SRC TR 85-38r1

An Optimal Adaptive Scheme for Two Competing Queues with Constraints

By

A. Shwarts and A. Makowski

Lecture Notes in Control and Information Sciences

Edited by M. Thoma and A. Wyner

83

Analysis and Optimization of Systems

Proceedings of the Seventh International Conference on Analysis and Optimization of Systems
Antibes, June 25-27, 1986

Edited by
A. Bensoussan and J. L. Lions

Springer-Verlag
Berlin Heidelberg New York Tokyo

AN OPTIMAL ADAPTIVE SCHEME FOR TWO COMPETING QUEUES WITH CONSTRAINTS

by

Adam Shwartz [1] and Armand M. Makowski [2]

ABSTRACT

Two types of traffic, e. g., voice and data, share a single synchronous and noisy communication channel. This situation is modelled as a system of two discrete-time queues with geometric service requirements which compete for the attention of a single server. The problem is cast as one in Markov decision theory with long-run average cost and constraint. An optimal strategy is identified that possesses a simple structure, and its implementation is discussed in terms of an adaptive algorithm of the stochastic-approximations type. The proposed algorithm is extremely simple, recursive and easily implementable, with no a priori knowledge of the actual values of the statistical parameters. The derivation of the results combines martingale arguments, results on Markov chains, O.D.E. characterization of the limit of stochastic approximations and methods from weak convergence. The ideas developed here are of independent interest and should prove useful in studying broad classes of constrained Markov decision problems.

1. INTRODUCTION:

Consider the simple situation where two types of traffic, e. g., voice and data, compete for the use of a single synchronous communication channel. Time is thus slotted with packets formatted so that their transmission time corresponds exactly to a time slot. Each type of traffic is received in an infinite capacity buffer. At the beginning of each time slot, the controller gives priority to one of the queues according to some prespecified dynamic priority assignment:

Channel rights are given to the corresponding traffic, and the selected queue attempts transmission during that slot. The channel is assumed noisy, possibly due to co-user interference, and with a positive probability, the transmission will fail; in that case, retransmission of the failed packet will be rescheduled at a later time in accordance with the channel control policy.

In this application, the quality of service provided by the channel can be measured by *packet delay* and depends heavily on the traffic type. Voice service is expected to deteriorate significantly when voice packets experience large delays, whereas data service is less sensitive to large delay fluctuations in data traffic. Owing to these natural time requirements on the various traffic types, it

[1]Electrical Engineering Department, Technion-Israel Institute of Technology, Haifa 32000, ISRAEL. The work of this author was supported partially through a Grant from AT&T Bell Laboratories, and partially through a grant from the Minta Martin Aeronautical Research Fund, College of Engineering, University of Maryland at College Park.

[2]Electrical Engineering Department and Systems Research Center, University of Maryland, College Park, Maryland 20742, U.S.A. The work of this author was supported partially through NSF Grant ECS-83-51836, partially through a Grant from AT&T Bell Laboratories and partially through ONR Grant N00014-84-K-0614

seems reasonable to seek a dynamic priority-based channel access mechanism which minimizes the average delay of data packets, under the constraint that the average delay of voice packets be bounded from above, say by $V \geq 0$.

In a recent paper [12] , Nain and Ross proposed a simple model to capture this situation, by formulating it as a system of two discrete-time queues with geometric service requirements that compete for the attention of a single server. A simple application of Little's result allows the cost function and the constraint to be easily recast as long-run averages of expressions which are linear in the queue sizes. Extending some of the optimality results from Baras, Dorsey and Makowski [3] , they show that if the bound can be met, then there exists an optimal policy with the following simple structure: If both queues are non-empty, a biased coin is flipped with bias η^*, and channel right is given to voice and data traffic according to the outcome with probability η^* and $1-\eta^*$, respectively. If one of the queues is empty, the other queue is automatically given channel use.

The optimal bias value η^* is determined so that the delay constraint on the voice packets is met. As argued in Section 4, even in the event the values of the various statistical parameters were available to the channel controller, non-trivial off-line computations are required in order to implement this seemingly simple policy [12]. To circumvent this difficulty, an *adaptive* scheme is proposed which uses ideas from stochastic approximations. This paper is devoted to the performance analysis of this adaptive policy which can be easily implemented on-line and which performs no worse than the optimal policy of Nain and Ross, in that it minimizes the average delay for data packets and meets the constraint on voice packets. Because of lack of space, proofs are only sketched, if not omitted all together, and are available in a lengthier version of this paper [17].

The paper is structured as follows: The model and basic assumptions are described in Section 2. The results of Nain and Ross are reviewed in Section 3 and the adaptive scheme is proposed in Section 4, and its main properties discussed. The stability properties of the two competing queues system are discussed in Section 5 for a class of randomized strategies, whereas Section 6 is devoted to the derivation of some useful properties of the invariant measure associated to the queue size process when operating under a simple randomized strategy of the type described above. The convergence of the proposed adaptive scheme is taken on in Section 7. Cost analysis of the proposed channel allocation scheme requires a useful extension to a well-known result of Mandl [11] on the optimality of adaptive policies, which is given in Section 8.

A word on the notation: The set of all non-negative integers is denoted by $\mathbb{N}$ while $\mathbb{R}$ stands for the set of all real numbers. The indicator function of a set A is denoted by $I(A)$ and the Kronecker delta is denoted by $\delta(\bullet,\bullet)$ with $\delta(a,b)=1$ if $a=b$ and $\delta(a,b)=0$ otherwise. An element x in $\mathbb{N}^2$ will sometimes be written as a pair (x^v, x^d), with the convention $0:=(0,0)$, and $|x| := x^v + x^d$. Moreover, for any x in $\mathbb{R}$, pose $\bar{x} := 1-x$, with a similar convention for $\mathbb{R}$-valued mappings. For any mapping $h : \mathbb{N}^2 \to \mathbb{R}$, it is convenient to pose $|h| := \sup_x |h(x)|$.

2. MODEL AND ASSUMPTIONS:

The canonical sample space

In this paper, all probabilistic structures are defined on a single sample space $\Omega := \mathbb{N}^2 \times \left(\{0,1\} \times \mathbb{N}^2 \times \{0,1\}^2 \right)^\infty$ which acts as the *canonical* space for the Markov decision problem under consideration. The *information* spaces $\{H_n\}_1^\infty$ are recursively defined by $H_1 := \mathbb{N}^2$ and

$H_{n+1} := H_n \times \left(\{0,1\} \times \mathbb{N}^2 \times \{0,1\}^2 \right)$ for all $n = 1,2,\ldots$. With a slight abuse of notation, Ω is clearly identified with H_∞.

A generic element ω of the sample space Ω is viewed as a sequence $(x, \omega_1, \omega_2, \cdots)$ with blocks x in $\mathbb{N}^2$ and ω_n in $\{0,1\} \times \mathbb{N}^2 \times \{0,1\}^2$ for all $n = 1,2,\ldots$. Each one of the blocks ω_n will be decomposed into a triplet (u_n, a_n, b_n), where u_n, a_n and b_n are elements of $\{0,1\}$, $\mathbb{N}^2$ and $\{0,1\}^2$, respectively. Finally, for each $n = 1,2,\ldots$, an element h_n in H_n is uniquely associated with the sample ω by $h_n := (x, \omega_1, \omega_2, \cdots, \omega_{n-1})$, with $h_1 := x$.

These quantities can be readily interpreted in the context of the situation described in the introduction: Let the sample $\omega = (x, \omega_1, \omega_2, \cdots)$ be realized. The number of voice and data packets initially in the system is set at x^v and x^d, respectively, with $x = (x^v, x^d)$. For each $n = 1.2,\ldots$, the state of the system is represented by a vector $x_n = (x_n^v, x_n^d)$ of integer components with the interpretation that x_n^v and x_n^d voice and data packets are stored in buffer awaiting transmission, at the beginning of the slot $[n, n+1)$. Thus at that time,

(i): control action u_n is selected with the convention that $u_n = 1$ (resp. $u_n = 0$) if voice (resp. data) is given channel right during that slot;

(ii): new packets arrive into the system according to the vector $a_n = (a_n^v, a_n^d)$, in that a_n^v new voice packets and a_n^d new data packets join their respective queues, and

(iii): completions of transmission are encoded in the binary vector $b_n = (b_n^v, b_n^d)$; here $b_n^v = 1$ (resp. $b_n^v = 0$) signifies successful completion (resp. abortion) of transmission for a voice packet which is given channel rights. A similar interpretation is given to the binary variable b_n^d relative to data packets.

As a result, the successive system states or queue sizes form an $\mathbb{N}^2$-valued sequence $\{x_n\}_1^\infty$ generated recursively by

$$x_{n+1}^v = x_n^v + a_n^v - I[x_n^v \neq 0] u_n b_n^v \qquad n = 1.2,\ldots (2.1a)$$

$$x_{n+1}^d = x_n^d + a_n^d - I[x_n^d \neq 0] \bar{u}_n b_n^d \qquad n = 1.2,\ldots (2.1b)$$

with $x_1 := x$.

At the beginning of each time slot $[n, n+1)$, the channel controller has access to the initial queue sizes x, the past arrival pattern $a_i, 1 \leq i < n$, the past decisions $u_i, 1 \leq i < n$ and the channel history $b_i, 1 \leq i < n$. Thus, the channel controller has knowledge of the information vector h_n, which is to be used for generating the control value u_n implemented in the slot $[n, n+1)$. The selection of this control value is done according to a prespecified mechanism, which may be either deterministic or random.

The basic random variables

The *coordinate* mappings $\mathbb{E}$, $\{U(n)\}_1^\infty$, $\{A(n)\}_1^\infty$ and $\{B(n)\}_1^\infty$ are defined on the sample space Ω by setting

$$\mathbb{E}(\omega) := x, \quad U(n, \omega) := u_n, \quad A(n, \omega) := a_n, \quad B(n, \omega) := b_n \qquad n = 1.2,\ldots (2.2)$$

for all ω in Ω. Moreover, the *information* mappings $\{H(n)\}_1^\infty$ are defined by

$$H(n,\omega):=(x,\omega_1,\omega_2,\cdots,\omega_{n-1})=:h_n \qquad n=1,2,\ldots (2.3)$$

for all ω in Ω.

For each $n=1,2,\ldots$, the mapping $H(n)$ generates a σ-field $\mathcal{F}_n$ on the sample space Ω, with $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. With standard notation, $\mathcal{F}:=\bigvee_{n=1}^{\infty}\mathcal{F}_n$ is simply the natural σ-field on the infinite cartesian product $\mathcal{H}_\infty$ generated by the mappings Ξ and $\{U(n),A(n),B(n)\}_1^\infty$. Throughout, the sample space Ω is always equipped with this σ-field $\mathcal{F}$ and in that event, the mappings Ξ, $\{U(n)\}_1^\infty$, $\{A(n)\}_1^\infty$, $\{B(n)\}_1^\infty$ and $\{H(n)\}_1^\infty$ are all *random variables* (RV) taking values in $\mathbb{N}^2$, $\{0,1\}$, $\mathbb{N}^2$, $\{0,1\}^2$ and $\mathcal{H}_n$ respectively. The sequence $\{X(n)\}_1^\infty$ of $\mathbb{N}^2$-valued RV's is now defined recursively by

$$X^v(n+1)=X^v(n)+A^v(n)-I[X^v(n)\neq 0]U(n)B^v(n) \qquad n=1,2,\ldots (2.4a)$$

$$X^d(n+1)=X^d(n)+A^d(n)-I[X^d(n)\neq 0]\overline{U}(n)B^d(n) \qquad n=1,2,\ldots (2.4b)$$

with $X(1):=\Xi$. The RV's $X(n)$ are clearly $\mathcal{F}_n$-measurable and can thus be recovered from knowledge of the RV's $H(n)$.

The probabilistic structure

Since randomized strategies are allowed, an admissible control policy π is defined as any collection $\{\pi_n\}_1^\infty$ of mappings $\pi_n:\mathcal{H}_n\to[0,1]$, with the interpretation that at times $n=1,2,\ldots$, voice (resp. data) packets are given channel rights with probability $\pi_n(h_n)$ (resp. $\overline{\pi}_n(h_n)$) whenever the information vector h_n is available to the channel controller. Denote the collection of all such admissible policies by Π.

Let $q(\cdot)$ and $q_A(\cdot)$ be two probability distributions on $\mathbb{N}^2$, and fix μ^v and μ^d in the interval $[0,1]$. Given an arbitrary policy π in Π, it is a simple exercise to show constructively, via the Daniell-Kolmogorov Consistency Theorem, that there exists a unique probability measure P^π on $\mathcal{F}$ which satisfies the requirements (R1)-(R3) below, i. e.,

(R1): For all x in $\mathbb{N}^2$;

$$P^\pi[\Xi=x]:=q(x),$$

(R2): For all a_n in $\mathbb{N}^2$ and b_n in $\{0,1\}^2$,

$$P^\pi[A(n)=a_n,B(n)=b_n \mid \mathcal{F}_n\vee\sigma\{U(n)\}]:=P^\pi[A(n)=a_n]P^\pi[B(n)=b_n]$$

$$:=q_A(a_n)\left(b_n^v\mu^v+\overline{b}_n^v\overline{\mu}^v\right)\left(b_n^d\mu^d+\overline{b}_n^d\overline{\mu}^d\right) \qquad n=1,2,\ldots$$

and

(R3):

$$P^\pi[U(n)=1\mid\mathcal{F}_n]:=\pi_n(H_n). \qquad n=1,2,\ldots$$

The reader will readily check that under each probability measure P^π, the following properties hold true.

(P1): The $\mathbb{N}^2$-valued RV Ξ and the sequences of RV's $\{A(n)\}_1^\infty$ and $\{B(n)\}_1^\infty$ are *mutually independent*;

(P2): The sequences $\{B^v(n)\}_1^\infty$ and $\{B^d(n)\}_1^\infty$ of $\{0,1\}$-valued RV's are *mutually independent Bernoulli* sequences with parameters μ^v and μ^d, respectively;

(P3): The $\mathbb{N}^2$-valued RV's $\{A(n)\}_1^\infty$ form a sequence of *i.i.d* RV's, with a common distribution $q_A(\bullet)$.

(P4): The probability transitions have the form

$$P^\pi[X(n+1)=y \mid \mathcal{F}_n]=p(X(n),y;\pi_n(H(n))) \tag{2.5}$$

where

$$p(x,y;\eta) := \eta Q^v(x,y)+(1-\eta)Q^d(x,y), \tag{2.6}$$

$$Q^v(x,y) := P^\pi[x^v+A^v(n)-I[x^v\neq 0]B^v(n)=y^v,\ x^d+A^d(n)=y^d] \tag{2.7a}$$

and

$$Q^d(x,y) := P^\pi[x^v+A^v(n)=y^v,\ x^d+A^d(n)-I[x^d\neq 0]B^d(n)=y^d] \tag{2.7b}$$

for all x and y in $\mathbb{N}^2$, and all η in $[0,1]$. Note that the right handsides of (2.7) are independent of n and of the policy π owing to the assumptions made earlier.

Several families of policies

Several subclasses of policies within Π will be of interest in the sequel.

A policy π in Π is said to be a *Markov* or *memoryless* policy if there exists a family of mappings $\{g_n\}_1^\infty$ where $g_n : \mathbb{N}^2 \to [0,1]$ such that

$$\pi_n(H(n))=g_n(X(n)) \quad P^\pi\text{-}a.s. \qquad n=1,2,\ldots \tag{2.8}$$

with $\{X(n)\}_1^\infty$ generated through the recursion (2.4). In the event all the mappings $\{g_n\}_1^\infty$ are identical to a given mapping g, the Markov policy π is termed *stationary* and can be identified with the mapping g itself, as will be done repeatedly in the sequel.

A policy π in Π will be said to be a *pure* strategy if there exists a family $\{f_n\}_1^\infty$ of mappings $f_n : \mathcal{H}_n \to \{0,1\}$ such that

$$\pi_n(H(n))=\delta(1,f_n(H(n))) \quad P^\pi\text{-}a.s. \qquad n=1,2,\ldots \tag{2.9}$$

A pure policy π can thus be identified with the sequence of deterministic mappings $\{f_n\}_1^\infty$. A *pure Markov stationary* policy π in Π is thus fully characterized by a single mapping $f : \mathbb{N}^2 \to \{0,1\}$ to which it is substituted in the notation.

A policy π in Π is said to be *non-idling* or *work-conserving* whenever the conditions

$$\pi_n(H(n)) = 1 \text{ on the event } [X^v(n)\neq 0,\ X^d(n)=0] \qquad n=1,2,\ldots \tag{2.10a}$$

$$\pi_n(H(n)) = 0 \text{ on the event } [X^v(n)=0,\ X^d(n)\neq 0] \qquad n=1,2,\ldots \tag{2.10b}$$

hold true P^π-*a.s.*, in which case, the P^π-*a.s.* equality

$$U(n)I[X^v(n)\neq 0]+\bar{U}(n)I[X^d(n)\neq 0] = 1-I[X(n)=0] \qquad n=1,2,\ldots \tag{2.11}$$

necessarily follows.

3. OPTIMALITY RESULTS IN THE NON-ADAPTIVE CASE:

For any admissible policy π in Π, pose

$$J^v(\pi) := \overline{\lim_{n\uparrow\infty}} \frac{1}{n} E^\pi \sum_{i=1}^{n} X^v(i) \tag{3.1}$$

and

$$J^d(\pi) := \overline{\lim_{n\uparrow\infty}} \frac{1}{n} E^\pi \sum_{i=1}^{n} X^d(i). \tag{3.2}$$

Moreover, given $V>0$, denote by Π_V the set of all admissible control policies π in Π which satisfy the constraint

$$J^v(\pi) \leq V. \tag{3.3}$$

In this paper, the discussion will revolve around the following family of *constrained* optimization problems $\{(P_V), V>0\}$, where

$$(P_V): \quad \textit{Minimize } J^d(\pi) \textit{ on } \Pi_V.$$

This problem was originally formulated and solved by Nain and Ross [12] through a Lagrangian argument. Here, the Lagrangian functional for problem (P_V) is naturally defined for any admissible policy π in Π to be

$$J^\gamma(\pi) := \overline{\lim_{n\uparrow\infty}} \frac{1}{n} E^\pi \sum_{i=1}^{n} X^d(i) + \gamma X^v(i) \tag{3.4}$$

with $\gamma \geq 0$, and attention is given to the *unconstrained* problems $\{(PL_\gamma), \gamma \geq 0\}$, where

$$(PL_\gamma): \quad \textit{Minimize } J^\gamma(\pi) \textit{ on } \Pi.$$

Each Lagrangian problem (PL_γ) is a two competing queues problem with a long-run average cost linear in the queue sizes, as studied by Baras, Dorsey and Makowski [3]. These authors considered only *non*-randomized policies and showed that when $\mu^d \neq \gamma\mu^v$ a fixed static priority assignment is optimal, whereas if $\mu^d = \gamma\mu^v$, *any non*-randomized policy which is *non*-idling is optimal. Nain and Ross strengthened these results by showing that if $\mu^d = \gamma\mu^v$, *any non*-idling policy in Π is optimal, be it randomized or not.

The mappings f^v and f^d are defined on $\mathbb{N}^2$ and take values in $\{0,1\}$ according to

$$f^v(x) := I[x^v \neq 0] \text{ and } f^d(x) := I[x^d = 0] \tag{3.5}$$

for all $x = (x^v, x^d)$ in $\mathbb{N}^2$. These mappings naturally induce non-idling pure Markov stationary policies, denoted again by f^v and f^d. According to f^v (resp. f^d), priority is always given to the voice (resp. data) packets (when available). These are fixed static priority assignments of which the well-known μc-rules constitute only one description.

Following Nain and Ross [12] , the class of *simply randomized* (SR) policies is defined as a one-parameter family of non-idling Markov stationary policies obtained by *randomizing* the static priority assignments f^v and f^d with a fixed bias. Formally, for any η with $0 \leq \eta \leq 1$, the SR policy with bias η is the non-idling Markov stationary policy denoted by f^η which is generated through the mapping $f^\eta: \mathbb{N}^2 \to [0,1]$ where

$$f^{\eta}(x) := \eta \text{ if } x^v \neq 0 \text{ and } x^d \neq 0 \tag{3.6a}$$

$$:= 1 \text{ if } x^d = 0 \text{ and } x^v \neq 0 \tag{3.6b}$$

$$:= 0 \text{ if } x^v = 0 \text{ and } x^d \neq 0 \tag{3.6c}$$

for all $x=(x^v, x^d)$ in $\mathbb{N}^2$, the convention being that $f^{\eta}(0)=0$. Note that $f^1 \equiv f^v$ and $f^0 \equiv f^d$ on $\mathbb{N}^2 - \{0\}$.

The basic result of Nain and Ross can then be summarized as follows. Let λ^v and λ^d be the common (possibly infinite) first moments of the sequences of $\mathbb{N}$-valued RV's $\{A^v(n)\}_1^{\infty}$ and $\{A^d(n)\}_1^{\infty}$, respectively. For future use, let ρ be the expression defined by

$$\rho := \frac{\lambda^v}{\mu^v} + \frac{\lambda^d}{\mu^d} \tag{3.7}$$

with the understanding that $\rho<1$ throughout this paper.

Theorem 3.1. *Assume that $J^v(f^v) \leq V \leq J^v(f^d)$, i. e., the class of constrained policies Π_V is non-empty. Under the foregoing assumptions (R1)-(R3), there exists a SR policy f^{η^*} that solves problem (P_V) and its bias value η^* is determined by the equation*

$$J^v(f^{\eta}) = V, \quad \eta \text{ in } [0,1]. \tag{3.8}$$

4. A STOCHASTIC APPROXIMATIONS IMPLEMENTATION:

The determination of the optimal bias value η^* requires the evaluation of the expression $J^v(f^{\eta})$ for all values of η in the unit interval $[0,1]$. This a non-trivial task for $0<\eta<1$ since in that case the two competing queues can be interpreted (under P^{η}) as a two processor-sharing system of the type studied by Fayolle and Iasnogorodski [1]. As shown there, the computation reduces to the solution of a Riemann-Hilbert problem, from which the optimal bias η^* could in principle be determined as a function of the arrival statistics $q_A(\bullet)$ and of the rates μ^v and μ^d. Even if these statistical parameters were available to the channel controller, the evaluation of η^* seems to constitute a non-trivial computational undertaking [12].

To avoid these difficulties, an alternate approach is now proposed for *adaptively* generating a control policy that solves the constrained problems $\{(P_V), V \geq 0\}$. The proposed scheme is driven by the very special form of the optimal policy given in Theorem 3.1 (which saturates the constraint) and by well-known ideas from the theory of Stochastic Approximations, of which the Robbins-Monro scheme is the archetypical example [13].

This adaptive scheme generates a sequence of bias values $\{\eta(n)\}_1^{\infty}$ through the recursion

$$\eta(n+1) = \left[\eta(n) - a_n \left(V - X^v(n+1) \right) \right]_0^1 \qquad n=1,2,\ldots \tag{4.1}$$

with $\eta(1)$ given in $[0,1]$, the convention being that $[x]_0^1 := 0 \vee (x \wedge 1)$ for all x in $\mathbb{R}$. As with most stochastic approximation algorithms, the step sizes $\{a_n\}_1^{\infty}$ form an $\mathbb{R}^+$-valued sequence which is assumed to satisfy the conditions

$$0<a_n\downarrow 0\,,\quad \sum_{n=1}^{\infty} a_n=\infty\,,\quad \sum_{n=1}^{\infty}\mid a_{n+1}-a_n\mid<\infty. \tag{4.2}$$

The RV $\eta(n)$ constitutes an *estimate* of the bias value η^* and is thus interpreted as the (conditional) probability of giving priority to voice packets in the slot $[n,n+1)$, given that the information $H(n)$ is available at time n to the channel controller. The adaptive non-idling policy so generated by the sequence $\{\eta(n)\}_1^{\infty}$ is denoted by α and can be formally defined through the mappings $\{\alpha_n\}_1^{\infty}$ with

$$\alpha_n(H(n)):=f^{\eta(n)}(X(n)). \qquad n=1,2,\ldots. \tag{4.3}$$

The policy α is structurally simple and easy to implement on-line; this simplicity of implementation is derived from the fact that the difficult step of *directly* solving (3.8) is completely bypassed. Moreover, since *no a priori* knowledge of the various statistics is required in order to implement it, this policy α is also of interest in the case where no such knowledge is available. As such, the proposed policy α constitutes an adaptive policy in the restricted technical sense understood in the literature on the non-Bayesian adaptive control problem for Markov chains [7].

This paper is devoted to the study of the estimation/control properties of the policy α. The basic results, which are summarized in the two propositions given below, will be derived under the following additional moment assumptions on the data of the problem: For every policy π in Π,

(R4a):

$$E^{\pi}\left[\mid \Xi^{v}\mid^{2}+\mid \Xi^{d}\mid^{2}\right]=\sum_{x\in\mathbb{N}^2}\left[\mid x^{v}\mid^{2}+\mid x^{d}\mid^{2}\right]q(x)<\infty$$

(R4b): For all $n=1,2,\ldots$,

$$E^{\pi}\left[\mid A^{v}(n)\mid^{2}+\mid A^{d}(n)\mid^{2}\right]=\sum_{a\in\mathbb{N}^2}\left[\mid a^{v}\mid^{2}+\mid a^{d}\mid^{2}\right]q_A(a)<\infty$$

Theorem 4.1. *Under the foregoing assumptions (R1)-(R4), the sequence of biases* $\{\eta(n)\}_1^{\infty}$ *generated by the recursion (4.1) converges in probability (under* P^{α}*) to the optimal bias* η^*.

Proof: Details are given in Section 7.

Theorem 4.2. *Under the foregoing assumptions (R1)-(R4) and the assumption that* $J^{v}(f^{v})\le V\le J^{v}(f^{d})$, *the policy* α *solves problem* (P_V) *with* $J^{v}(\alpha)=V$.

Proof: The result follows readily by combining Theorems 3.1, 4.1 and 8.1. Δ.

A closely related and somewhat more intuitive stochastic approximation scheme could have been considered; it is the one that generates the sequence of bias values $\{\eta(n)\}_1^{\infty}$ through the recursion

$$\eta(n+1)=\left[\eta(n)-a_n\left(V-\frac{1}{n+1}\sum_{i=1}^{n+1}X^{v}(i)\right)\right]_0^1 \qquad n=1,2,\ldots. \tag{4.4}$$

with $\eta(1)$ given in $[0,1]$. Here, in a closer analogy with the Robbins-Monro scheme, the terms

$$\frac{1}{n+1}\sum_{i=1}^{n+1} X^{v}(i) \qquad n=1,2,...(4.5)$$

can be viewed as *noisy observations* of the steady-state average delay $J^{v}(f^{v^*})$.

Although the adaptive control policy associated with this second recursion is expected to exhibit essentially the same properties as the policy α, the corresponding analysis in that case is much more involved as existing techniques for handling stochastic approximation algorithms are either not applicable nor can they be readily extended without considerable additional work.

The salient feature of the problem treated here is that although the optimal control policy is known to exist and its structure has been completely identified, it is not readily implemented even if the model parameters were known. Many control problems for Markov chains exhibit this feature and thus warrant similar studies that combine ideas from the theory of stochastic approximations to specific optimality results. Indeed, as Ross showed in his doctoral dissertation [14] , Markov decision processes with constraints naturally lead to simple randomized strategies obtained by mixing at most two pure Markov stationary strategies through a simple flip of a biased coin; the optimal bias is determined by solving a (typically complicated) equation that expresses the fact that at optimality the constraint is saturated. This result alone indicates that constrained Markov decision problems will constitute a rich source of concrete problems for which the approach proposed here could provide a viable alternative for generating adaptive controls.

5. STABILITY UNDER NON-IDLING RANDOMIZED POLICIES:

In order to study the stability properties of the process $\{X(n)\}_1^{\infty}$ under the probability measure associated with an arbitrary admissible (not necessarily Markov stationary) policy, it is convenient to introduce the mapping $Z: \mathbb{N}^2 \to \mathbb{R}$ defined by

$$Z(x) := \frac{x^{v}}{\mu^{v}} + \frac{x^{d}}{\mu^{d}} \qquad (5.1)$$

for all $x=(x^{v},x^{d})$ in $\mathbb{N}^2$.

Throughout this section, let π be a *non*-idling policy in Π. The results that follow extend to randomized policies some of the results obtained by Baras, Dorsey and Makowski ([2] , Sections 4) for non-randomized policies only; details are available in [17] :

The properties (P1)-(P4) and remark (2.11) readily imply the P^{π}-*a.s.* relations

$$E^{\pi}[Z(X(n+1)) \mid \mathcal{F}_n] = Z(X(n))-(1-\rho)+I[X(n)=0]. \qquad n=1.2,...(5.2)$$

As in [2] , the sequence of $\mathbb{R}$-valued RV's $\{M(n)\}_1^{\infty}$, defined by

$$M(n+1) := Z(X(n+1))+(1-\rho)n - \sum_{i=1}^{n} I[X(i)=0] \qquad n=1,2,...(5.3)$$

with $M(1) := Z(X(1))$, turns out to be an $(P^{\pi},\mathcal{F}_n)$-martingale by construction.

Let σ be an arbitrary $\mathcal{F}_n$-stopping time. The $\mathbb{N}$-valued RV $\nu(\sigma)$ is defined by

$$\nu(\sigma) := \inf\{n \geq 1: X(\sigma+n)=0\} \quad \text{if } \sigma<\infty. \qquad (5.4)$$

with the convention that $\nu(\sigma)=\infty$ whenever this set is empty or when $\sigma=\infty$; the RV $\tau(\sigma) := \sigma + \nu(\sigma)$ is clearly an $\mathcal{F}_n$-stopping time.

The sequence of $\mathbb{R}$-valued RV's $\{z^{M(n)}\}_1^\infty$ is a non-negative $(P^\pi, \mathcal{F}_n)$-submartingale for $0<z<1$, owing to Jensen's inequality. From Doob's optional sampling theorem applied to this submartingale with the stopping times σ and $\tau(\sigma)$, via a standard localization argument, it follows that $\sigma<\infty$, P^π-*a.s.* implies $\nu(\sigma)<\infty$ and thus $\tau(\sigma)<\infty$, both P^π-*a.s.*

Similar arguments, this time with the $(P^\pi, \mathcal{F}_n)$-martingale $\{M(n)\}_1^\infty$, readily yield the moment formula

$$E^\pi[\nu(\sigma) \mid \mathcal{F}_\sigma] = \frac{1}{1-\rho}\Big(Z(X(\sigma)) + I[X(\sigma)=0]\Big) \quad P^\pi\text{-}a.s. \tag{5.5}$$

As in Section 4 of [2], define the clearing times $\{\tau_n\}_0^\infty$ for the queue size process $\{X(n)\}_1^\infty$ as the successive epochs at which the process visits the empty state 0, with $\tau_0=0$ and $\tau_1=\nu(1)$. Here too, it can be argued ([2], Thm. 4.5) that the sequence $\{\tau_{n+1}-\tau_n\}_0^\infty$ is a (possibly delayed) *renewal* sequence with *finite means* given by

$$E^\pi[\tau_{n+1}-\tau_n] = \frac{1}{(1-q_A(0))(1-\rho)}, \qquad n=1,2,\ldots \tag{5.6}$$

the latter taking place *independently* of the initial state distribution.

The crucial technical results for the development of the material discussed in this paper are contained in the following propositions whose proofs are available in [17].

Theorem 5.1. *Under the foregoing assumptions (R1)-(R4),*

$$\sup_n E^\pi\Big[\,|\tau_{n+1}-\tau_n|^2\,\Big] < \infty \tag{5.7}$$

Theorem 5.2. *Under the foregoing assumptions (R1)-(R4), there exists a constant K such that for every non-idling policy π in Π,*

$$\sup_n E^\pi\Big[\,|X^v(n)|^2 + |X^d(n)|^2\,\Big] \le K < \infty \tag{5.8}$$

and the RV's $\{X(n)\}_1^\infty$ thus form a uniformly integrable sequence under P^π.

6. INVARIANT MEASURE-EXISTENCE AND CONTINUITY:

For each $0\le\eta\le1$, the sequence $\{X(n)\}_1^\infty$ is a *homogeneous* Markov chain over the state-space $\mathbb{N}^2$ under the probability measure P^η induced by the SR policy f^η. This chain is clearly *aperiodic* and all states communicate, whence the chain is *irreducible*. Moreover, the finite mean property (5.6) readily implies that the empty state is *positive recurrent* and so are all the states by virtue of the irreducibility of the chain.

It now follows from standard results on Markov chains [6] that the Markov chain $\{X(n)\}_1^\infty$ admits under P^η a *unique invariant* measure, which is denoted throughout by $\mathbb{P}^\eta$ with corresponding expectation operator $\mathbb{E}^\eta$. Moreover, for any *bounded* mapping $d: \mathbb{N}^2\to\mathbb{R}$,

$$\lim_{n\to\infty}\frac{1}{n}E^\eta\sum_{i=1}^n d(X(i)) = \lim_{n\to\infty}E^\eta d(X(n)) = \mathbb{E}^\eta d(X), \tag{6.1}$$

and this *independently* of the intial state distribution, where $X=(X^v, X^d)$ denotes a generic $\mathbb{N}^2$-

valued RV. This last fact is useful for evaluating the long-run average cost (3.2) and the constraint (3.1) under any SR policy.

Lemma 6.1. *Under the SR policy f^η, the convergences*

$$\lim_{n\to\infty}\frac{1}{n}E^\eta\sum_{i=1}^{n}X^v(i)=\mathbb{E}^\eta X^v \tag{6.2a}$$

and

$$\lim_{n\to\infty}\frac{1}{n}E^\eta\sum_{i=1}^{n}X^d(i)=\mathbb{E}^\eta X^d \tag{6.2b}$$

take place.

Proof: The result follows readily by standard arguments from (6.1) and from the uniform integrability of the sequence $\{X(n)\}_1^\infty$. Δ

The following properties will prove useful in the sequel.

Lemma 6.2. *The mapping $\eta\to\mathbb{E}^\eta X^v$ is continuous and strictly monotone decreasing on the interval* [0,1].

Proof: The continuity has already been established by Nain and Ross ([12] , Thm. 5.3, p. 23). The strict monotonicity property follows by an easy stochastic coupling argument that uses the form (2.5)-(2.7) of the transition probabilities. Full details are available in [17]. Δ

7. CONVERGENCE RESULTS FOR THE ESTIMATES $\{\eta(n)\}_1^\infty$:

In this section, arguments are outlined for showing that the sequence of estimates $\{\eta(n)\}_1^\infty$ generated by the recursive scheme (4.3) does indeed converge in probability to the optimal bias value η^* under P^a. In the discussion of Theorem 4.1 that follows, the assumptions (R1)-(R4) are enforced throughout. Results from the work of Kushner and Shwartz [8] will be heavily used, and the reader unfamiliar with the terminology and basic results of the theory of weak convergence is invited to consult the monograph by Billingsley [4]. The proof of the convergence is articulated in three basic steps:

Step 1: The ODE

$$\dot{\eta}(t)=\mathbb{E}^{\eta(t)}X^v - V,\ \eta(0)\ in\ [0,1] \tag{7.1}$$

is *asymptotically stable* and *any* one of its solutions $\eta(\cdot)$ converges *monotonically* to η^*, where η^* is the *unique* solution of

$$\mathbb{E}^\eta X^v=\lim_{n\to\infty}\frac{1}{n}E^\eta\sum_{k=1}^{n}X^v(k)=V. \tag{7.2}$$

The statement on the stability of (7.1) is readily obtained from the strict monotonicity of the mapping $\eta\to\mathbb{E}^\eta X^v$ (Lemma 6.2.), whereas the second statement follows from Lemma 6.1 and Theorem 3.1.

Step 2: The monotone sequence $\{t_n\}_1^\infty$ is given by $t_n:=\sum_{i=1}^{n}a_i$ for all $n=1,2,\ldots$ Define the

[0,1]-valued continuous-time process $\{\eta^0(t),\ t \geq 0\}$ through piecewise constant interpolation, i. e., $\eta^0(t) := \eta(n)$ for all t in the interval $[t_n, t_{n+1})$, and pose $\eta^n(t) = \eta^0(t+t_n)$ for all $t \geq 0$. The stochastic processes $\{\eta^n(\bullet)\}_1^\infty$ form a *tight* sequence in the space $D_{\mathbb{R}}[0,\infty)$, and *any* limit $\eta(\bullet)$ of a convergent subsequence $\{\eta^m(\bullet)\}_1^\infty$ (in the sense of weak convergence) *satisfies* the ODE (7.1).

The validity of these statements will be established by a direct application of the results of Kushner and Shwartz ([8] , Thms. 1,3,4 and the discussion following Thm. 4), with the relevant hypotheses from this reference being listed in their appropriate form and their validity shown to hold here. To summarize, under the conditions A1-A5 and A7-A9 listed below, the sequence of continuous-time processes $\{\eta^n(\bullet)\}_1^\infty$ is tight in $D_{\mathbb{R}}[0,\infty)$ and any weak limit $\eta(\bullet)$ satisfies

$$\dot{\eta}(t) = proj\,[\,\eta(t),\ E^{\eta(t)}X^v - V\,],\ \eta(0)\ in\ [0,1] \tag{7.3}$$

where $proj\,(\bullet,\bullet)$ is the *vector field projection* on the interval [0,1] ([8] , (2.1), p. 20). This projection is defined by

$$proj\,(\eta, f) = 0 \quad \text{if } \eta=0 \text{ and } f<0 \tag{7.4a}$$

$$= 0 \quad \text{if } \eta=1 \text{ and } f>0 \tag{7.4b}$$

$$= f \quad \text{otherwise}, \tag{7.4c}$$

and has the role of keeping the solution $\eta(\bullet)$ in [0,1] at all times. Here, however, the mapping $\eta \to \mathbb{E}^\eta[X^v] - V$ is strictly monotone decreasing with a single zero in the interval [0,1] and the initial conditions are in [0,1], whence this projection has no effect on the solution and can thus be discarded.

A1. The sequence $\{(\eta(n), X(n))\}_1^\infty$ is a *Markov process* under P^α: This was pointed out to be the case here in Section 5.

A2. The sequence $\{X(n)\}_1^\infty$ is *tight* under P^α: This follows from the uniform integrability of $\{X(n)\}_1^\infty$ established in Theorem 5.2.

A3. The one-step transition probabilities functions

$$\mathbb{N}^2 \times [0,1] \to [0,1]\colon (x,\eta) \to P^\alpha[\,X(n+1) \in A \mid X(n)=x,\ \eta(n)=\eta\,],\ A \subseteq \mathbb{N}^2 \tag{7.5}$$

do *not* depend on n, and are *weakly continuous* in η for each x in $\mathbb{N}^2$: Indeed, for all $n=1,2,\ldots,$

$$P^\alpha[\,X(n+1)=y \mid X(n)=x,\ \eta(n)=\eta\,] = p\,(x,y,f^\eta(x)), \tag{7.6}$$

as shown in (2.5). Now, for any bounded $g : \mathbb{N}^2 \to \mathbb{R}$, the relation

$$\sum_y g\,(y) P^\alpha[\,X(n+1)=y \mid X(n)=x,\ \eta(n)=\eta\,]$$

$$= f^\eta(x) \sum_y g\,(y) Q^v(x,y) + (1-f^\eta(x)) \sum_y g\,(y) Q^d(x,y) \tag{7.7}$$

holds for every x in $\mathbb{N}^2$ and η in [0,1], and the weak continuity readily follows owing to (3.6).

A4. Under each SR policy f^η, $0 \leq \eta \leq 1$, the Markov chain $\{X(n)\}_1^\infty$ has a *unique invariant* measure $\mathbb{P}^\eta$, and the family of probability measures $\{\mathbb{P}^\eta,\ 0 \leq \eta \leq 1\}$ is *tight*: The first statement follows from the discussion preceding (6.1), while the second follows from the uniform bound established in Theorem 5.1 (coupled to Tchebychev's inequality).

A5. The conditions

$$0<a_n \downarrow 0\ ,\ \sum_{n=1}^{\infty} a_n = \infty\ ,\ \sum_{n=1}^{\infty} |\, a_{n+1} - a_n \,| < \infty \tag{7.8}$$

hold on the step-size sequence $\{a_n\}_1^\infty$ by the very assumptions (4.2).

Define the mapping $F: \mathbb{N}^2 \to \mathbb{R}$ by $F(x) := x^v - V$ for all x in $\mathbb{N}^2$.

A7. The mapping $\eta \to \sum_y p(x,y;f^\eta(x))F(y)$ is continuous: This readily follows from the relation (7.7) (with F instead of g) and from the assumption (R4).

For all $T>0$ and all n in $\mathbb{N}$, define

$$m(n,T) = max\,\{j : \sum_{i=n}^{j} a_i \le T\}. \tag{7.9}$$

A8. For all j in $\mathbb{N}$, there exists an integer $N(j)$ such that for each $T>0$,

$$P^a[\,|X(n+m)|>K \;\;|\;\; X(n)=x\,,\,\eta(n)=\eta\,]I(|x|<j) \le \epsilon_{K,T}\ , \tag{7.10}$$

for all $m(n,T)-n \ge m \ge N(j)$, where $\epsilon_{K,T} \downarrow 0$ as $K \uparrow \infty$: This easily follows from the stability property of this two competing queues system, which is *uniform* in the non-idling control policies.

A9. There exists a constant $\delta \ge 0$ such that $|F(x)| \le \delta + |x|$ for all x in $\mathbb{N}^2$ and $\sup E(|X(n)|^{1+a}) < \infty$: This clearly holds by the definition of F and from Theorem 5.2 with $\alpha=1$.

Thus all hypotheses are satisfied and the second step of the proof is completed.

Step 3: The sequence $\{\eta(n)\}_1^\infty$ converge in probability to η^* under P^a.

This convergence follows from a shifting argument as in [9, 10]. For every $T>0$, define the process $\eta_T^n(\cdot)$ by $\eta_T^n(t) := \eta^n(t-T)$ for all $t \ge 0$ and assume that for some subsequence

$$\{\eta^m(\cdot)\,,\,\eta_T^m(\cdot)\} \Rightarrow \{\eta(\cdot).\,,\,\eta_T(\cdot)\}, \tag{7.11}$$

where $\Rightarrow$ denotes weak convergence [4]. By a standard argument based on Skorohod representation, the convergence may be assumed to hold for every ω not in A, where A is a null set. Fix ω not in A. The limit $\eta_T(\cdot)$ satisfies an asymptotically stable ODE, with stable point η^*, and the initial condition $\eta_T(0)$ lies in the interval [0,1], therefore it is possible to choose T large enough so that $\eta_T(T)=\eta(0)$ is arbitrarily close to η^*. Note that the appropriate T can be chosen *independently* of the subsequence and of the sample ω not in A. Since the solutions converge to η^* *monotonically*, $\eta(t)$ will remain close to η^* for all $t>0$ for all samples ω not in A. Therefore, $\eta(t) \equiv \eta^*$ and along this subsequence, $\eta(n) \Rightarrow \eta^*$. The subsequence being arbitrary and the limit being non-random and independent of the subsequence, it follows that $\eta(n) \to \eta^*$ in probability, and the third step of the proof is thus completed. Δ

8. THEOREM 4.2 AND MANDL'S RESULTS EXTENDED

The set-up adopted in this section is somewhat more general than is needed for studying the particular adaptive scheme introduced in Section 4, and the results obtained here apply to many other adaptive schemes as well. Throughout this section, η will denote a fixed bias value in the interval [0,1], which enters the definition of a given SR policy. Let $\{\eta(n)\}_1^\infty$ be any sequence of [0,1]-valued $\mathcal{F}_n$-adapted RV's which serve as estimates for the value η, and let α denote again, with a

slight abuse of notation, the admissible policy in Π which is generated through (4.3).

Theorem 8.1. *Under the foregoing assumptions (R1)-(R4), the convergence in probability under P^α of the estimates $\{\eta(n)\}_1^\infty$ to some non-random value η in [0,1] implies the convergences*

$$J^v(\alpha) = \lim_{n\to\infty} \frac{1}{n} E^\alpha \sum_{i=1}^{n} X^v(i) = \lim_{n\to\infty} \frac{1}{n} E^\eta \sum_{i=1}^{n} X^v(i) = J^v(f^\eta) \tag{8.1a}$$

$$J^d(\alpha) = \lim_{n\to\infty} \frac{1}{n} E^\alpha \sum_{i=1}^{n} X^d(i) = \lim_{n\to\infty} \frac{1}{n} E^\eta \sum_{i=1}^{n} X^d(i) = J^d(f^\eta) \tag{8.1b}$$

In order to prove this result, it is necessary to extend an argument due to Mandl [11] to the case of *unbounded* costs over *countable* state-spaces and *randomized* strategies. The discussion that follows leads via two lemmas of independent interest to a proof of Theorem 8.1 which is delayed till the end of this section. Throughout this section, d denotes a mapping $\mathbb{N}^2 \to \mathbb{R}^+$ which is assumed to be *monotone increasing* in each component of its argument. The constant J^η is defined to be the long-run average cost associated with the SR policy f^η for the one-step cost d, i. e.,

$$J^\eta = \lim_{n\to\infty} \frac{1}{n} E^\eta \sum_{i=1}^{n} d(X(i)) \tag{8.2}$$

That this limit exists, even when d is unbounded, will become apparent from the discussion given below.

Lemma 8.2. *If the mapping $h : \mathbb{N}^2 \to \mathbb{R}$ and the constant J solve the vector equation*

$$h(x) + J = \sum_y p(x, y; f^\eta(x)) h(y) + d(x) \tag{8.3a}$$

with

$$\lim_{k\to\infty} \frac{1}{k} E^\eta [h(X(k))] = 0 \tag{8.3b}$$

then $J = J^\eta$.

Proof: The formula (2.5) for the transition probabilities allows a rewriting of (8.3a) in the form

$$h((X(k)) + J = E^\eta [h(X(k+1)) \mid \mathcal{F}_k] + d(X(k)), \qquad k=1,2,\ldots. \tag{8.4}$$

and direct iteration then gives

$$E^\eta [h(\Xi)] + kJ = E^\eta [h(X(k+1))] + E^\eta [\sum_{i=1}^{k} d(X(i))]. \qquad k=1,2,\ldots. \tag{8.5}$$

Now dividing by k and letting $k \to \infty$ gives the result owing to the condition (8.3b). Δ

The existence of at least one solution pair (h, J) which satisfies the conditions of Lemma 8.2 is established by an argument due to Ross [16, 15] : The expected discounted cost function associated with the one-step cost d is denoted by D_β whenever the policy f^η is used over the infinite horizon and the discount factor is $\beta < 1$. In that case, it is plain that for all x in $\mathbb{N}^2$,

$$D_\beta(x) = d(x) + \beta \sum_y p(x,y;f^\eta(x)) D_\beta(y) \qquad (8.6)$$

and with the definition $h_\beta(x) := D_\beta(x) - D_\beta(0)$ for all x in $\mathbb{N}^2$, it follows by inspection

$$(1-\beta)D_\beta(0) + h_\beta(x) = d(x) + \beta \sum_y p(x,y;f^\eta(x)) h_\beta(y). \qquad (8.7)$$

This last remark is useful for establishing the following

Lemma 8.3. *Assume d to be bounded. With the notation and definition given above,*

(1): *There exists some constant $C > 0$ such that*

$$0 \le h_\beta(x) \le CZ(x) \qquad (8.8)$$

for all x in $\mathbb{N}^2$ and all β in $(0,1)$,

(2): *The convergence*

$$\lim_{\beta \uparrow 1} (1-\beta) D_\beta(0) = J^\eta = \mathbb{E}^\eta d(X) \qquad (8.9)$$

takes place, and

(3): *There exists a pair (h, J) which satisfies (8.8), with $J = J^\eta$ and*

$$h(x) := \lim_{\beta \uparrow 1} h_\beta(x) \qquad (8.10)$$

for all x in $\mathbb{N}^2$, the limit being taken along a subsequence.

Proof: (1): Fix $x \neq 0$ in $\mathbb{N}^2$. The monotonicity of the mapping d readily implies that

$$E^\eta[d(X(n)) \mid X(1)=0] \le E^\eta[d(X(n)) \mid X(1)=x], \qquad n=1,2,\ldots (8.11)$$

whence $0 \le h_\beta(x)$. Since $\tau_1 = \nu(1)$, standard arguments yield

$$D_\beta(x) - D_\beta(0) = E^\eta\left[\sum_{i=1}^{\tau_1 - 1} \beta^i d(X(i)) \mid X(1)=x\right] + D_\beta(0) E^\eta\left[\beta^{\tau_1} - 1 \mid X(1)=x\right] \qquad (8.12)$$

It is clear that

$$0 \le E^\eta\left[\sum_{i=1}^{\tau_1 - 1} \beta^i d(X(i)) \mid X(1)=x\right] \le |d| E^\eta[\tau_1 \mid X(1)=x] \qquad (8.13)$$

whereas the easy bound

$$0 \le D_\beta(0) \le \frac{|d|}{1-\beta} \qquad (8.14)$$

immediately implies

$$|D_\beta(0) E^\eta\left[\beta^{\tau_1} - 1 \mid X(1)=x\right]| \le |d| E^\eta[\tau_1 \mid X(1)=x] \qquad (8.15)$$

by standard properties of geometric series. Combining (8.12), (8.13) and (8.15) readily leads to

$$0 \le h_\beta(x) \le 2|d| E^\eta[\tau_1 \mid X(1)=x] \le \frac{2|d|}{1-\rho} Z(x), \qquad (8.16)$$

where the last inequality follows from (5.5), and (8.8) is thus established.

(2). This result readily follows from (6.1), (8.2) and the version of the Tauberian Theorem stated in Prop. 4-7 of ([5] , pp. 173).

(3): Owing to (8.8), the mapping $\beta \rightarrow h_\beta(x)$ is bounded for all x in $\mathbb{N}^2$. A simple diagonalization argument then implies the existence of a subsequence $\{\beta_n\}_1^\infty$ in $[0,1]$, with $\beta_n \uparrow 1$ as $n \uparrow \infty$, along which the sequence $\{h_{\beta_n}(x)\}_1^\infty$ has a well-defined limit for all x in $\mathbb{N}^2$. This well-defined limit, denoted by h, clearly satisfies (8.8) and so, by dominated convergence,

$$\lim_{n\to\infty} \beta_n \sum_y p(x,y;f^\eta(x))h_{\beta_n}(y) = \sum_y p(x,y;f^\eta(x))h(y) \tag{8.17}$$

for all x in $\mathbb{N}^2$. Upon taking the limit in (8.7), these remarks and (8.9) readily imply that the pair (h, J^η) indeed solves the equation (8.3a). That h satisfies the condition (8.3b) is easily obtained from the form of the bound (8.7) it satisfies and from the uniform integrability of the RV's $\{X(n)\}_1^\infty$ [17].

Δ

Lemma 8.4. *Under the assumptions of Theorem 8.1 and Lemma 8.3., the convergence*

$$\lim_{n\to\infty} \frac{1}{n} E^\alpha \sum_{i=1}^n d(X(i)) = \lim_{n\to\infty} \frac{1}{n} E^\eta \sum_{i=1}^n d(X(i)) = \mathbb{E}^\eta d(X) \tag{8.18}$$

takes place.

Proof: The notation being the one introduced in the discussion of Lemma 8.3, the sequence $\{\Phi(n)\}_1^\infty$ of $\mathbb{R}$-valued RV's is defined to be

$$\Phi(n) = E^\alpha[h(X(n+1)) \mid \mathcal{F}_n] - E^\eta[h(X(n+1)) \mid \mathcal{F}_n]. \qquad n=1,2,\ldots (8.19)$$

This definition makes sense owing to Lemma 8.3 and to the uniform integrability stated in Theorem 5.2.

As in the work of Mandl ([11] p. 46), the sequence $\{Y(n)\}_1^\infty$ of $\mathbb{R}$-valued RV's is defined by

$$Y(n+1) := h(X(n+1)) - E^\alpha[h(X(n+1)) \mid \mathcal{F}_n] \qquad n=1,2,\ldots (8.20)$$

with $Y(1) = h(X(1)) - E^\alpha[h(X(1))]$. This sequence $\{Y(n)\}_1^\infty$ plays a key role owing to the fact that it forms a $(P^\alpha, \mathcal{F}_n)$-*martingale difference* sequence. Moreover, (8.4), (8.19) and (8.20) readily yield

$$\Phi(n) = d(X(n)) - J^\eta + h(X(n+1)) - h(X(n)) - Y(n+1), \tag{8.21}$$

for all $n = 1,2,\ldots$.

The bounds (8.8) and Theorem 5.2 combine to imply the *uniform integrability* of the sequence of RV's $\{h(X(n))\}_1^\infty$ under the probability measure associated to *any* non-idling admissible policy, in particular any SR policy and the adaptive policy α. The estimate

$$E^\alpha\left(\sum_{n=1}^\infty \frac{Y^2(n)}{n^2}\right) < 4C^2 E^\alpha\left(\sum_{n=1}^\infty \frac{|Z(X(n))|^2}{n^2}\right) < \infty \tag{8.22}$$

readily follows, and a martingale version of the Law of Large Numbers ([11] , Theorem 3) thus applies to yield

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}Y(i)=0, \qquad P^{\alpha}\text{-a.s.} \tag{8.23}$$

whence

$$\lim_{n\to\infty}E^{\alpha}\frac{1}{n}\sum_{i=1}^{n}Y(i)=0 \tag{8.24}$$

by the noted uniform integrability of the RV's $\{h(X(n))\}_1^{\infty}$ under the probability measure P^{α}.

The convergence in probability to the non-random value η (under P^{α}) of the estimate sequence $\{\eta(n)\}_1^{\infty}$ readily implies that the sequence $\{\Phi(n)\}_1^{\infty}$ converges in probability to 0 under P^{α}; see [17] for details. Therefore,

$$\lim_{n\to\infty}E^{\alpha}\frac{1}{n}\sum_{i=1}^{n}\Phi(i)=0 \tag{8.25}$$

owing to the uniform integrability of the sequence of RV's $\{\Phi(n)\}_1^{\infty}$.

Iteration of (8.21) implies

$$J^{\eta}+\frac{1}{n}E^{\alpha}\sum_{i=1}^{n}\Phi(i)+\frac{1}{n}E^{\alpha}\sum_{i=1}^{n}Y(i)$$

$$=\frac{1}{n}E^{\alpha}\sum_{i=1}^{n}d(X(i))+\frac{1}{n}\{E^{\alpha}h(X(n+1))-E^{\alpha}h(X(1))\} \qquad n=1,2,\ldots \tag{8.26}$$

The result (8.24) is now easily obtained upon taking the limit in (8.26), owing to (8.3b), (8.24) and (8.25). Δ

Proof of Theorem 8.1:

Start with d given by $d(x)=x^{v}$ and for each $N>0$, define the bounded mapping $d^{N}:\mathbb{N}^2\to\mathbb{R}$ by $d^{N}(x):=x^{v}\wedge N$. Since $\eta(n)\to\eta$ in probability under P^{α}, Lemma 8.4 implies that

$$\lim_{n\to\infty}\frac{1}{n}E^{\alpha}\sum_{i=1}^{n}d^{N}(X(i))=\mathbb{E}^{\eta}d^{N}(X). \tag{8.27}$$

On the other hand, the Monotone Convergence Theorem gives

$$\lim_{n\to\infty}\frac{1}{n}E^{\alpha}\sum_{i=1}^{n}d(X(i))=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\lim_{N\uparrow\infty}E^{\alpha}d^{N}(X(i)) \tag{8.28}$$

and the RV's $\{d(X(n))\}_1^{\infty}$ being uniformly integrable, it follows that the inner convergence is clearly uniform in N. Consequently, (8.27) and (8.28) yield

$$\lim_{n\to\infty}E^{\alpha}\frac{1}{n}\sum_{i=1}^{n}d(X(i))=\lim_{N\uparrow\infty}\lim_{n\to\infty}\frac{1}{n}E^{\alpha}\sum_{i=1}^{n}d^{N}(X(i))=\lim_{N\to\infty}\mathbb{E}^{\eta}d^{N}(X). \tag{8.29}$$

Monotone convergence is again invoked to obtain

$$\lim_{N\to\infty}\mathbb{E}^{\eta}d^{N}(X)=\mathbb{E}^{\eta}d(X). \tag{8.30}$$

and the result (8.1a) is thus obtained by combining (8.29) and (8.30). A similar argument could be made to get (8.1b), this time with $d(x)=x^{\delta}$. Δ

REFERENCES:

[1] G. Fayolle and R. Iasnogorodski, "Two coupled processors: The reduction to a Riemann-Hilbert problem," *Z. Wahr. verw. Gebiete* vol. 47, pp. 325-351 (1979).

[2] J. S. Baras, A. J. Dorsey, and A. M. Makowski, "Discrete time competing queues with geometric service requirements: stability, parameter estimation and adaptive control," *SIAM J. Control Opt.*, Under revision. Invited paper to the ORSA/TIMS National Meeting, San-Francisco, California, (May 1984).

[3] J. S. Baras, A. J. Dorsey, and A. M. Makowski, "Two competing queues with geometric service requirements and linear costs: the mu-c rule is often optimal," *Adv. Appl. Prob.* vol. 17 , pp. 186-209 (March 1985).

[4] P. Billingsley, *Convergence of Probability Measures*, John Wiley, New York (1968).

[5] D. P. Heyman and M. J. Sobel, *Stochastic Models in Operations Research, Volume II: Stochastic Optimization*, MacGraw-Hill, New York (1984).

[6] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, Academic Press, New York (1975).

[7] P. R. Kumar, "A survey of some results in stochastic adaptive control," *SIAM J. Control Opt.* vol. 23, no. 3, pp. 329-380 (May 1985).

[8] H. J. Kushner and A. Shwartz, "An invariant measure approach to the convergence of Stochastic Approximations with state-dependent noise," *SIAM J. Control Opt.* vol. 22, no. 1, pp. 13-27 (January 1984).

[9] H. J. Kushner and A. Shwartz, "Weak convergence and asymptotic properties of adaptive filters with constant gains," *IEEE Trans. Info. Theory* vol. IT-30, no. 1, pp. 177-182 (March 1984).

[10] H. J. Kushner and A. Shwartz, "Stochastic Approximations in Hilbert space: identification and optimization of linear continuous-parameter systems," *SIAM J. Control Opt.* vol. 23, no. 5, pp. 774-793 (September 1985).

[11] P. Mandl, "Estimation and control in Markov chains," *Adv. Appl. Prob.* vol. 6, pp. 40-60 (1974).

[12] P. Nain and K. W. Ross, *Optimal Priority Assignment with Constraints*, Rapport de Recherche No. 346, INRIA - Rocquencourt, France (November 1984).

[13] H. Robbins and S. Monro, "A Stochastic Approximation method," *Ann. Math. Stat.* vol. 22, pp. 400-407 (1951).

[14] K. W. Ross, *Constrained Markov Decision Processes with Queueing Applications*, Ph. D. thesis, Computer, Information and Control Engineering, University of Michigan (1985).

[15] S. M. Ross, *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco (1970).

[16] S. M. Ross, *Introduction to Stochastic Dynamic Programming*, Academic Press (1984).

[17] A. Shwartz and A. M. Makowski, *Adaptive schemes for server allocation in systems of competing queues with constraints*, Systems Research Report, In preparation, Systems Research Center, University of Maryland at College Park. (1986)