABR rates in the presence of delays, and with observations and information feedback available only at periodic time instants.

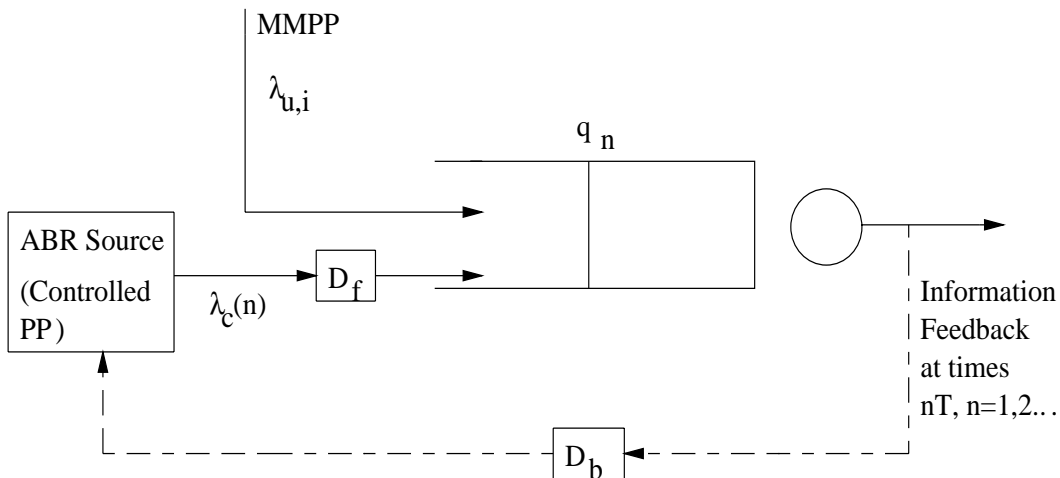


Figure 1: The ABR Model

The model we consider is a single queue (or switch) with finite buffer (of size B) fed with two input streams. This queue may represent a bottleneck node on the virtual circuit of an ABR source. One of the input streams (the uncontrolled stream) is a Markov modulated Poisson process (MMPP). The ABR source is represented by an infinite data source and the packets from this source are extracted as a controlled Poisson process. The queue length process is observed at instants nT ,

$n = 0, 1, \dots$, and from which the desired ABR rate for the next time interval is computed at the node. Delays in receiving the new rate information and in transmitting packets are included in our model. The objective is to find an optimal control policy for the ABR source that balances various performance measures such as throughput, mean and variance of delay and probability of overflow. Often this problem is addressed by minimizing the distance of stationary mean queue length from a given fixed constant N_0 [34], [18], [30].

In this paper, we consider parametrized policies that have several levels of control. We develop a simultaneous perturbation stochastic approximation (SPSA) [33] variant of a two timescale stochastic approximation algorithm [6] to obtain the optimal policy with this structure (see also [15] for application of SPSA to optimization of discrete event systems). The two timescale stochastic approximation algorithm developed in [6] for simulation based parametric optimization had the advantage that it updates the parameter at increasing deterministic instants [2], without the need for regeneration as in [11], [13]. This it achieves using different timescales (or step-size schedules). On the other hand, like other finite difference schemes, it requires $N + 1$ simulations for any N -vector parameter to obtain the gradient estimate. A proposed alternative was to update the parameter in ‘cycles’, in which only one component of the parameter is updated at a time using only two parallel simulations at any instant. This, however, slows down the convergence considerably. In [5], this algorithm was used for the ABR problem with the uncontrolled arrival stream Poisson: As a result of slow convergence, only three level parametrized policies were considered in the numerical experiments. The simultaneous perturbation stochastic approximation (SPSA) technique developed by Spall [33] requires only two simulations for any N -vector parameter, with all the N -components of the parameter vector updated simultaneously. The gradient is obtained from two performance measurements taken at randomly perturbed settings of the parameter components, most commonly by using i.i.d. symmetric Bernoulli random variables. Our numerical experiments indicate that the two timescale SPSA algorithm is particularly effective in obtaining optimal multilevel feedback policies (with as many levels as one wants) for our model. In sum, our work contributes to the ABR literature by developing a computationally efficient simulation based algorithm for ABR flow control using a queueing model that incorporates the important practical features of information and propagation delays. The algorithm is orders of magnitude faster than a previously proposed algorithm. Furthermore, our numerical experiments highlight the substantial performance gains obtained by employing the structured feedback policies proposed here over open loop policies.

The rest of the paper is organized as follows. In Section 2, we describe the model and the two timescale SPSA scheme for obtaining the optimal structured policy, and compare it with the original two timescale stochastic approximation scheme of [6]. The convergence of the algorithm is shown in the Appendix. In Section 3, we use numerical experiments to illustrate the algorithm and compare performance of multilevel closed loop optimal feedback policies with optimal open loop performance. Finally, in Section 4, we provide concluding remarks and extensions.

2 The Optimization Problem

The model that we consider, shown in Fig.1, is a bottleneck node with two input streams, one controlled (representing the traffic from the ABR source) and the other uncontrolled (representing all the other traffic in the network passing through this node). The ABR stream is modelled as a controlled Poisson process with instantaneous intensity specified by a feedback control law defined below. The uncontrolled stream is modelled as a Markov modulated Poisson process (MMPP). Let

$\{X(t), t \geq 0\}$ be a finite state, irreducible, aperiodic Markov process with state space S_u . When $X(t) = i \in S_u$, the instantaneous rate of the uncontrolled stream is $\lambda_{u,i}$. Let $X_n \triangleq X(nT)$ represent the state of the modulating chain of the uncontrolled MMPP at time nT . Let $p(i; j)$, $i, j \in S_u$ represent the transition probabilities of $\{X_n\}$. The size of the bottleneck buffer is B and could be large (e.g., in the simulation experiments that we illustrate in Section 3, B is taken as 5×10^5). Let the instantaneous queue length at time t be represented by $q(t)$ and let $q_n \triangleq q(nT)$, $n \geq 0$, represent the queue length at times nT , $n \geq 0$. We assume that the ABR rate is held fixed in the time intervals $[nT, (n+1)T)$, $n \geq 0$, with $\lambda_c(n)$ representing this ABR rate (in the n th interval) and is computed using the queue length (q_n) observed at the node. The new rate is then fed back to the source. Our model thus incorporates explicit rate feedback. This information however reaches the ABR source with a delay D_b , whereupon the ABR source starts sending packets with the new rate. Further, there is a propagation delay D_f in packets to arrive at the bottleneck node from the source. We assume all through that the quantities T , D_b and D_f are constants. Let $S = \{0, 1, \dots, B\}$ be the set of possible queue length values. Let θ represent the N -dimensional parameter vector to be optimized that takes values in a compact set $C \subset \mathcal{R}^N$. We shall assume in particular that C is of the form $\prod_{i=1}^N [\lambda_{i,\min}, \lambda_{i,\max}]$, with $\lambda_{i,\min} > 0$, for all $i = 1, \dots, N$. Then $\lambda_c(n)$ takes values in some closed set $D \subset [\min_{i \in \{1, \dots, N\}} \lambda_{i,\min}, \max_{i \in \{1, \dots, N\}} \lambda_{i,\max}]$ (depending on θ). Let $h : S \rightarrow \mathcal{Z}^+$ be a given bounded and nonnegative cost function (\mathcal{Z}^+ is the space of nonnegative integers). Our aim is to minimize the average cost

$$J(\theta) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n h(q_i). \quad (2.1)$$

The feedback policy that governs the rate $\lambda_c(n)$ is given in (2.2) below. In the Appendix, it is shown that under the type of feedback policies (2.2), the limit in (2.1) exists w.p.1 and is deterministic for each θ . Let a_i , $i = 0, 1, \dots, N$, be integers such that $-1 = a_0 < a_1 < a_2 < \dots < a_{N-1} < a_N = B$. Define the following subsets of S : $S_i = \{a_{i-1} + 1, \dots, a_i\}$, $i = 1, \dots, N$. Then for $i = 1, \dots, N$,

$$\lambda_c(n) = \lambda_i \text{ if } q_n \in S_i. \quad (2.2)$$

In the above, $\theta = (\lambda_1, \lambda_2, \dots, \lambda_N)^T$ is the given parameter. Let us now briefly motivate our choice of the form of policy (2.2). Note that if $N = B$, $S_i = \{i\}$, we have the most general case possible (viz., allocating one control to each possible state). Markov decision processes (MDP) [26] represent a general framework for dealing with such problems. However, a numerical solution using MDP based on standard MDP solution techniques like policy iteration and value iteration [26] normally faces the ‘curse of dimensionality’ for large state spaces. Our numerical experiments in Section 3 for instance take $B = 5 \times 10^5$. In addition MDP solution techniques also require explicit specification of the transition probabilities of the continuous time chain at instants $T, 2T$, etc. In the presence of non-zero delays D_b and D_f , these techniques become computationally prohibitive. Heuristic simulation based alternatives to standard MDP solution methods based on reinforcement learning techniques (also sometimes called Neuro Dynamic Programming (NDP) [4]) provide another possible solution procedure that are not considered here.

We now briefly motivate the use of stochastic approximation, i.e., when $J(\theta)$ (as defined in (2.1)) is not available analytically but must be estimated by simulation (see [14], [21]). A stochastic approximation algorithm recursively updates θ using gradient descent with decreasing step-sizes and an appropriate estimate of $\nabla J(\theta)$. The efficiency of the algorithm usually depends on the quality and computational requirements of the gradient estimate.

The advantages of using the SPSA approach are best appreciated after first presenting the original two timescale algorithm of [6]. We begin with the scalar parameter case. Let $\delta > 0$ be a small fixed constant. Let \bar{C} be the closed interval in which θ takes values. Let $\pi(\cdot)$ represent the projection map onto \bar{C} , i.e., $\pi(x)$ is the closest point in \bar{C} from x . Define sequences $\{a(n)\}$ and $\{b(n)\}$ in $(0, 1]$ as follows: $a(0) = b(0) = 1$, $a(i) = i^{-1}$, $b(i) = i^{-\alpha}$, $i \geq 1$, for some $\alpha \in (1/2, 1)$. Then,

$$\frac{a(n+1)}{a(n)}, \frac{b(n+1)}{b(n)} \rightarrow 1, \text{ as } n \rightarrow \infty, \quad (2.3)$$

$$\sum_n a(n) = \sum_n b(n) = \infty, \sum_n a(n)^2, \sum_n b(n)^2 < \infty, a(n) = o(b(n)). \quad (2.4)$$

Let $\theta(m) \triangleq (\lambda_1(m), \dots, \lambda_N(m))^T$ represent the m th update of the parameter θ . Define $\{n_m, m \geq 0\}$ as follows: $n_0 = 1$ and $n_{m+1} = \min\{j > n_m \mid \sum_{i=n_m+1}^j a(i) \geq b(m)\}$, $m \geq 1$. Let $\theta(m)$ represent the m th update of θ . Consider now the process $\{(q_j, X_j)\}$ governed by $\{\tilde{\theta}_j\}$ with $\tilde{\theta}_j = \theta(m)$ for $j = n_m, n_m + 1, \dots, n_{m+1} - 1$, $m \geq 0$. Similarly consider a parallel process $\{(\bar{q}_j^1, \bar{X}_j^1)\}$ governed by $\{\tilde{\theta}_j^1\}$ with $\tilde{\theta}_j^1 = \pi(\theta(m) + \delta)$ for $j = n_m, n_m + 1, \dots, n_{m+1} - 1$, $m \geq 0$. The two timescale algorithm of [6] then is

$$\theta(m+1) = \pi \left(\theta(m) + \sum_{j=n_m+1}^{n_{m+1}} a(j) \left(\frac{h(q_j) - h(\bar{q}_j^1)}{\delta} \right) \right). \quad (2.5)$$

We need more notation to handle the vector parameter case $\theta = (\lambda_1, \dots, \lambda_N)^T$. Let $\pi_i(\lambda)$ denote the point closest to $\lambda \in \mathcal{R}$ in the interval $[\lambda_{i,\min}, \lambda_{i,\max}] \subset \mathcal{R}$ (defined earlier) and $\pi(\theta)$ be defined by $\pi(\theta) = (\pi_1(\lambda_1), \pi_2(\lambda_2), \dots, \pi_N(\lambda_N))^T$. The vector case would ordinarily require $N + 1$ parallel simulations $\{(q_j, X_j)\}$ and $\{(\bar{q}_j^i, \bar{X}_j^i)\}$, $i = 1, \dots, N$, respectively governed by $\{\tilde{\theta}_j\}$, $\{\tilde{\theta}_j^i\}$, $i = 1, \dots, N$, where $\tilde{\theta}_j = \theta(m)$, $\tilde{\theta}_j^i = \pi(\theta(m) + \delta e_i)$, $i = 1, \dots, N$, $m \geq 0$, and where e_i is the unit vector with 1 in the i th direction. Then the algorithm is as follows: For $i = 1, \dots, N$,

$$\lambda_i(m+1) = \pi_i \left(\lambda_i(m) + \sum_{j=n_m+1}^{n_{m+1}} a(j) \left(\frac{h(q_j) - h(\bar{q}_j^i)}{\delta} \right) \right). \quad (2.6)$$

An alternative (proposed and used in [6]) to using $N + 1$ parallel simulations is to move the algorithm in cycles during each of which only two simulations are used as follows: The first simulation corresponds to $\{(q_j, X_j)\}$ and is governed by $\{\tilde{\theta}_j\}$ defined as earlier, and the second simulation is represented as $\{(\bar{q}_j, \bar{X}_j)\}$ which is governed by $\{\hat{\theta}_j\}$ defined by $\hat{\theta}_j = \pi(\theta(m) + \delta e_i)$ for $j = n_{Nm+i-1}, n_{Nm+i-1} + 1, \dots, n_{Nm+i} - 1$, $i = 1, \dots, N$, $m \geq 0$. The algorithm then is

$$\lambda_i(m+1) = \pi_i \left(\lambda_i(m) + \sum_{j=n_{Nm+i-1}+1}^{n_{Nm+i}} a(j) \left(\frac{h(q_j) - h(\bar{q}_j)}{\delta} \right) \right). \quad (2.7)$$

Thus instead of all components being updated every n_m steps, $m \geq 1$, as in (2.6), only one component is updated now every n_m steps and the algorithm thus moves in bigger loops or cycles of n_{Nm} with all components updated once at the end of the bigger loop. It is clear that one needs only two simulations in this manner but there is a tradeoff with speed of convergence. We return to this issue after we present the SPSA version of the two timescale algorithm next.

Let for any $i \geq 0$, $\Delta(i) \in \mathcal{R}^N$ be a vector of mutually independent and mean zero random variables $\{\Delta_{i,1}, \dots, \Delta_{i,N}\}$, taking values in a compact set $E \subset \mathcal{R}^N$ and having a common distribution. We assume that these random variables satisfy condition (A) below.

Condition (A) There exists a constant $\bar{K} < \infty$ such that for any $l \geq 0$ and $i \in \{1, \dots, N\}$, $E[\Delta_{l,i}^{-2}] \leq \bar{K}$.

Further, we assume that $\{\Delta(i)\}$ is a mutually independent sequence with $\Delta(i)$ independent of $\sigma(\theta(l), l \leq i)$. In what follows, we shall address the problem of minimizing the average cost $J(\theta)$ within the constraint set C . We make the following assumption about the local minima of $J(\theta)$.

Assumption (B) There exists atleast one local minima of the average cost $J(\theta)$ within the set C . Further, all local minima of $J(\theta)$ within C lie in C^0 (the interior of C).

Condition (A) is a standard condition in SPSA algorithms. Minor variants of this are for instance available in [33], [10]. Note that distributions like Gaussian and Uniform are precluded while using (A). An important consequence of $E[\Delta_{l,i}^{-2}] < \infty$ is that $P(\Delta_{l,i} = 0) = 0$.

We now proceed with our SPSA algorithm. Define parallel processes $\{(q_j^1, X_j^1)\}$ and $\{(q_j^2, X_j^2)\}$ such that for $n_m < j \leq n_{m+1}$, $\{(q_j^1, X_j^1)\}$ is governed by $\pi(\theta(m) - \delta\Delta(m)) \triangleq (\pi_1(\lambda_1(m) - \delta\Delta_{m,1}), \dots, \pi_N(\lambda_N(m) - \delta\Delta_{m,N}))^T$. Similarly, $\{(q_j^2, X_j^2)\}$ is governed by $\pi(\theta(m) + \delta\Delta(m))$ defined analogously. In the above (as mentioned earlier), $\theta(m) \triangleq (\lambda_1(m), \dots, \lambda_N(m))^T$ is the value of the parameter update that is governed by the following recursion equations. For $i = 1, \dots, N$,

$$\lambda_i(m+1) = \pi_i \left(\lambda_i(m) + \sum_{j=n_m+1}^{n_{m+1}} a(j) \left(\frac{h(q_j^1) - h(q_j^2)}{2\delta\Delta_{m,i}} \right) \right). \quad (2.8)$$

In the above, $\{n_m, m \geq 0\}$ is the same as before with $\{a(i), i \geq 0\}$, $\{b(i), i \geq 0\}$ as those defined earlier and which satisfy (2.3)-(2.4). We now discuss the reasons for the SPSA scheme in (2.8) to be computationally more efficient than both schemes (2.6) and (2.7). We begin with (2.7) first. It was shown in [6] that the scheme (2.7) tracks trajectories of an o.d.e. similar to (2.9) below but with a factor of $1/N$ multiplying the RHS of it. Also, the scheme (2.6) tracks trajectories of (2.9) as is. This means that even though the qualitative behaviour of the algorithm (2.7) is the same as that of (2.6) and also (2.8) (as shown in Appendix), the factor of $1/N$ on the RHS of (2.9) essentially serves to slow down its rate of convergence. Hence Theorem 2.1 essentially serves to indicate that we no longer need $N + 1$ parallel simulations for an N -vector parameter as (2.6) would require while at the same time we do not compromise on the speed of convergence. The gain in computational efficiency comes about because generating $N - 1$ ‘extra’ simulation samples is much more computationally expensive than generating N i.i.d. Bernoulli random variables for the process $\{(q_j, X_j)\}$. Thus in some sense the computational burden has been shifted from the numerator of the finite difference gradient estimate term on the RHS of (2.6) to the denominator of the same on the RHS of (2.8). One can also see intuitively that (2.8) updates the whole parameter vector every n_m steps, as does (2.6), which is the reason for its fast convergence (unlike (2.7)). The convergence analysis proceeds through a sequence of steps and is given in detail in the Appendix. We state here our main result, the detailed proof of which is given in the Appendix.

The ordinary differential equation (o.d.e.) technique is commonly used to prove convergence of stochastic approximation algorithms. Here, we show that the algorithm (2.8) asymptotically

converges to the stable points of the o.d.e. (2.9) below. Let $\tilde{Z}(t) \triangleq (\tilde{Z}_1(t), \dots, \tilde{Z}_N(t)) \in \mathcal{R}^N$, where $\tilde{Z}_i(t)$, $i = 1, \dots, N$, satisfy the o.d.e.

$$\dot{\tilde{Z}}_i(t) = \tilde{\pi}_i(-\nabla_i J(\tilde{Z}(t))), \quad t \geq 0, \quad \tilde{Z}(0) \in C, \quad (2.9)$$

where for any bounded, continuous, real valued function $v(\cdot)$,

$$\tilde{\pi}_i(v(y)) = \lim_{0 < \Delta \rightarrow 0} \left(\frac{\pi_i(y + \Delta v(y)) - \pi_i(y)}{\Delta} \right).$$

For $x = (x_1, \dots, x_N)$, let $\tilde{\pi}(x) = (\tilde{\pi}_1(x_1), \dots, \tilde{\pi}_N(x_N))^T$. The role played by the operator $\tilde{\pi}(\cdot)$ is in some sense to force the o.d.e. (2.9) to evolve within the constraint set C . Then the o.d.e. (2.9) has the set $K \triangleq \{\theta \in C \mid \tilde{\pi}(\nabla J(\theta)) = 0\}$ as its asymptotically stable attractor with $J(\cdot)$ itself as its strict Liapunov function. Also for $\eta > 0$, let $K^\eta = \{\theta \in C \mid \exists \theta' \in K \text{ s.t. } \|\theta - \theta'\| \leq \eta\}$ represent the set of points within a distance η of local optima.

Theorem 2.1 Given $\eta > 0$, $\exists \bar{\delta} > 0$ such that for any $\delta \in (0, \bar{\delta}]$, the algorithm (2.8) converges to K^η a.s.

Remark Note that K is the set of all critical points of (2.9), and not just the set of local minima. However, points in K that are not local minima will be unstable equilibria, and because of the presence of noise (randomness) in the algorithm, under fairly general conditions, the algorithm will converge to the η -neighborhood of $K_0 (\triangleq \text{the set of local minima of } J(\cdot)) \subset K$ (cf. pp.127-128 of [21]). Note that we assumed the cost function to be merely bounded and continuous. If on the other hand, we assume the cost function $h(\cdot)$ to be in addition convex, the average cost $J(\cdot)$ will be convex as well. Moreover, if $J(\cdot)$ is strictly convex, it will have a unique minimum, to which our algorithm will a.s. converge within an η -neighborhood.

3 Numerical Results

In this section, we provide numerical results to illustrate the two timescale SPSA scheme and to illustrate how much improvement in performance is achieved by using the structured feedback policies for ABR flow control. As mentioned earlier, flow control in ABR service requires balancing various conflicting performance criteria such as mean and variance of delay and throughput. Often this is addressed by minimizing the distance of stationary mean queue length from a given fixed constant N_0 [34], [18], [30], [5]. We adopt a similar approach here, i.e., $h(x) = |x - N_0|$, where N_0 is assumed given. In the concluding section, we shall also indicate ways to obtain an optimal such N_0 . We compare the performance of optimal structured closed loop feedback policies of type (2.2) obtained by applying the two timescale SPSA algorithm given by (2.8), with the optimal open loop policy, defined by setting $\lambda_c(n) = \lambda^*$ for all n , where λ^* is obtained by applying the two timescale algorithm given by (2.5). Note that the optimal open loop policy has a fixed rate and thus does not use any queue length information.

For the closed loop policies, we consider experiments with policies that have five and eleven parameter levels. We assume throughout that both D_b and D_f are integral multiples of T . The

form of the five level policies for obtaining $\lambda_c(n)$, is as follows.

$$\lambda_c(n) = \begin{cases} \lambda_1^* & \text{if } q_n < N_0 - 2\epsilon \\ \lambda_2^* & \text{if } N_0 - 2\epsilon \leq q_n < N_0 - \epsilon \\ \lambda_3^* & \text{if } N_0 - \epsilon \leq q_n \leq N_0 + \epsilon \\ \lambda_4^* & \text{if } N_0 + \epsilon < q_n \leq N_0 + 2\epsilon \\ \lambda_5^* & \text{if } q_n > N_0 + 2\epsilon. \end{cases} \quad (3.1)$$

In the above (as well as below), ϵ is also a given fixed constant in addition to N_0 . The eleven level policies are given by

$$\lambda_c(n) = \begin{cases} \lambda_1^* & \text{if } q_n < N_0 - 5\epsilon \\ \lambda_2^* & \text{if } N_0 - 5\epsilon \leq q_n < N_0 - 4\epsilon \\ \lambda_3^* & \text{if } N_0 - 4\epsilon \leq q_n < N_0 - 3\epsilon \\ \lambda_4^* & \text{if } N_0 - 3\epsilon \leq q_n < N_0 - 2\epsilon \\ \lambda_5^* & \text{if } N_0 - 2\epsilon \leq q_n < N_0 - \epsilon \\ \lambda_6^* & \text{if } N_0 - \epsilon \leq q_n \leq N_0 + \epsilon \\ \lambda_7^* & \text{if } N_0 + \epsilon < q_n \leq N_0 + 2\epsilon \\ \lambda_8^* & \text{if } N_0 + 2\epsilon < q_n \leq N_0 + 3\epsilon \\ \lambda_9^* & \text{if } N_0 + 3\epsilon < q_n \leq N_0 + 4\epsilon \\ \lambda_{10}^* & \text{if } N_0 + 4\epsilon < q_n \leq N_0 + 5\epsilon \\ \lambda_{11}^* & \text{if } q_n > N_0 + 5\epsilon. \end{cases} \quad (3.2)$$

In the form of the policies above, we have for simplicity chosen all the regions S_1, \dots, S_N , in the state space defined in (2.2), in terms of N_0 and ϵ alone. In the numerical experiments, we actually consider a generalization of the model in Fig.1, where rate feedback is done at instants nF_b , $n \geq 1$, for F_b a fixed multiple of T . This gives us added flexibility in studying the effect of changes in F_b in addition to those in T . The role played by F_b is in some sense that of an additional delay. The sequence of events is thus as follows: The ABR rate $\lambda_c(\cdot)$ is computed at times nT , $n \geq 1$, at the node, using feedback policies above. These rates are fed back to the source every F_b units of time. The source receives this rate information with a delay D_b and upon receiving it immediately starts sending packets with the new rate. The packets arrive at the node with a propagation delay D_f .

For the SPSA algorithm (2.8), we choose α in the definition of $\{b(i)\}$ in (2.3)-(2.4) as $\alpha = 2/3$. Thus, we have $a(0) = b(0) = 1$, $a(i) = i^{-1}$, $b(i) = i^{-2/3}$ and $\{n_m\}$ is thus obtained. Also, the random variables $\Delta_{l,i}$, $i = 1, \dots, N$; $l \geq 1$, are chosen to be i.i.d. Bernoulli distributed with $\Delta_{l,i} = \pm 1$ w.p. $1/2$, $i = 1, \dots, N$, $l \geq 1$.

The uncontrolled process is an MMPP with the underlying Markov chain chosen for simplicity to be an irreducible two state chain. To simplify the simulation code, we assume that the underlying chain undergoes state transitions every T units of time. The buffer size B is 5×10^5 . We tested our SA scheme on various combinations of the parameters D_b , D_f , N_0 , ϵ , T , F_b , $\lambda_{u,i}$, $p(i; j)$. We also conducted experiments with two controlled sources feeding into the same bottleneck node but with rate ($\lambda_c(n)$) information fed back with different delays (almost without any delay to the first and with a significant delay to the second). We observed that the bandwidth is shared equally by the two sources. This amounts to our scheme showing fairness in performance. However, we are aware of the fact that the appropriate framework to study fairness is in tandem queues with different ABR sources feeding packets through different sets of nodes [32].

Let θ^* denote the parameter value for the corresponding optimal policy, i.e., $\theta^* = (\lambda_1^*, \dots, \lambda_l^*)^T$ for the l -level closed loop policy and $\theta^* = \lambda^*$ for the open loop policy. In the following, subscript

θ^* is used in the definition of various performance measures to indicate θ^* -parametrized stationary distributions of the various quantities. Thus, $Var_{\theta^*}(q_n)$ represents the stationary variance of $\{q_n\}$ parameterized by θ^* . Let B_d represent the segment or band (of queue length values) $[N_0 - \epsilon, N_0 + \epsilon]$. We compare performance in terms of parameters of queue length distributions and throughput rate; \bar{q} , P_{band} , σ_q , $\bar{\lambda}_c$, P_{idle} and $J(\theta^*)$. These quantities and their estimates are defined as follows:

$$\begin{aligned}\bar{q} &\triangleq E_{\theta^*}[q_n] \approx \frac{1}{N} \sum_{i=0}^{N-1} q_i, \quad \sigma_q \triangleq \text{Var}_{\theta^*}(q_n) \approx \left(\frac{1}{N} \sum_{i=0}^{N-1} q_i^2 \right) - (\bar{q})^2, \\ P_{band} &\triangleq P_{\theta^*}(q_n \in B_d) \approx \frac{1}{N} \sum_{i=0}^{N-1} I\{q_i \in B_d\}, \quad P_{idle} \triangleq P_{\theta^*}(q_n = 0) \approx \frac{1}{N} \sum_{i=0}^{N-1} I\{q_i = 0\}, \\ \bar{\lambda}_c &\triangleq E_{\theta^*}[\lambda_c(n)] \approx \frac{1}{N} \sum_{i=0}^{N-1} \lambda_c(i), \quad J(\theta^*) \approx \frac{1}{N} \sum_{i=0}^{N-1} |q_i - N_0|,\end{aligned}$$

where N is taken as 10^5 in our experiments. The last performance measure is the one that the algorithm seeks to minimize, but clearly the others are closely related and are included here because they are often taken as measures of performance in ABR service. One desires P_{band} to be high in order to satisfy the various other performance criteria. One expects as a consequence of the above minimization that this quantity will be maximized. The measure P_{idle} gives the stationary probability of the server lying idle and should be close to zero. The average ABR throughput rate $\bar{\lambda}_c$ is often considered the most important measure of performance in ABR because it is this measure which tells us whether the available bandwidth has been properly utilized or not.

In the simulations for the five-level policies, $\lambda_1, \dots, \lambda_4 \in [0.10, 3.0]$ and $\lambda_5 \in [0.10, 0.90]$. For the eleven-level policies, we have $\lambda_1, \dots, \lambda_{10} \in [0.10, 3.0]$ and $\lambda_{11} \in [0.10, 0.90]$. Also, we take the service time process to be i.i.d. exponential with rate $\mu = 1.0$ and $\delta = 0.12$ in Tables 1 to 9. In the following, we considered two settings for the uncontrolled traffic: (a) $\lambda_{u,1} = 0.05$, $\lambda_{u,2} = 0.15$, $p(1;1) = p(1;2) = p(2;1) = p(2;2) = 0.5$, and (b) $\lambda_{u,1} = 0.2$, $\lambda_{u,2} = 0.4$, $p(1;1) = p(2;1) = 0.6$, $p(1;2) = p(2;2) = 0.4$. In Tables 1 - 9, the two settings are summarized by the value of $\bar{\lambda}_u$ (the mean rate of the uncontrolled MMPP) which is 0.10 and 0.28 in cases (a) and (b) respectively.

We performed a broad range of experiments for both the cases viz., $D_b = D_f = 0$ and $D_b, D_f > 0$, under non-zero T and F_b , for both the five level and eleven level policies. We show here experiments with five level policies in greater detail since the observations for experiments with eleven level policies are similar to those with five level policies. Throughout, ‘O.L.’ represents the optimal open loop policy. For $D_b = D_f = 0$, we performed various sets of experiments with fixed N_0, ϵ and the uncontrolled MMPP parameters (Tables 1 to 4 for five level and Table 8 for eleven level policies) and varying T and F_b in each. Also in Tables 6 and 9 (for the five level and eleven level policies respectively), we chose T and F_b fixed along with N_0, ϵ and the uncontrolled MMPP parameters and varied D_b and D_f . In Table 5, we study the effect of varying N_0 with all other parameters fixed. For small D_b, D_f, T and F_b , for five-level policies, our algorithm converges in about 130-150 iterations while for eleven-level policies, it takes about 150-180 iterations to converge. For large D_b, D_f, T and F_b , the algorithm takes about 200-250 iterations for five-level policies and about 220-270 iterations for eleven-level policies. It is shown in [6] (Lemma 3.1, pp.513) that $\{n_m\}$ grows exponentially as $n_{m+1} \approx \exp(am^{1/3})$ for some $a > 0$. In order to give an idea of the amount of computation required for the algorithm, the numbers n_m , $m \geq 0$, defined before (2.5)

take the following values for some integers m : $n_{100} \approx 4.3 \times 10^5$, $n_{150} \approx 3.3 \times 10^6$, $n_{200} \approx 1.6 \times 10^7$, $n_{250} \approx 6.5 \times 10^7$, $n_{300} \approx 2.1 \times 10^8$, $n_{350} \approx 6.1 \times 10^8$. On a Sun Ultra10 UNIX workstation, it takes about 5-10 minutes for five-level and 10-20 minutes for eleven-level policies for small D_b, D_f, T and F_b . For large D_b, D_f, T and F_b , it takes about 30-50 minutes for five-level policies and 40-70 minutes for eleven-level policies. We also tried running the two timescale stochastic approximation algorithm of [6], for five level policies with $D_b = D_f = 0$. It did not converge even after 350 iterations (but was close to it) after almost 200 minutes. This confirms that the SPSA version of the two timescale stochastic approximation scheme shows faster convergence than the original two timescale scheme.

For the open loop policy, the ABR throughput rate is simply the value of the parameter selected. The results indicate a significantly lower value for this performance measure compared with the closed loop policy. To get a feel for the dependence of the average cost $J(\theta)$ on the ABR rate, we varied the open loop parameter over a range for the settings in Table 1. The results are shown in Fig.2. The graph clearly shows the steep degradation in average cost when the ABR rate exceeds the O.L. value. For comparison purposes we also plot the values of the average cost obtained from the corresponding closed loop policies, where significantly higher ABR throughput is achieved along with superior average cost performance. Furthermore, a closer look at Table 1 also reveals that compared with the open loop optimal policy, the closed loop policy (for small enough F_b) leads to a stationary queue length with a significantly smaller variance while staying much closer to the target N_0 , as further supported by the values of \bar{q} and P_{band} . This comparison is indicative of the results for all of the other cases considered as well.

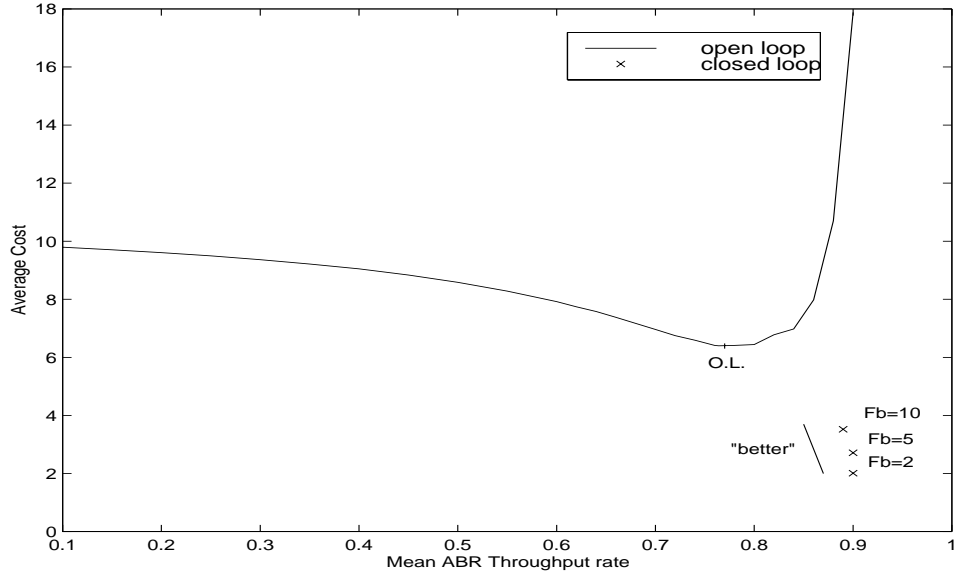
We also considered the situation in which the uncontrolled arrival process is Poisson. In this case, one can directly compute theoretically the various performance metrics for a given ABR rate (for the open loop case) since it becomes an $M/M/1/B$ queueing system now (see pp.62, [27] for the various expressions). $\bar{\lambda}_u$ for this case represents just the arrival rate of the uncontrolled Poisson stream. Here, we take $\bar{\lambda}_u = 0.10$ and $T = 1$. Note that by PASTA, the same results will hold for any T in this case (which is not so in the other cases where we have an MMPP uncontrolled stream instead of a Poisson stream). The algorithm converged in this case to $\bar{\lambda}_c = 0.74$. Fig.3 below gives the values for the various performance metrics.

Figure 2

Scheme	\bar{q}	P_{band}	σ_q	P_{idle}	$J(\theta)$
Theoretical	5.4	0.09	30.1	0.16	6.0
Algorithm	5.4	0.09	30.3	0.15	6.1

Figure 3

We now discuss the simulation results in more detail. The more important highlights are as follows:



1. The closed loop solution utilizes almost the entire bandwidth ($\bar{\lambda}_c + \bar{\lambda}_u \approx \mu$) even when D_b , D_f , T and F_b are sufficiently high.
2. The performance degrades when the delays D_b and D_f increase, but remains better than the optimal open loop case even when D_b and D_f become significantly high. For $N_0 = 10$, $\epsilon = 1$, $T = 1$ and $F_b = 2$ (Tables 6 and 9), performance is better than the optimal open loop case even for $D_b + D_f = 150$.
3. We considered the case of two controllers feeding arrivals into the same bottleneck node in addition to the uncontrolled MMPP stream (Table 7). Explicit rate information was fed back to the two sources with different delays D_{b1} and D_{b2} . Further there were different delays D_{f1} and D_{f2} in customers arriving to the bottleneck node from the two sources. We observed that the stationary mean rates $\bar{\lambda}_{c1}$ and $\bar{\lambda}_{c2}$ for the two sources are almost the same even when the difference in the delays is significantly large. This amounts to our scheme showing ‘fairness’ in performance. We also observed the other performance metrics in this case and found that the performance here is not as good as that of a single source with the lower delays (in the original setting of Fig.1) and also it is not as bad as that of a single source with the higher delays. Thus in some sense the performance of the two sources here is getting averaged, which is possibly the reason for getting fairness in performance. We did not present in Table 7 the average cost measure $J(\theta^*)$ in order to save space.
4. In Table 5, we vary N_0 and fix other parameters with $D_b = D_f = 0$, to see the effect on performance. As expected for small N_0 , $\bar{\lambda}_c$ is low and subsequently P_{idle} is high. But as N_0 increases, $\bar{\lambda}_c$ becomes high and P_{idle} becomes close to zero. In the light of this observation, we discuss in the concluding section a method of finding an optimal N_0 .

Other general observations are as follows:

1. As expected, we get the best performance for lower values of T and F_b , see Tables 1 to 4 and 8 below. For the same T , lower F_b gives better performance. Also, the difference between the

lowest and the highest rates (λ_5^* and λ_1^* for the five level case and λ_{11}^* and λ_1^* for the eleven level case respectively) decreases as T and F_b increase.

2. When the settings of the uncontrolled MMPP stream are changed such that the mean rate $\bar{\lambda}_u$ of the stream is increased, the performance degrades as is seen upon comparing values in Table 1 with the corresponding values in Table 2 for the same sets of other input data in both. It is intuitively clear that this will happen since for higher $\bar{\lambda}_u$, the controller has less control on the performance.
3. The performance improves when ϵ is increased as is seen upon comparing values in Table 1 with the corresponding ones in Table 3.
4. In Table 4, we take $N_0 = 20$, $\epsilon = 2$ with the rest of the parameters the same as in Table 2. The observations with regards T and F_b are the same here as before.

4 Conclusions and Extensions

Using a continuous time queueing framework, we studied the problem of ABR rate based flow control in the presence of information and propagation delays, by developing a numerically efficient two timescale SPSA algorithm. The convergence of this algorithm was theoretically proven, and numerical experiments were conducted to investigate the performance of the structured feedback policies. The results indicate that as expected, closed loop policies lead to a significant improvement in performance over open loop policies, for reasonable values of information and propagation delays. We considered feedback policies with five and eleven levels. It was found that the convergence time increases only marginally when the number of parameter levels is increased from five to eleven, and the scheme converges orders of magnitude faster than the original two timescale stochastic approximation scheme of [6]. We also considered experiments with two ABR sources sharing the same bottleneck node but with the two sources experiencing significantly different propagation and information delays. We found that the sources under stationarity share the bandwidth equally between them. This interesting result amounts to our scheme exhibiting fairness in performance, but further experiments on tandem queues [32] are needed to conclusively demonstrate this claim.

One natural extension of this work is to apply similar methods for selecting N_0 , which in our numerical experiments was assumed given. By incorporating N_0 into the parameter vector θ , i.e., θ is now represented as $\theta = (\lambda_1, \dots, \lambda_N, N_0)^T$, an optimal N_0 can be determined by optimizing θ as earlier. However note that if we continue with the form of the cost function that we chose earlier viz., $h(q_n) = |q_n - N_0|$, then this would in fact give rise to a family of parametrized cost functions (parametrized by N_0), which would complicate matters unnecessarily. Table 5 in our numerical experiments suggests choosing a band $[a, b]$ (depending upon acceptable levels of performance) within which one can expect N_0 to lie. One can then select a cost function that takes value zero on $[a, b]$ and increases sharply outside. The two timescale SPSA algorithm (2.8) applied to suitable parametrized policies similar to (2.2) with the parameter $\theta = (\lambda_1, \dots, \lambda_N, N_0)^T$ can then give rise to an optimal N_0 within that class of policies. As a simple example, one could consider policies of the

following type:

$$\lambda_c(n) = \begin{cases} \lambda_1 & \text{if } q_n < a \\ \lambda_2 & \text{if } a \leq q_n < N_0 \\ \lambda_3 & \text{if } N_0 \leq q_n \leq b \\ \lambda_4 & \text{if } q_n > b. \end{cases} \quad (4.1)$$

There is still one problem with this: The updates of N_0 are continuous valued now while queue length observations are discrete valued. In fact, updates of N_0 in the two parallel simulations in the two timescale SPSA scheme of (2.8) would be $N_0(n) - \delta\Delta_{5,n}$ and $N_0(n) + \delta\Delta_{5,n}$ respectively, where $\{\Delta(n)\}$, $n \geq 1$ is now defined by $\Delta(n) = (\Delta_{1,n}, \dots, \Delta_{5,n})^T$ are independent random vectors and where for each $i = 1, \dots, 5$, $\Delta_{i,n}$, $n \geq 1$, are symmetric i.i.d. Bernoulli random variables as before. However, all one needs to do is to use the integral parts of $N_0 - \delta\Delta_{5,n}$ and $N_0(n) + \delta\Delta_{5,n}$ in the policy updates in the two parallel simulations.

Finally, we mention an open problem here. The problem is to prove Theorem A.1 (see Appendix) for the system in Fig.1 with i.i.d. general service times (we assumed exponential distribution) and with a finite or an infinite buffer. The rest of the convergence analysis for such a system can be shown as remarked at the end of Theorem A.1.

Appendix: Convergence Analysis

We assume here that the service time process is i.i.d. with exponential distribution. This assumption is however only required in the proof of Theorem A.1 (below) which along with Corollary A.1 establishes the preliminary hypotheses for convergence of algorithm (2.8). The remark at the end of Theorem A.1 explains the difficulty with the general service time case.

When $D_b = D_f = 0$, the rate $\lambda_c(n)$ becomes effective in the time interval $[nT, (n+1)T)$. Then under the type of policies (2.2), for $D_b = D_f = 0$, it is clear that $\{(q_n, X_n)\}$, $n \geq 0$, is a Markov chain. When D_b, D_f are non-zero, we will assume for simplicity that $D_b + D_f = MT$ for some integer $M > 0$. Simple modifications can however take care of the case when $D_b + D_f \neq MT$ for any $M > 0$. Now, it can be seen that when $D_b + D_f = MT$ for some $M > 0$, the ABR rate $\lambda_c(n)$ computed at time nT at the node is in fact effective in the time interval $[(n+M)T, (n+M+1)T)$. Thus, in the interval $[nT, (n+1)T)$, packets from the ABR source that arrive at the node were in fact sent from the source with rate $\lambda_c(n-M)$ computed at time $(n-M)T$ at the node. For such a system it can be seen that the joint process $\{(q_n, X_n, q_{n-1}, X_{n-1}, \dots, q_{n-M}, X_{n-M})\}$, $n \geq 0$, is a Markov chain. In a related work [1], it is shown that a system as in Fig.1 with (say) $D_b + D_f = MT$, is equivalent to one with $D_f = 0$ and $D_b = MT$. We shall call any Markov process that is aperiodic, irreducible and positive recurrent as ergodic. It is easy to see that since we have a finite buffer system and because $\{X_n\}$ is ergodic, for $D_b = D_f = 0$, the joint process $\{(q_n, X_n)\}$, under policies (2.2) is ergodic. Similarly for the delayed case (when $D_b + D_f = MT$), the joint process $\{(q_n, X_n, q_{n-1}, X_{n-1}, \dots, q_{n-M}, X_{n-M})\}$, under policies (2.2) is ergodic as well. For ease of exposition, we will consider the case $D_b = D_f = 0$ in detail from now on and explain the changes necessary for nonzero D_b, D_f as we proceed. Thus for $D_b = D_f = 0$, for any given θ , $\{(q_n, X_n)\}$ is ergodic Markov with $\{\lambda_c(n)\}$ as in (2.2). Let $\mu_\theta(q, x)$ be the stationary distribution of this Markov chain on $S \times S_u$, for given $\theta \in C$. Let $\nu_\theta(q)$ be the marginal of $\mu_\theta(q, x)$ on S that corresponds to the stationary distribution of $\{q_n\}$ alone. Thus $\nu_\theta(q) = \sum_{x \in S_u} \mu_\theta(q, x)$. Let $h : S \rightarrow \mathcal{Z}^+$ be a given bounded and nonnegative cost function (\mathcal{Z}^+ is the space of nonnegative integers). The average

cost $J(\theta)$ in (2.1) is now well defined and can be further written as:

$$J(\theta) = \sum_{i \in S} h(i) \nu_\theta(i) = \sum_{i \in S} \sum_{x \in S_u} h(i) \mu_\theta(i, x). \quad (\text{A.1})$$

We now establish some preliminary hypotheses necessary to prove Theorem 2.1. Let $p_\theta(i, x; i', x')$, $i, i' \in S$, $x, x' \in S_u$ represent the transition probabilities for the Markov chain $\{(q_n, X_n)\}$ for given θ . Let D_n denote the number of departures from the queue in the time interval $[nT, (n+1)T)$, A_n^c denote the number of arrivals from the controlled source in $[nT, (n+1)T)$ and A_n^u be the number of arrivals from the uncontrolled stream during the same time interval.

Theorem A.1 Under all policies of type (2.2), $J(\theta)$ is continuously differentiable in θ .

Proof Let us first consider the case $D_b = D_f = 0$. For $J(\theta)$ to be continuously differentiable, it is enough to show that $\mu_\theta(\cdot, \cdot)$ is continuously differentiable in θ . For ease of exposition let us consider for the moment that θ is a scalar. Writing in matrix notation, let for fixed θ , $P(\theta) := [[p_\theta(i, x; j, y)]]$ be the transition probability matrix of $\{(q_n, X_n)\}$ and $\mu(\theta) := [\mu_\theta(i, x)]$ denote the vector of stationary probabilities. Also let $Z(\theta) := [I - P(\theta) - P^\infty(\theta)]^{-1}$, where I is the identity matrix and $P^\infty(\theta) = \lim_{m \rightarrow \infty} (P(\theta) + \dots + P^m(\theta))/m$. Then from Theorem 2, pp.402-403 of [29], we can write

$$\mu(\theta + h) = \mu(\theta)(I + (P(\theta + h) - P(\theta))Z(\theta) + o(h)). \quad (\text{A.2})$$

Thus,

$$\mu'(\theta) = \mu(\theta)P'(\theta)Z(\theta).$$

Hence $\mu'(\theta)$ (the derivative of $\mu(\theta)$) exists if $P'(\theta)$ (the derivative of $P(\theta)$) does. Then we have,

$$\begin{aligned} |\mu'(\theta + h) - \mu'(\theta)| &\leq |\mu(\theta + h)P'(\theta + h)Z(\theta + h) - \mu(\theta)P'(\theta + h)Z(\theta + h)| \\ &\quad + |\mu(\theta)P'(\theta + h)Z(\theta + h) - \mu(\theta)P'(\theta)Z(\theta + h)| \\ &\quad + |\mu(\theta)P'(\theta)Z(\theta + h) - \mu(\theta)P'(\theta)Z(\theta)|. \end{aligned}$$

Now, from Theorem 2, pp.402-403 of [29], we can write $Z(\theta + h)$ as

$$Z(\theta + h) = Z(\theta)H(\theta, \theta + h) - P^\infty(\theta)H(\theta, \theta + h)U(\theta, \theta + h)Z(\theta)H(\theta, \theta + h),$$

where,

$$H(\theta, \theta + h) = [I - (P(\theta + h) - P(\theta))]^{-1} \rightarrow I \quad \text{as } |h| \rightarrow 0$$

and

$$U(\theta, \theta + h) = (P(\theta + h) - P(\theta))Z(\theta) \rightarrow \bar{0} \quad \text{as } |h| \rightarrow 0.$$

In the above $\bar{0}$ is the matrix (of appropriate dimension) with all zero elements. It thus follows that $Z(\theta + h) \rightarrow Z(\theta)$ as $|h| \rightarrow 0$. Moreover from (A.2), $\mu(\theta)$ is continuous. Thus from above, $\mu'(\theta)$ is continuous in θ and the claim follows. For vector θ , a similar proof as above verifies the claim.

We finally show that $P'(\theta)$ exists and is continuous in θ . Let us first consider the case when $j \geq i$ (below). Then, the transition probability $p_\theta(i, x_1; j, x_2)$ can be written as

$$\begin{aligned} p_\theta(i, x_1; j, x_2) &= \Pr(q_{n+1} = j, X_{n+1} = x_2 \mid q_n = i, X_n = x_1, \theta) \\ &= \Pr(q_{n+1} = j \mid q_n = i, X_n = x_1, X_{n+1} = x_2, \theta) \Pr(X_{n+1} = x_2 \mid q_n = i, X_n = x_1, \theta) \end{aligned}$$

$$\begin{aligned}
&= \sum_{m=0}^{\infty} \sum_{l=0}^{j-i+m} \Pr(D_n = m, A_n^c = j - i + m - l, A_n^u = l \mid q_n = i, X_n = x_1, X_{n+1} = x_2, \theta) p(x_1; x_2) \\
&= \sum_{m=0}^{\infty} \sum_{l=0}^{j-i+m} \Pr(D_n = m, A_n^c = j - i + m - l \mid q_n = i, X_n = x_1, X_{n+1} = x_2, A_n^u = l, \theta) \times \\
&\quad \Pr(A_n^u = l \mid X_n = x_1, X_{n+1} = x_2, q_n = i, \theta) p(x_1; x_2) \\
&= \sum_{m=0}^{\infty} \sum_{l=0}^{j-i+m} \frac{(\exp(-\mu T)) (\mu T)^m (\exp(-\lambda_c(n)T)) (\lambda_c(n)T)^{j-i+m-l}}{m! (j-i+m-l)!} \Pr(A_n^u = l \mid X_n = x_1, X_{n+1} = x_2) \\
&\quad \times p(x_1; x_2).
\end{aligned}$$

Now it can be seen that the derivative of $p_\theta(i, x_1; j, x_2)$ w.r.t. $\lambda_c(n)$ exists and is continuous. Similar conclusions can be seen to hold for $j < i$. Now since $\theta = (\lambda_1, \dots, \lambda_N)^T$, for any $m = 1, \dots, N$,

$$\frac{dp_\theta(i, x_1; j, x_2)}{d\lambda_m} = \frac{\partial p_\theta(i, x_1; j, x_2)}{\partial \lambda_c(n)} \frac{d\lambda_c(n)}{d\lambda_m},$$

where $\lambda_c(n) = \lambda_m$, if $i \in S_m$, $m = 1, \dots, N$. Hence $d\lambda_c(n)/d\lambda_j = 1$, if $j = m$ (corresponding to $i \in S_m$) and is 0 otherwise. Thus the derivative of these transition probabilities w.r.t. θ exists and is continuous. The proof for the delayed case $D_b, D_f \neq 0$ follows in a similar manner. \square

Remark Let V_n represent the work load at instant nT . For a system as in Fig.1 (but) with general i.i.d. service times and an infinite buffer, under policies (2.2), $\{(V_n, X_n)\}$, for $D_b = D_f = 0$, can be shown to be ergodic Markov under an additional standard stability assumption (cf. [30]) which is not required for a finite buffer system. Similarly, $\{(V_n, X_n, V_{n-1}, X_{n-1}, \dots, V_{n-M}, X_{n-M})\}$ can be shown to be ergodic Markov for $D_b + D_f = MT$. The above Markov chains are however uncountable and Theorem 2 of [29] (which holds for a finite state system) is no longer valid. However, Corollary A.1 below can be shown quite easily for this system using sample path arguments. Moreover, the remainder of the analysis can be shown in a similar manner as follows but under an extra Liapunov stability hypothesis which in turn can be proven for our system under the extra stability assumption (mentioned above). Note however, that from a practical view point, we do not gain much by showing our results for an infinite buffer system (since we have shown experimental results for a system with large buffer). For proving Theorem A.1 for a finite buffer system with general i.i.d. service times, one could discretize the work load process and the result of [29] could still be used. However, the problem remains of showing that the transition probabilities are differentiable in the parameter. The problem of proving Theorem A.1 for i.i.d. general service times for a finite or an infinite buffer system remains an open problem.

Corollary A.1 Under all policies of type (2.2), the map $\theta \rightarrow p_\theta(i, x_1; j, x_2)$ is continuous in θ .

Proof The above map is differentiable in θ (see proof of Theorem A.1), and hence continuous. \square

We now proceed with the rest of the convergence analysis. Let $\mathcal{F}_n \triangleq \sigma(q_j^1, q_j^2, X_j^1, X_j^2, \tilde{\theta}_j, \tilde{\Delta}_j, 1 \leq j \leq n)$ represent the σ -algebra associated with information upto period nT and where $\tilde{\theta}_j = \theta(m)$ and $\tilde{\Delta}_j = \Delta(m)$, for $n_m \leq j < n_{m+1}$. Let us consider the undelayed case ($D_b = D_f = 0$) first. We shall later comment on the changes necessary for the delayed case. For any sets $A \subset S$

and $D \subset S_u$, define sequences $\{M_{1,n}(A \times D)\}$ and $\{M_{2,n}(A \times D)\}$ as follows. For $k = 1, 2$,

$$M_{k,n}(A \times D) = \sum_{m=0}^{n-1} b(m)^{-1} \left[\sum_{j=n_m+1}^{n_{m+1}} a(j) (I\{q_j^k \in A, X_j^k \in D\} - E[I\{q_j^k \in A, X_j^k \in D\} | \mathcal{F}_{j-1}]) \right],$$

$n \geq 1$, where $I\{\cdot\}$ is the indicator or characteristic function. Then one can proceed to show as in [6] that $\{M_{1,n}(A \times D)\}$ and $\{M_{2,n}(A \times D)\}$ are zero mean, square integrable martingale sequences with a.s. convergent quadratic variation processes. Thus, we have

Lemma A.1 For any $A \subset S$ and $D \subset S_u$, $\{M_{1,n}(A \times D)\}$ and $\{M_{2,n}(A \times D)\}$ converge a.s.

Proof Follows from Proposition VII.2.3(c), pp.149-150 of [23]. \square

Let $\mathcal{P}(S \times S_u)$ be the space of all probability measures on $S \times S_u$ endowed with the Prohorov topology [8]. Now for $m \geq 1$, define $\mathcal{P}(S \times S_u)$ -valued random variables $\{\mu_m^1\}$ and $\{\mu_m^2\}$ as follows: Let $A \subset S$ and $D \subset S_u$ be any two sets. Then,

$$\mu_m^k(A \times D) = \frac{\sum_{j=n_m+1}^{n_{m+1}} a(j) I\{q_j^k \in A, X_j^k \in D\}}{\sum_{j=n_m+1}^{n_{m+1}} a(j)},$$

$k = 1, 2$. For ease of exposition we assume in the following theorem that θ is a scalar. The necessary changes for vector θ are remarked at the end of it. Recall that $\mu_\theta(i, x)$ represents the invariant measure corresponding to the ergodic Markov process $\{(q_n, X_n)\}$ with parameter $\theta \in C$ (fixed), that has transition probabilities $[[p_\theta(i, x; j, y)]]$. Then $\mu_{\pi(\theta-\delta\Delta)}(i, x)$ (resp. $\mu_{\pi(\theta+\delta\Delta)}(i, x)$) is the invariant measure of the ergodic Markov process $\{(q_n, X_n)\}$ that has transition probabilities $[[p_{\pi(\theta-\delta\Delta)}(i, x; j, y)]]$ (resp. $[[p_{\pi(\theta+\delta\Delta)}(i, x; j, y)]]$).

Theorem A.2 Almost surely, $(\mu_m^1, \mu_m^2, \theta(m), \Delta(m))$, $m \geq 0$ converges in $\mathcal{P}(S \times S_u) \times \mathcal{P}(S \times S_u) \times C \times E$ to the compact set $\{(\mu_{\pi(\theta-\delta\Delta)}, \mu_{\pi(\theta+\delta\Delta)}, \theta, \Delta) \mid \theta \in C, \Delta \in E\}$.

Proof Let us begin with the undelayed case ($D_b = D_f = 0$) and consider $\{M_{1,n}(A \times D)\}$ first. From Lemma A.1 and the fact that $\sum_{j=n_m+1}^{n_{m+1}} a(j)/b(m) \rightarrow 1$ as $m \rightarrow \infty$, one has

$$\frac{\sum_{j=n_m+1}^{n_{m+1}} a(j) [I\{q_j^1 \in A, X_j^1 \in D\} - E[I\{q_j^1 \in A, X_j^1 \in D\} | \mathcal{F}_{j-1}]]}{\sum_{j=n_m+1}^{n_{m+1}} a(j)} \rightarrow 0 \text{ a.s.},$$

for any $A \subset S$ and $D \subset S_u$. Hence,

$$\frac{\sum_{j=n_m+1}^{n_{m+1}} a(j) [I\{q_j^1 \in A, X_j^1 \in D\} - \sum_{i \in S} \sum_{x \in S_u} p_{\pi(\theta(m)-\delta\Delta(m))}(i, x; A, D) I\{q_{j-1}^1 = i, X_{j-1}^1 = x\}]}{\sum_{j=n_m+1}^{n_{m+1}} a(j)} \rightarrow 0 \text{ a.s.} \tag{A.3}$$

Now any limit point of $\{(\mu_m^1, \pi(\theta(m) - \delta\Delta(m)))\}$ must be of the form $(\mu, \pi(\theta - \delta\Delta))$. From (A.3), Corollary A.1, the definition of $\{\mu_m^1\}$ and (2.3), it follows that μ must satisfy

$$\mu(A \times D) = \sum_{i \in S} \sum_{x \in S_u} p_{\pi(\theta-\delta\Delta)}(i, x; A, D) \mu(i, x).$$

Thus $\mu = \mu_{\pi(\theta-\delta\Delta)}$. An analogous argument applies to $\{\mu_m^2\}$. The claim now follows from the fact that the continuous image of a compact set is compact. Let us now consider the delayed case

with $D_b + D_f = MT$ for some integer $M \geq 1$. Then $\{Y_j^1\}$ and $\{Y_j^2\}$ defined by $Y_j^1 \triangleq (q_j^1, X_j^1, q_{j-1}^1, X_{j-1}^1, \dots, q_{j-M}^1, X_{j-M}^1)$ and $Y_j^2 \triangleq (q_j^2, X_j^2, q_{j-1}^2, X_{j-1}^2, \dots, q_{j-M}^2, X_{j-M}^2)$, $j \geq 1$, are Markov chains such that for any $j \geq 1$, Y_j^1 and Y_j^2 are governed by parameters $\pi(\tilde{\theta}_j - \delta\tilde{\Delta}_j)$ and $\pi(\tilde{\theta}_j + \delta\tilde{\Delta}_j)$ respectively. Consider now for any set A in the new state space $(S^M \times S_u^M)$, sequences $\{\bar{\mu}_{1,n}(A)\}$ and $\{\bar{\mu}_{2,n}(A)\}$ defined by

$$\bar{\mu}_{k,n}(A) = \sum_{m=0}^{n-1} b(m)^{-1} \left[\sum_{j=n_m+1}^{n_{m+1}} a(j) (I\{Y_j^k \in A\} - E[I\{Y_j^k \in A\} | \mathcal{F}_{j-M-1}]) \right],$$

$k = 1, 2$, respectively. Note that $\bar{\mu}_{k,n}(A)$ can be written as

$$\begin{aligned} \bar{\mu}_{k,n}(A) &= \sum_{m=0}^{n-1} b(m)^{-1} \left[\sum_{j=n_m+1}^{n_{m+1}} a(j) (I\{Y_j^k \in A\} - E[I\{Y_j^k \in A\} | \mathcal{F}_{j-1}]) \right] + \\ &\quad \sum_{m=0}^{n-1} b(m)^{-1} \left[\sum_{j=n_m+1}^{n_{m+1}} a(j) (E[I\{Y_j^k \in A\} | \mathcal{F}_{j-1}] - E[I\{Y_j^k \in A\} | \mathcal{F}_{j-2}]) \right] + \dots + \\ &\quad \sum_{m=0}^{n-1} b(m)^{-1} \left[\sum_{j=n_m+1}^{n_{m+1}} a(j) (E[I\{Y_j^k \in A\} | \mathcal{F}_{j-M}] - E[I\{Y_j^k \in A\} | \mathcal{F}_{j-M-1}]) \right], \end{aligned}$$

$k = 1, 2$. Let us represent for any $k \in \{1, 2\}$ fixed, each of the $M + 1$ individual summations on the RHS above by $M_1^k(n), \dots, M_{M+1}^k(n)$. Then it is easy to see that $(M_1^k(m), \mathcal{F}_{n_m}), (M_2^k(m), \mathcal{F}_{n_{m-1}}), \dots, (M_{M+1}^k(m), \mathcal{F}_{n_{m-M}})$ are martingale sequences each of which can be shown to converge by showing that their quadratic variation processes are a.s. convergent in a manner as earlier. Then, we also have

$$\frac{\sum_{j=n_m+1}^{n_{m+1}} a(j) [I\{Y_j^k \in A\} - E[I\{Y_j^k \in A\} | \mathcal{F}_{j-M-1}]]}{\sum_{j=n_m+1}^{n_{m+1}} a(j)} \rightarrow 0 \text{ a.s.},$$

and similar conclusions as for the undelayed case can be obtained in the same manner as before. \square

Corollary A.2 We have

$$\left| \frac{\sum_{j=n_m+1}^{n_{m+1}} a(j) h(q_j^1)}{b(m)} - J(\pi(\theta(m) - \delta\Delta(m))) \right| \rightarrow 0 \text{ a.s.},$$

and

$$\left| \frac{\sum_{j=n_m+1}^{n_{m+1}} a(j) h(q_j^2)}{b(m)} - J(\pi(\theta(m) + \delta\Delta(m))) \right| \rightarrow 0 \text{ a.s.},$$

as $m \rightarrow \infty$.

Proof Immediate from Theorem A.2, the definition of $J(\theta)$ and the fact that $\sum_{j=n_m+1}^{n_{m+1}} a(j)/b(m) \rightarrow 1$, as $m \rightarrow \infty$. \square

The proof of Theorem 2.1 proceeds through a sequence of steps. We go through these in detail. Let for $m \geq 1$, $\mathcal{F}'_m \triangleq \sigma(\theta(0), \theta(1), \dots, \theta(m), \Delta(0), \Delta(1), \dots, \Delta(m-1))$. Then $\Delta(m)$ is independent of \mathcal{F}'_m , $\forall m \geq 1$. Define sequences $\{N_i^1(p), p \geq 1\}$, $\{N_i^2(p), p \geq 1\}$; $i = 1, \dots, N$, as follows:

$$N_i^1(p) = \sum_{j=0}^p b(j) \left(\frac{J(\pi(\theta(j) - \delta\Delta(j)))}{\Delta_{j,i}} - E \left[\frac{J(\pi(\theta(j) - \delta\Delta(j)))}{\Delta_{j,i}} \mid \mathcal{F}'_j \right] \right),$$

and

$$N_i^2(p) = \sum_{j=0}^p b(j) \left(\frac{J(\pi(\theta(j) + \delta\Delta(j)))}{\Delta_{j,i}} - E \left[\frac{J(\pi(\theta(j) + \delta\Delta(j)))}{\Delta_{j,i}} \mid \mathcal{F}'_j \right] \right).$$

Then, we have

Lemma A.2 For every $i = 1, \dots, N$, $\{N_i^1(p)\}$ and $\{N_i^2(p)\}$ converge a.s.

Proof We prove here the convergence of $\{N_i^1(p)\}$, $i = 1, \dots, N$. The proof of $\{N_i^2(p)\}$, $i = 1, \dots, N$, is similar. Note that for every $i = 1, \dots, N$, $\{N_i^1(p), \mathcal{F}'_{p+1}\}$ are zero mean martingales. Let $\{\langle N_i^1 \rangle(p)\}$, $i = 1, \dots, N$, represent their quadratic variation processes. Then by definition, for $i = 1, \dots, N$,

$$\begin{aligned} \langle N_i^1 \rangle(p) &= \sum_{j=0}^p E \left[(N_i^1(j+1) - N_i^1(j))^2 \mid \mathcal{F}'_p \right] + E \left[(N_i^1(0))^2 \right] \\ &= \sum_{j=0}^p E \left[(b(j))^2 \left(\frac{J(\pi(\theta(j) - \delta\Delta(j)))}{\Delta_{j,i}} - E \left[\frac{J(\pi(\theta(j) - \delta\Delta(j)))}{\Delta_{j,i}} \mid \mathcal{F}'_j \right] \right)^2 \mid \mathcal{F}'_p \right] \\ &\quad + E \left[(N_i^1(0))^2 \right]. \end{aligned}$$

Also,

$$\langle N_i^1 \rangle(\infty) = \lim_{p \rightarrow \infty} \langle N_i^1 \rangle(p).$$

Now recall that $\sum_{j=0}^{\infty} b(j)^2 < \infty$ (cf. (2.4)), $E \left[\Delta_{j,i}^{-2} \right] \leq \bar{K} < \infty$, $i = 1, \dots, N$; $j \geq 1$ (Condition (A))

and $J(\cdot) < \infty$ since the cost function $h(\cdot)$ is bounded. Thus $\langle N_i^1 \rangle(\infty) < \infty$, $i = 1, \dots, N$. Now, by Proposition VII.2.3(c), pp.149-150 of [23], the claim follows. \square

As a consequence of Lemma A.1, we have

$$\begin{aligned} &\sum_{j=0}^{\infty} b(j) \left(\frac{J(\pi(\theta(j) - \delta\Delta(j))) - J(\pi(\theta(j) + \delta\Delta(j)))}{\Delta_{j,i}} \right. \\ &\quad \left. - E \left[\frac{J(\pi(\theta(j) - \delta\Delta(j))) - J(\pi(\theta(j) + \delta\Delta(j)))}{\Delta_{j,i}} \mid \mathcal{F}'_j \right] \right) < \infty \text{ a.s.} \end{aligned}$$

We shall use a key result from [16] stated as Lemma A.3 below. Consider an o.d.e. in \mathcal{R}^N

$$\dot{x}(t) = F(x(t)), \tag{A.4}$$

which has an asymptotically stable attracting set \bar{G} . Let \bar{G}^ϵ denote the ϵ -neighborhood of \bar{G} . For $T > 0$, $\gamma > 0$, say that $y(\cdot)$ is a (T, γ) -perturbation of (A.4) if there exist $0 = T_0 < T_1 < T_2 < \dots$, such that $T_{i+1} - T_i \geq T$, $\forall i$, and on each interval $[T_i, T_{i+1}]$, there exists a solution $x^i(\cdot)$ of (A.4) such that

$$\sup_{t \in [T_i, T_{i+1}]} |x^i(t) - y(t)| < \gamma.$$

The following result is adapted from [16], pp.339. The proof of this can be found in the appendix of [9].

Lemma A.3 For given $\epsilon > 0$, $T > 0$, there exists a $\bar{\gamma}$ such that for all $\gamma \in [0, \bar{\gamma}]$, any (T, γ) -perturbation of (A.4) converges to \bar{G}^ϵ . \square

Now let $t(0) = 0$, $t(n) = \sum_{i=1}^n b(i)$, $n \geq 0$. Define $\tilde{\theta}(\cdot) \triangleq (\tilde{\theta}_1(\cdot), \dots, \tilde{\theta}_N(\cdot))^T : \mathcal{R}^+ \rightarrow C \subset \mathcal{R}^N$ by $\tilde{\theta}(t(n)) = \theta(n) \triangleq (\lambda_1(n), \dots, \lambda_N(n))^T$, $n \geq 0$, with linear interpolation on intervals $\{[t(n), t(n+1)]\}$. Let $\Delta_{t,i} = \Delta_{n,i}$, for $t \in [t(n), t(n+1)]$, $n \geq 1$. Consider the following o.d.e. in C : For $i = 1, \dots, N$,

$$\dot{\theta}_i(t) = \tilde{\pi}_i \left(E \left[\frac{J(\pi(\theta(t) - \delta\Delta_{t,i})) - J(\pi(\theta(t) + \delta\Delta_{t,i}))}{2\delta\Delta_{t,i}} \right] \right), \quad (\text{A.5})$$

where the operator $E[\cdot]$ represents the expectation w.r.t. the common c.d.f. of $\{\Delta_{t,i}\}$. We then have

Lemma A.4 For any $T, \gamma > 0$, $\tilde{\theta}(t(n) + \cdot)$ is a bounded (T, γ) -perturbation of (A.5) for sufficiently large n .

Proof Rewrite the SPSA algorithm (2.8) as follows: For $i = 1, \dots, N$,

$$\begin{aligned} \lambda_i(m+1) = \pi_i(\lambda_i(m) + b(m) E \left[\frac{J(\pi(\theta(m) - \delta\Delta(m))) - J(\pi(\theta(m) + \delta\Delta(m)))}{2\delta\Delta_{m,i}} \mid \mathcal{F}'_m \right] \\ + \eta_1(m) + \eta_2(m)), \end{aligned} \quad (\text{A.6})$$

where,

$$\begin{aligned} \eta_1(m) = b(m) \left(\frac{J(\pi(\theta(m) - \delta\Delta(m))) - J(\pi(\theta(m) + \delta\Delta(m)))}{2\delta\Delta_{m,i}} \right. \\ \left. - E \left[\frac{J(\pi(\theta(m) - \delta\Delta(m))) - J(\pi(\theta(m) + \delta\Delta(m)))}{2\delta\Delta_{m,i}} \mid \mathcal{F}'_m \right] \right), \end{aligned}$$

and

$$\eta_2(m) = b(m) \left(\frac{\sum_{j=n_{m+1}}^{n_{m+1}} a(j) \left(\frac{h(q_j^1) - h(q_j^2)}{2\delta\Delta_{m,i}} \right)}{b(m)} - \frac{J(\pi(\theta(m) - \delta\Delta(m))) - J(\pi(\theta(m) + \delta\Delta(m)))}{2\delta\Delta_{m,i}} \right).$$

Now both $\eta_1(m)$ and $\eta_2(m)$ become asymptotically negligible as $m \rightarrow \infty$ by Lemma A.2 and Corollary A.2 respectively. The algorithm (A.6) can then be viewed as a discretization of the o.d.e. (A.5) except that (as already mentioned) it has in addition asymptotically diminishing error terms $\eta_1(m)$ and $\eta_2(m)$. Now a standard argument as in pp.191-194 of [20] proves the claim. \square

Recall that $\nabla J(\theta) \triangleq (\nabla_1 J(\theta), \dots, \nabla_N J(\theta))^T$ represents the gradient of $J(\theta)$, where $\nabla_i J(\theta)$ represents the i th partial derivative. Also, C^0 represents the interior of the set C .

Lemma A.5 For any $\theta(m) \in C^0$, for all $i = 1, \dots, N$,

$$\lim_{\delta \downarrow 0} \left| E \left[\frac{J(\pi(\theta(m) - \delta\Delta(m))) - J(\pi(\theta(m) + \delta\Delta(m)))}{2\delta\Delta_{m,i}} \mid \mathcal{F}'_m \right] - \nabla_i J(\theta(m)) \right| = 0.$$

Proof Since $\theta(m) \in C^0$, choose $\delta > 0$ small enough such that for all $i = 1, \dots, N$, $\pi(\theta(m) - \delta\Delta_{m,i}) = \theta(m) - \delta\Delta_{m,i}$ and $\pi(\theta(m) + \delta\Delta_{m,i}) = \theta(m) + \delta\Delta_{m,i}$ respectively. Now an application of Taylor series on $J(\theta(m) + \delta\Delta(m))$ around the point $\theta(m)$ gives

$$J(\theta(m) + \delta\Delta(m)) = J(\theta(m)) + \delta \sum_{j=1}^N \Delta_{m,j} \nabla_j J(\theta(m)) + O(\delta^2 \sum_{j=1}^N \Delta_{m,j}^2).$$

Similarly, one obtains

$$J(\theta(m) - \delta\Delta(m)) = J(\theta(m)) - \delta \sum_{j=1}^N \Delta_{m,j} \nabla_j J(\theta(m)) + O(\delta^2 \sum_{j=1}^N \Delta_{m,j}^2).$$

From the above two equations it follows that

$$\begin{aligned} E \left[\frac{J(\theta(m) + \delta\Delta(m)) - J(\theta(m) - \delta\Delta(m))}{2\delta\Delta_{m,i}} \mid \mathcal{F}'_m \right] &= E \left[\frac{\sum_{j=1}^N \Delta_{m,j} \nabla_j J(\theta(m))}{\Delta_{m,i}} \mid \mathcal{F}'_m \right] \\ &\quad + O \left(\delta E \left[\frac{\sum_{j=1}^N \Delta_{m,j}^2}{\Delta_{m,i}} \mid \mathcal{F}'_m \right] \right) \\ &= \nabla_i J(\theta(m)) + \sum_{j=1, j \neq i}^N \nabla_j J(\theta(m)) E \left[\frac{\Delta_{m,j}}{\Delta_{m,i}} \right] + O \left(\delta E \left[\frac{\sum_{j=1}^N \Delta_{m,j}^2}{\Delta_{m,i}} \right] \right). \end{aligned}$$

The last step follows from the fact that $\theta(m)$ is measurable w.r.t. \mathcal{F}'_m and $\Delta_{m,i}$, $i = 1, \dots, N$, are independent of \mathcal{F}'_m . Now, $E \left[\frac{\Delta_{m,j}}{\Delta_{m,i}} \right] = 0$, for $j \neq i$, since for any m , $\{\Delta_{m,k}\}$, $k = 1, \dots, N$, are zero mean, independent random variables satisfying Condition (A). Further, note that $E \left[\Delta_{m,k}^2 \right] < \infty$ (since $\{\Delta_{m,k}\}$ takes values in a compact set E) and $E \left[|\Delta_{m,k}|^{-1} \right]$ is uniformly bounded (by Condition (A)). We thus have

$$E \left[\frac{J(\theta(m) + \delta\Delta(m)) - J(\theta(m) - \delta\Delta(m))}{2\delta\Delta_{m,i}} \mid \mathcal{F}'_m \right] = \nabla_i J(\theta(m)) + O(\delta).$$

The claim now follows. \square

Consider finally the o.d.e. (2.9). Recall that $K \triangleq \{\theta \in C \mid \tilde{\pi}(\nabla J(\theta)) = 0\}$ is the asymptotically stable attractor set for the o.d.e. (2.9) with $J(\cdot)$ itself serving as the strict Liapunov function. Also, $K^\eta \triangleq \{\theta \in C \mid \exists \theta' \in K \text{ s.t. } \|\theta - \theta'\| \leq \eta\}$ represents the set of points that are within an η -distance from K . Let $\eta_0 > 0$ be such that $K^{\eta_0} \subset C^0$ (interior of C). Note that this is possible by Assumption (B). We now have

Lemma A.6 Given $0 < \eta < \eta_0$, there exists a $\delta_0 > 0$ such that for $\delta \in (0, \delta_0)$, K^η is an asymptotically stable attracting set for the o.d.e. (A.5).

Proof Note that as a consequence of Assumption (B) and Lemmas A.4 and A.5, for sufficiently small δ , $J(\cdot)$ will serve as a strict Liapunov function for (A.5) outside the set K^η , for $\eta \leq \eta_0$. \square

We finally come to the proof of Theorem 2.1.

Proof of Theorem 2.1 Follows directly from Lemmas A.3, A.4 and A.6. \square

Five Level Policies with $D_b = D_f = 0$

Table.1: $N_0 = 10, \epsilon = 1, \bar{\lambda}_u = 0.1$

T	F_b	λ_5^*	λ_4^*	λ_3^*	λ_2^*	λ_1^*	\bar{q}	P_{band}	σ_q	λ_c	P_{idle}	$J(\theta^*)$
1	2	0.1	0.3	0.7	1.7	2.6	10.1	0.45	7.3	0.90	0	2.0
1	5	0.2	0.5	0.8	1.3	1.8	9.5	0.34	12.8	0.90	0	2.7
1	10	0.3	0.6	0.9	1.2	1.5	10.0	0.26	21.3	0.89	0.01	3.5
1	O.L.	-	-	-	-	-	5.3	0.08	32.3	0.76	0.15	6.4
5	10	0.3	0.7	0.9	1.1	1.4	9.8	0.26	21.6	0.89	0.02	3.6
5	25	0.4	0.8	0.8	0.9	1.1	8.2	0.17	34.1	0.82	0.07	4.8
5	50	0.3	0.4	0.6	0.9	0.9	6.4	0.12	38.3	0.75	0.16	6.0
5	O.L.	-	-	-	-	-	5.1	0.09	31.9	0.75	0.16	6.4

Table.2: $N_0 = 10, \epsilon = 1, \bar{\lambda}_u = 0.28$

T	F_b	λ_5^*	λ_4^*	λ_3^*	λ_2^*	λ_1^*	\bar{q}	P_{band}	σ_q	λ_c	P_{idle}	$J(\theta^*)$
1	2	0.1	0.2	0.5	2.0	2.5	10.2	0.44	7.9	0.72	0	2.2
1	5	0.1	0.2	0.7	1.1	1.6	10.4	0.33	13.2	0.72	0	2.7
1	10	0.1	0.5	0.7	0.9	1.3	9.9	0.26	20.5	0.71	0.01	3.5
1	O.L.	-	-	-	-	-	5.3	0.08	33.2	0.60	0.13	6.5
5	10	0.1	0.5	0.7	0.9	1.3	9.2	0.26	20.5	0.70	0.02	3.6
5	25	0.3	0.6	0.7	0.8	1.0	7.3	0.16	27.9	0.64	0.07	4.8
5	50	0.3	0.6	0.6	0.7	0.8	7.3	0.12	39.4	0.59	0.14	5.6
5	O.L.	-	-	-	-	-	5.3	0.09	31.1	0.58	0.15	6.4

Table.3: $N_0 = 10, \epsilon = 2, \bar{\lambda}_u = 0.1$

T	F_b	λ_5^*	λ_4^*	λ_3^*	λ_2^*	λ_1^*	\bar{q}	P_{band}	σ_q	λ_c	P_{idle}	$J(\theta^*)$
1	2	0.1	0.1	0.8	2.1	2.7	10.0	0.64	7.7	0.90	0	2.2
1	5	0.1	0.3	0.8	1.6	2.2	9.7	0.52	12.7	0.89	0	2.7
1	10	0.1	0.5	0.9	1.3	1.6	9.8	0.43	19.8	0.88	0.01	3.4
1	O.L.	-	-	-	-	-	5.3	0.14	32.2	0.77	0.13	6.4
5	10	0.1	0.5	0.9	1.3	1.6	10.0	0.42	20.6	0.87	0.01	3.5
5	25	0.4	0.7	0.8	1.0	1.2	9.1	0.30	36.3	0.85	0.05	4.7
5	50	0.3	0.6	0.8	1.0	1.0	9.3	0.24	39.8	0.82	0.11	5.7
5	O.L.	-	-	-	-	-	5.2	0.14	30.0	0.76	0.14	6.4

Table.4: $N_0 = 20, \epsilon = 2, \bar{\lambda}_u = 0.28$

T	F_b	λ_5^*	λ_4^*	λ_3^*	λ_2^*	λ_1^*	\bar{q}	P_{band}	σ_q	λ_c	P_{idle}	$J(\theta^*)$
1	2	0.1	0.1	0.5	1.9	2.9	19.3	0.63	8.0	0.72	0	2.2
1	5	0.1	0.2	0.6	1.5	2.1	19.9	0.51	14.2	0.71	0	2.9
1	10	0.1	0.4	0.8	1.2	1.5	20.9	0.41	21.8	0.71	0	3.6
1	O.L.	-	-	-	-	-	7.4	0.05	58.0	0.63	0.09	11.8
5	10	0.1	0.4	0.7	1.0	1.5	20.2	0.40	24.3	0.71	0	3.8
5	25	0.3	0.6	0.6	1.0	1.2	19.0	0.26	58.7	0.71	0	5.7
5	50	0.4	0.4	0.7	1.2	1.1	21.1	0.18	137.2	0.69	0.03	8.6
5	O.L.	-	-	-	-	-	7.1	0.05	51.6	0.64	0.10	12.0

Table.5: $T = 1, F_b = 2, \epsilon = 1, \bar{\lambda}_u = 0.1$

N_0	λ_5^*	λ_4^*	λ_3^*	λ_2^*	λ_1^*	\bar{q}	P_{band}	σ_q	λ_c	P_{idle}	$J(\theta^*)$
2	0.1	0.1	0.5	1.2	2.6	1.5	0.55	2.3	0.55	0.34	1.3
3	0.1	0.1	0.5	1.6	1.8	2.7	0.51	4.0	0.74	0.14	1.6
4	0.1	0.2	0.6	1.2	2.0	3.8	0.48	4.9	0.83	0.07	1.8
5	0.1	0.3	0.6	1.9	2.2	4.9	0.46	5.8	0.85	0.04	1.9
6	0.1	0.4	0.6	1.8	2.3	5.7	0.46	6.4	0.87	0.02	2.0
7	0.1	0.2	0.7	1.6	2.5	6.9	0.46	6.5	0.89	0.01	2.0
8	0.1	0.3	0.8	1.5	2.6	8.2	0.44	6.9	0.89	0.01	2.0
9	0.1	0.3	0.8	1.3	2.6	8.9	0.45	7.2	0.90	0	2.0
10	0.1	0.3	0.7	1.7	2.6	10.1	0.45	7.3	0.90	0	2.0
11	0.1	0.3	0.8	1.4	2.6	11.4	0.43	7.4	0.90	0	2.1

Five Level Policies with $D_b, D_f > 0$ **Table.6:** $T = 1, F_b = 2, N_0 = 10, \epsilon = 1, \bar{\lambda}_u = 0.1$

D_b	D_f	λ_5^*	λ_4^*	λ_3^*	λ_2^*	λ_1^*	\bar{q}	P_{band}	σ_q	λ_c	P_{idle}	$J(\theta^*)$
1	0	0.1	0.3	0.7	1.8	2.2	10.1	0.38	9.7	0.89	0	2.4
1	1	0.1	0.4	0.8	1.3	2.2	9.8	0.38	9.7	0.88	0	2.4
5	5	0.2	0.7	0.9	1.1	1.5	9.7	0.26	20.0	0.88	0.01	3.6
10	10	0.4	0.7	0.9	1.0	1.2	8.8	0.20	28.0	0.87	0.03	4.5
20	10	0.4	0.5	0.6	0.6	1.0	8.4	0.14	39.5	0.84	0.06	5.2
20	40	0.5	0.8	0.9	1.1	1.0	8.2	0.13	44.0	0.80	0.06	5.4
30	20	0.4	0.6	0.8	0.8	0.9	8.1	0.13	48.0	0.79	0.08	5.7
40	10	0.5	0.6	0.8	0.8	0.9	6.4	0.12	52.7	0.78	0.12	5.8
40	30	0.5	0.5	0.6	0.9	0.9	6.9	0.12	56.4	0.78	0.12	5.9
50	50	0.5	0.5	0.9	0.9	0.9	7.1	0.12	65.2	0.78	0.10	6.0
50	100	0.5	0.8	0.8	0.8	0.9	6.7	0.11	59.5	0.77	0.13	6.1
O.L.	-	-	-	-	-	-	5.3	0.08	32.3	0.76	0.15	6.4

Table.7 (Two ABR Sources): $T = 1, F_b = 2, N_0 = 10, \epsilon = 1, \bar{\lambda}_u = 0.1$

D_{b1}	D_{b2}	D_{f1}	D_{f2}	λ_5^*	λ_4^*	λ_3^*	λ_2^*	λ_1^*	\bar{q}	P_{band}	σ_q	λ_{c1}	λ_{c2}	P_{idle}
1	10	4	10	0.1	0.1	0.3	0.6	0.8	8.6	0.26	18.1	0.44	0.45	0.02
1	30	4	10	0.1	0.2	0.3	0.7	0.7	8.9	0.23	29.4	0.44	0.44	0.02
1	50	4	20	0.1	0.2	0.3	0.5	0.7	8.1	0.21	29.1	0.42	0.43	0.03
1	80	4	100	0.1	0.1	0.6	0.5	0.6	8.7	0.18	35.9	0.41	0.42	0.04

Eleven Level Policies with $D_b = D_f = 0$

Table.8: $N_0 = 10, \epsilon = 1, \bar{\lambda}_u = 0.1$

T	F_b	λ_{11}^*	λ_{10}^*	λ_9^*	λ_8^*	λ_7^*	λ_6^*	λ_5^*	λ_4^*	λ_3^*	λ_2^*	λ_1^*
1	2	0.1	0.1	0.1	0.4	0.7	0.9	0.9	1.2	1.2	1.3	2.7
1	5	0.1	0.1	0.1	0.3	0.4	0.8	1.2	1.5	1.9	2.1	2.4
1	10	0.1	0.3	0.4	0.5	0.6	0.9	1.1	1.2	1.3	1.5	1.7
1	O.L.	-	-	-	-	-	-	-	-	-	-	-
5	10	0.1	0.3	0.4	0.5	0.7	0.9	1.1	1.2	1.3	1.5	1.6
5	25	0.3	0.6	0.7	0.7	0.7	0.9	1.0	1.0	1.1	1.0	1.2
5	50	0.4	0.7	0.6	0.4	0.6	0.8	0.9	0.9	0.9	0.9	1.0
5	O.L.	-	-	-	-	-	-	-	-	-	-	-

Table.8 (Contd.)

T	F_b	\bar{q}	P_{band}	λ_c	σ_q	P_{idle}	$J(\theta)$
1	2	9.4	0.45	0.91	7.1	0	2.4
1	5	10.1	0.35	0.90	11.4	0	2.6
1	10	9.1	0.27	0.90	17.6	0.01	3.3
1	O.L.	5.3	0.08	32.3	0.76	0.15	6.4
5	10	9.6	0.27	0.90	19.2	0.01	3.4
5	25	9.4	0.18	0.88	36.1	0.04	4.6
5	50	7.2	0.12	0.82	34.2	0.10	5.5
5	O.L.	5.1	0.09	0.75	31.9	0.16	6.4

Eleven Level Policies with $D_b, D_f > 0$

Table.9: $T = 1, F_b = 2, N_0 = 10, \epsilon = 1, \bar{\lambda}_u = 0.1$

D_b	D_f	λ_{11}^*	λ_{10}^*	λ_9^*	λ_8^*	λ_7^*	λ_6^*	λ_5^*	λ_4^*	λ_3^*	λ_2^*	λ_1^*
5	5	0.1	0.1	0.4	0.5	0.7	0.8	1.4	1.5	1.0	1.1	1.9
10	20	0.2	0.5	0.5	0.6	0.6	0.8	0.7	1.4	1.1	1.4	1.4
20	30	0.3	0.6	0.6	1.0	0.9	0.8	0.9	1.0	1.0	1.1	1.0
30	40	0.2	0.3	0.6	0.7	0.7	0.8	0.9	1.0	0.9	1.1	1.0
40	50	0.2	0.5	0.5	0.6	0.6	0.6	0.9	1.0	1.1	0.9	0.9
50	100	0.5	0.7	0.7	1.0	0.8	0.6	1.0	0.8	1.0	0.9	0.9
O.L.	-	-	-	-	-	-	-	-	-	-	-	-

Table.9 (Contd.)

D_b	D_f	\bar{q}	P_{band}	$\bar{\lambda}_c$	σ_q	P_{idle}	$J(\theta)$
5	5	9.8	0.26	0.90	19.5	0.01	3.6
10	20	9.3	0.20	0.87	29.2	0.03	4.4
20	30	8.6	0.15	0.85	42.7	0.06	5.3
30	40	7.8	0.12	0.81	47.5	0.11	5.8
40	50	7.3	0.11	0.78	51.1	0.12	6.0
50	100	6.4	0.10	0.78	69.1	0.12	6.0
O.L.	-	5.3	0.08	0.76	32.3	0.15	6.4

References

- [1] E. Altman, T. Basar, and R. Srikant. Robust rate control for ABR sources. In *Proceedings INFOCOM 98, San Francisco, CA*, March 1998.
- [2] J. D. Bartusek and A. M. Makowski. On stochastic approximation driven by sample averages: convergence results via the ODE method. Technical Report, Institute for Systems Research, University of Maryland, http://www.isr.umd.edu/TechReports/ISR/1994/TR_94-4/, 1994.
- [3] L. Benmohamed and S. M. Meerkov. Feedback control of congestion in packet switching networks: the case of a single congested node. *IEEE/ACM Trans. Network.*, 1(6):693–707, 1993.
- [4] D. P. Bertsekas and J. N. Tsitsiklis. *Neuro Dynamic Programming*. Athena Scientific, Belmont, 1996.
- [5] S. Bhatnagar. Multiscale stochastic approximation schemes with applications to ABR service in ATM networks. Doctoral dissertation, Dept. of Electrical Engineering, Indian Institute of Science, Bangalore, India, July 1997.
- [6] S. Bhatnagar and V. S. Borkar. Multiscale stochastic approximation for parametric optimization of hidden Markov models. *Probability in the Engineering and Informational Sciences*, 11:509–522, 1997.
- [7] F. Bonomi and K. W. Fendick. The rate based flow control framework for the available bit rate ATM service. *IEEE Network*, pages 25–39, 1995.
- [8] V. S. Borkar. *Probability Theory: An Advanced Course*. Springer Verlag, New York, 1995.
- [9] V. S. Borkar. Stochastic approximation with two time scales. *Systems and Control Letters*, 29:291–294, 1997.
- [10] H. F. Chen, T. E. Duncan, and B. P.-Duncan. A Kiefer-Wolfowitz algorithm with randomized differences. *IEEE Trans. Autom. Contr.*, 44(3):442–453, 1999.
- [11] E. K. P. Chong and P. J. Ramadge. Stochastic optimization of regenerative systems using infinitesimal perturbation analysis. *IEEE Trans. on Autom. Contr.*, 39(7):1400–1410, 1994.
- [12] A. I. Elwalid and D. Mitra. Statistical multiplexing with loss priorities in rate based congestion control of high speed networks. *IEEE Trans. on Comm.*, 42(11):2989–3002, 1994.
- [13] M. C. Fu. Convergence of a stochastic approximation algorithm for the $GI/G/1$ queue using infinitesimal perturbation analysis. *J. Optim. Theo. Appl.*, 65:149–160, 1990.
- [14] M. C. Fu. Optimization via simulation: a review. *Annals of Oper. Res.*, 53:199–248, 1994.
- [15] M. C. Fu and S. D. Hill. Optimization of discrete event systems via simultaneous perturbation stochastic approximation. *IIE Trans.*, 29(3):233–243, 1997.
- [16] M. W. Hirsch. Convergent activation dynamics in continuous time networks. *Neural Networks*, 2:331–349, 1987.

- [17] R. Jain, S. Kalyanaraman, R. Viswanathan, and R. Goyal. A sample switch algorithm. *ATM Forum/95-0178*, Feb 1995.
- [18] S. Kalyanaraman. Traffic management for the available bit rate (ABR) service in asynchronous transfer mode (ATM) networks. Ph.D. dissertation, Dept. of Computer and Information Sciences, The Ohio State University, August 1997.
- [19] A. Kolarov and G. Ramamurthy. A control theoretic approach for high speed ATM networks. In *Proceedings of IEEE Infocom97*, pages 293–301, Apr 1997.
- [20] H. J. Kushner and D. S. Clark. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer Verlag, New York, 1978.
- [21] H. J. Kushner and G. G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer Verlag, New York, 1997.
- [22] S. Mascolo, D. Cavendish, and M. Gerla. ATM rate based congestion control using a smith predictor: an EPRCA implementation. In *Proceedings of IEEE Infocom96*, pages 569–576, Mar 1996.
- [23] J. Neveu. *Discrete Parameter Martingales*. North Holland, Amsterdam, 1975.
- [24] R. S. Pazhyannur and R. Agrawal. Feed-back based flow control of B-ISDN/ ATM networks. *IEEE JSAC*, pages 1252–1266, 1995.
- [25] R. S. Pazhyannur and R. Agrawal. Rate based flow control with delayed feedback in integrated services networks. Technical Report ECE-97-4 ECE Dept., Univ. of Wisconsin - Madison, July 1997.
- [26] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York, 1994.
- [27] T. G. Robertazzi. *Computer Networks and Systems: Queueing Theory and Performance Evaluation*. Springer-Verlag, New York, 1990.
- [28] L. Roberts. Enhanced proportional rate control algorithm (PRCA). *ATM Forum/94-0735R1*, Aug 1994.
- [29] P. J. Schweitzer. Perturbation theory and finite Markov chains. *J. Appl. Prob.*, 5:401–413, 1968.
- [30] V. Sharma and J. Kuri. Stability and performance analysis of rate based feedback flow control of ATM networks. *To appear in QUESTA*, 1999.
- [31] V. Sharma and R. R. Mazumdar. Stability and performance of some window and credit based flow control in the presence of background traffic. *Submitted*, 1997.
- [32] K. Y. Siu and H. Y. Tzeng. Intelligent congestion control for ABR service in ATM networks. *ACM SIGCOMM Computer Communication Review*, pages 81–106, 1995.

- [33] J. C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Autom. Cont.*, 37(3):332–341, 1992.
- [34] Y. T. Wang and B. Sengupta. Performance analysis of a feed back congestion control policy. *ACM SIGCOMM Computer Communication*, pages 149–157, 1991.
- [35] H. Zhang, O. W. Yang, and H. Mouftah. Design of robust congestion controllers for ATM networks. In *Proceedings INFOCOM 97*, pages 302–309, April 1997.