TR-86-64

ADAPTIVE POLICIES FOR A SYSTEM OF COMPETING QUEUES I: CONVERGENCE RESULTS FOR THE LONG-RUN AVERAGE COST

by

Adam Shwartz & Armand M. Makowski

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **1986**	2. REPORT TYPE	3. DATES COVERED **00-00-1986 to 00-00-1986**
4. TITLE AND SUBTITLE **Adaptive Policies for a System of Competing Queues I: Convergene Results for the Long-Run Average Cost**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Maryland,Electrical Engineering Department,College Park,MD,20742**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT **see report**		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**41**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

ADAPTIVE POLICIES FOR A SYSTEM OF COMPETING QUEUES I: CONVERGENCE RESULTS FOR THE LONG-RUN AVERAGE COST

by

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

ABSTRACT

This paper considers a system of discrete-time queues competing for the attention of a single geometric server. The problem of implementing a given Markov stationary service allocation policy g through an adaptive allocation policy α is posed and convergence of the long-run average cost under such adaptive policy α to the long-run average cost under the policy g is investigated. Such question typically arises in the context of Markov decision problems associated with this queueing system, say when some of the model parameters are not available [1, 20] , or when the optimality criterion incorporates constraints [14, 21, 20] .

Conditions are given so that the long-run average cost under the policy α converges to the corresponding cost under the policy g , provided a natural condition on the relative asymptotic behavior of the policies g and α holds. Applications of the results developed here are discussed in a companion paper [20] . However, the ideas of this paper are of independent interest and should prove useful in studying implementation and adaptive control issues for broad classes of Markov decision problems [12] .

[1]Electrical Engineering Department, Technion-Israel Institute of Technology, Haifa 32000, ISRAEL. The work of this author was supported partially a Grant from AT&T Bell Laboratories and partially through United-States Israel Binational Science Foundation Grant BSF 85-00306.

[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

1. INTRODUCTION:

Consider the following system of K+1 infinite-capacity queues that compete for the use of a single server: Time is slotted with the service requirement of each customer corresponding exactly to one time slot. At the beginning of each time slot, the controller gives priority to one of the queues according to some prespecified dynamic priority assignment, on the basis of available information, and the selected queue is given service attention during that slot. However, due to a variety of reasons ranging from server failure to exogenous interferences, with a positive probability, the service will fail; in that case, the service of that customer will be rescheduled at a later time in accordance with the service allocation policy. When the service does not fail in a given time slot, the customer is declared serviced and leaves the system at the end of the slot. In the present paper, the failures are assumed generated through *independent Bernoulli* processes, with possibly class-dependent parameters, and this *independently* of the arrival mechanism. New customers may arrive in batches, which are modelled as an arbitrary $(K+1)$-dimensional *renewal* process, to capture partial correlations between arrivals from different classes in a given slot.

For this system of competing queues, the selection of a service allocation strategy with good performance properties has been discussed in a series of recent papers: Baras, Dorsey and Makowski [2] discussed the model in the case $K=1$, and showed the optimality of the μc-rule when the cost is linear in the queue sizes. This result was further extended to an arbitrary number of customer classes, under weaker statistical assumption on the arrival stream, in the works of Baras, Ma and Makowski [3] and of Buyyukkoc, Varaiya and Walrand [4] .

In [14] , Nain and Ross considered the situation where several types of traffic, e. g., voice, video and data, compete for the use of a single synchronous communication channel. They formulate this situation as a system of $K+1$ discrete-time queues that compete for the attention of a single server, and solve for the service allocation strategy that minimizes the long-run average of a linear expression in the queue sizes of K customer classes, under the constraint that the long-run average queue size of the remaining customer class does not exceed a certain value. Extending some of the optimality results from Baras, Ma and Makowski [3] , they show that if the constraint can be met, then the following policy is optimal: There exists a pair of *static* work-conserving service assignment policies (of which μc-rules are only one description), say f^0 and f^1, with the property that if there are customers in the system, a biased coin is flipped with bias η^*, and channel right is implemented according to the outcome via f^0 and f^1 with probability η^* and $1-\eta^*$, respectively.

Typically, as these two examples indicate, analysis will identify a Markov stationary policy g which exhibits optimality properties with respect to a long-run average cost criterion. Unfortunately, this Markov stationary policy is usually *not* readily *implementable*, despite its strong structural properties, with the encountered difficulties falling essentially into one the two following categories:

(i): The form of the policy g is a function of the various parameters determining the statistical description of the model. The actual *values* of these parameters are often *not* available to the decision-maker and need to be estimated as part of the system operation. Such a situation was considered by Baras, Dorsey and Makowski [1] for a long-run average cost linear in queue sizes, when the failure rates where not available.

(ii): Even in the event the actual parameter values are available, the Markov stationary policy g need still not be implementable due to *computational* difficulties inherent to its definition. The situation treated by Nain and Ross [14] is a good case in point, for non-trivial off-line computations are required in order to compute the actual value of η^* and implement the seemingly simple randomized policy discussed earlier.

Various methods have been proposed in the literature to overcome these implementation difficulties. In most cases, the solution amounts to generating an alternate policy α through the Certainty Equivalence Principle via a specific estimation scheme that exploits the specific structure of the Markov stationary policy g. Such a policy α will be referred to as an *adaptive* implementation of the policy g, thus broadening the technical meaning of the word "adaptive" as understood in the literature on the non-Bayesian adaptive control problem for Markov chains [11] .

In this context, it is then natural to investigate when the performance measures of interest coincide under these two policies. The main result of this paper, given in Theorem 3.1, can be viewed as an extension of a result by Mandl [13] to *randomized* strategies and *countable* state spaces. It gives sufficient conditions for the performance measures, when taken in the long-run average sense, to coincide under the two policies. Although this result is discussed in the context of competing queues systems, the methodology has broader applicability to various issues in the theory of Markov decision processes. The ideas and results presented here are applied on various problems in a companion paper [20] .

The paper is organized as follows: The model and basic assumptions are described in Section 2. In Section 3, some motivation is provided for the issues discussed in this paper, and the main convergence result for the cost given as Theorem 3.1. Its proof underlies most of the material discussed in subsequent sections. The structure of passage times to the empty state is studied in Section 4 under the action of arbitrary admissible non-idling strategies, and the results are exploited in Section 5 to derive the statistical properties of the busy cycles. In Section 6, bounds on various moments of the queue size process are established through a renewal argument that exploits the statistical properties of the busy periods; moreover, a representation is obtained for the cost, in terms of the invariant measure associated with the policy g. The proof of the main result, Theorem 3.1, is then an easy consequence of a useful extension to a well-known result of Mandl [13] on the optimality of adaptive policies. This topic is discussed in Section 7.

A few words on the notation used throughout the paper: The set of all non-negative integers is denoted by $\mathbb{N}$ while $\mathbb{R}$ stands for the set of all real numbers. An element x in

$\mathbb{R}^{K+1}$ will sometimes be written as a $(K+1)$-tuple $(x_0, x_1, \ldots, x_K)$, with the notation $|x| := \sum_{k=0}^{K} x_k$. As a natural convention, the k-th component of any element of $\mathbb{R}^{K+1}$ is denoted by the same symbol as this element, but subscripted by k, $0 \le k \le K$, with a similar convention for random variables. In particular, the element in $\mathbb{R}^{K+1}$ whose components are all zero is also denoted by 0. The standard basis for $\mathbb{R}^{K+1}$ is denoted by $\mathbb{B} = \{e_k\}_0^K$, while S is the standard K-simplex, i. e.,

$$S := \{ p \in \mathbb{R}^{K+1} : \sum_{k=0}^{K} p_k = 1 \text{ and } 0 \le p_k \le 1, \; 0 \le k \le K \}.$$

The indicator function of a set A is denoted by $I(A)$ and the Kronecker delta is denoted by $\delta(\cdot,\cdot)$, with $\delta(a,b)=1$ if $a=b$ and $\delta(a,b)=0$ otherwise. For any mapping $h : \mathbb{N}^{K+1} \to \mathbb{R}$, it is convenient to pose $|h| := \sup_x |h(x)|$.

2. MODEL AND ASSUMPTIONS:

The basic random variables

In this paper, all probabilistic elements are defined on a single sample space Ω equipped with the σ-field of events $\mathbb{F}$. This sample space carries the basic random variables (RV's) Ξ, $\{U(n)\}_1^\infty$, $\{A(n)\}_1^\infty$ and $\{B(n)\}_1^\infty$ which take values in $\mathbb{N}^{K+1}$, $\mathbb{B}$, $\mathbb{N}^{K+1}$ and $\{0,1\}^{K+1}$, respectively. It is convenient to introduce the *information* RV's $\{H(n)\}_1^\infty$, which are recursively defined by $H(1) := \Xi$ and

$$H(n+1) := (H(n), U(n), A(n), B(n)) \qquad n=1,2,\ldots (2.1)$$

and which take values in the corresponding *information* spaces $\{\mathbb{H}_n\}_1^\infty$ where $\mathbb{H}_1 := \mathbb{N}^{K+1}$ and $\mathbb{H}_{n+1} := \mathbb{H}_n \times \mathbb{B} \times \mathbb{N}^{K+1} \times \{0,1\}^{K+1}$ for all $n=1,2,\ldots$.

These quantities have a ready interpretation in the context of the situation described in the introduction: The number of customers initially in the k-th queue is set at Ξ_k and for each $n=1,2,\ldots$, the state of the system is represented by a RV $X(n)$ of integer components with the interpretation that at the beginning of the slot $[n, n+1)$, $X_k(n)$ customers are stored in the k-th buffer, including the one receiving service. Thus at that time,

(i): control action $U(n)$ is selected with the convention that $U_k(n)=1$ (resp. $U_k(n)=0$) if the k-th queue is (resp. is not) given service attention during that slot;

(ii): new packets arrive into the system according to the RV $A(n)$ in that $A_k(n)$ new customers join the k-th queue, and

(iii): completions of transmission are encoded in the binary RV $B(n)$; here $B_k(n)=1$ (resp. $B_k(n)=0$) signifies successful completion (resp. abortion) of service for the k-th queue conditioned on it being served.

As a result, the successive system states or queue sizes form a sequence $\{X(n)\}_1^\infty$ of $\mathbb{N}^{K+1}$-valued RV's which are generated componentwise through the recursion

$$X_k(n+1) = X_k(n) + A_k(n) - I[X_k(n) \neq 0] U_k(n) B_k(n), \quad 0 \leq k \leq K, \qquad n=1,2,\ldots (2.2)$$

with $X_1 := \Xi$.

At the beginning of each time slot $[n, n+1)$, the channel controller has access to the initial queue sizes Ξ, the past arrival pattern $A(i)$, $1 \leq i < n$, the past decisions $U(i)$, $1 \leq i < n$, and the service history $B(i)$, $1 \leq i < n$. Thus, the channel controller has knowledge of the RV $H(n)$ which is used to generate 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 probabilistic structure

Since randomized strategies are allowed, an admissible control policy π is defined as any collection $\{\pi_n\}_1^\infty$ of mappings $\pi_n : \mathbb{H}_n \to S$, with the interpretation that at times $n=1,2,\ldots$ the k-th queue is given service attention with probability $\pi_n(k; h_n)$ whenever the information vector h_n is available to the system controller. Denote the collection of all such admissible policies by Π.

For each $n=1,2,\ldots$ let $\mathbb{F}_n$ denote the σ-field on the sample space Ω generated by the RV $H(n)$, with $\mathbb{F}_n \subseteq \mathbb{F}_{n-1}$.

Let $q_\Xi(\cdot)$ and $q(\cdot)$ be two probability distributions on $\mathbb{N}^{K+1}$, with $q(0) < 0$, and fix a service rate vector μ in $(0.1]^{K+1}$. The model is now completely specified by ***postulating*** the existence of a family $\{P^\pi, \pi \in \Pi\}$ of probability measures on the σ-field $\mathbb{F}$ which satisfies the requirements (R1)-(R3) below, i. e., for every policy π in Π,

(R1): For all x in $\mathbb{N}^{K-1}$;

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

(R2): For all a in $\mathbb{N}^{K-1}$ and b in $\{0,1\}^{K+1}$,

$$P^\pi[A(n)=a, B(n)=b \mid \mathbb{F}_n \vee \sigma\{U(n)\}] := P^\pi[A(n)=a] P^\pi[B(n)=b]$$

$$:= q(a) \prod_{k=0}^{K} \left(b_k \mu_k + (1-b_k)(1-\mu_k) \right) \qquad n=1,2,\ldots$$

and

(R3): For all $e_k, 0 \leq k \leq K$, in $\mathbb{B}$,

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

The existence of a sample space $(\Omega, \mathbb{F})$ that carries such a family of probability measures $\{P^\pi, \pi \in \Pi\}$ is easily established via the Kolmogorov extension theorem, by taking Ω to be the *canonical* space $:= \mathbb{N}^{K+1} \times \left(\mathbb{B} \times \mathbb{N}^{K+1} \times \{0.1\}^{K+1} \right)^\infty$ equipped with its natural σ-

field. This modelling approach for the Markov decision process under consideration was adopted in [21] to which the reader is referred for additional information.

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

(P1): The $\mathbb{N}^{K+1}$-valued RV Ξ and the sequences of RV's $\{A(n)\}_1^{\infty}$ and $\{B(n)\}_1^{\infty}$ are *mutually independent;*

(P2): The sequences $\{B_k(n)\}_1^{\infty}$ of $\{0,1\}$-valued RV's are *mutually independent Bernoulli* sequences with parameters μ_k , $0 \leq k \leq K$;

(P3): The $\mathbb{N}^{K+1}$-valued RV's $\{A(n)\}_1^{\infty}$ form a sequence of *i.i.d* RV's, with a common distribution $q(\cdot)$; and

(P4): The probability transitions have the form

$$P^{\pi}[X(n+1)=y \mid \mathbb{F}_n]=p(X(n),y;\pi_n(H(n))) \qquad (2.3)$$

where, for all q in S,

$$p(x,y;q) := \sum_{k=0}^{K} q_k \; Q_k(x,y) , \qquad (2.4)$$

as x and y range over $\mathbb{N}^{K+1}$, with the definitions

$$Q_k(x,y) := P^{\pi}[x_k + A_k(n) - I[x_k \neq 0]B_k(n) = y_k ; \; x_j + A_j(n) = y_j , \; 0 \leq j \neq k \leq K] \qquad (2.5)$$

for all $0 \leq k \leq K$. Note that the right handsides of (2.5) are independent of n and of the policy π owing to the assumptions made earlier.

For future notational use, it will be convenient to assume that the sample space Ω carries an additional $\mathbb{N}^{K+1}$-valued RV $A(\infty)$ which is distributed according to $q(\cdot)$ and is *independent* of all other basic RV's introduced so far, under the probability measure associated with any policy π in Π.

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 $\{g_n\}_1^{\infty}$ of mappings $g_n : \mathbb{N}^{K+1} \to S$ such that

$$\pi_n(H(n)) = g_n(X(n)) \qquad P^{\pi}\text{-}a.s. \qquad n=1,2,\ldots(2.6)$$

with $\{X(n)\}_1^{\infty}$ generated through the recursion (2.2). In the event all the mappings $\{g_n\}_1^{\infty}$ are identical to a given mapping $g : \mathbb{N}^{K+1} \to S$, 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 $f_n : \mathbb{H}_n \to \mathbb{B}$ such that for all $0 \leq k \leq K$,

$$\pi_n(k\,;H(n)) = \delta(e_k, f_n(H(n))), \qquad P^{\pi}\text{-}a.s. \qquad n=1,2,...(2.7)$$

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 $f : \mathbb{N}^{K+1} \to \mathbb{B}$ to which it is substituted in the notation.

A policy π in Π is said to be *non-idling* or *work-conserving* whenever for all $0 \le k \le K$, the condition

$$\pi_n(k\,; H(n)) > 0 \textit{ implies either } X_k(n) \ne 0 \textit{ or } X(n)=0 \qquad n=1,2,...(2.8)$$

holds true P^{π}-$a.s.$, in which case, the P^{π}-$a.s.$ equality

$$\sum_{k=0}^{K} U_k(n) I[X_k(n) \ne 0] = 1 - I[X(n)=0] \qquad n=1,2,...(2.9)$$

necessarily follows.

3. CONVERGENCE FOR THE LONG-RUN AVERAGE COST:

Let λ_k be the first moment of the sequence of i.i.d RV's $\{A_k(n)\}_1^{\infty}$, $0 \le k \le K$, and for future use, define the *traffic coefficient* ρ to be

$$\rho := \sum_{k=0}^{K} \frac{\lambda_k}{\mu_k}. \tag{3.1}$$

Throughout this paper, the discussion is carried out under the assumption that $\rho < 1$, which expresses *stability* of the queueing system under any non-idling policy (as discussed in Sections 4-5).

Let c denote a mapping $\mathbb{N}^{K+1} \to \mathbb{R}$ and for any admissible policy π in Π, pose

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

with the usual interpretation that the quantity $J(\pi)$ is a measure of system performance when the policy π is in use. Analysis often identifies a Markov stationary policy g in Π which exhibits suitable performance properties with respect to the cost function (3.2). Various examples are now discussed in some detail so as to motivate the developments of the paper.

Some examples

When the cost-per-stage is *linear*, i.e., for all x in $\mathbb{N}^{K+1}$,

$$c(x) = \sum_{k=0}^{K} c_k x_k, \tag{3.3}$$

with $c_k \ge 0$, $0 \le k \le K$, several authors [2, 3, 4] showed that the μc-rule minimizes (3.2)-(3.3) over the class of all admissible policies Π. Here the μc-rule is the non-randomized Markov stationary policy g given by

$$g(k,x) = 1 \text{ if } x_l = 0 \text{ for } 0 \leq l < k \text{ and } x_k \neq 0, \quad 0 \leq k \leq K \tag{3.4}$$

for all $x \neq 0$ in $\mathbb{N}^{K+1}$, under the convenient assumption that $\mu_0 c_0 \geq \mu_1 c_1 \geq \dots \geq \mu_K c_K$. It is clear that implementation of this policy is conditioned on knowledge of the rate parameters μ_k, $0 \leq k \leq K$. The situation where such knowledge is not available was studied by Baras, Dorsey and Makowski [1] through use of the Certainty Equivalence Principle; the proposed adaptive μc-rule α based on the maximum likelihood estimates for the rate parameters was shown to be optimal in that $J(\alpha) = J(g)$.

For each V in $\mathbb{R}$, consider now the set Π_V of all admissible control policies in Π which satisfy the constraint

$$J(\pi) \leq V. \tag{3.5}$$

With the interpretation given earlier for the quantity $J(\pi)$, desirable system behavior may then be conveniently expressed in the form of a constraint (3.5), in that only admissible policies in Π_V need to be considered when running the system. This viewpoint was taken by Nain and Ross [14] in the study of a simple problem of channel allocation. There, as in many other Markov decision problems with constraints [17] , Lagrangian arguments often reduce the search of a constrained (optimal) policy to finding a Markov stationary policy g in Π that *saturates* the constraint, i.e.,

$$J(g) = V. \tag{3.6}$$

It should be pointed out that this last problem is of independent interest for it can be viewed more generally as one of *steering* the cost (3.2) to a particular value V determined through various design considerations.

This line of arguments typically proceeds by identifying two Markov stationary policies in Π, possibly randomized, say f^0 and f^1, with the property that

$$J(f^0) \leq V \leq J(f^1). \tag{3.7}$$

It then remains to construct from the policies f^0 and f^1 an admissible policy g in Π that satisfies (3.6). This construction is often achieved by *randomizing* between the policies f^0 and f^1. To that end, consider for any η with $0 \leq \eta \leq 1$, the *randomized* Markov stationary policy f^η with bias η generated through the mapping $f^\eta : \mathbb{N}^{K+1} \rightarrow S$ where

$$f^\eta(x) := \eta f^0(x) + \overline{\eta} f^1(x) \tag{3.8}$$

for all x in $\mathbb{N}^{K+1}$. Note that for $\eta=0$ (resp. $\eta=1$), f^η is identical to the original policy f^0 (resp. f^1). If the mapping $\eta \rightarrow J(f^\eta)$ is *strictly monotone* and *continuous* on the interval [0,1], then exactly one randomized strategy f^{η^*} meets the constraint, and its bias value η^* is the *unique* solution of the equation

$$J(f^{\eta}) = V, \ \eta \ in \ [0,1], \tag{3.9}$$

whence the identification $g = f^{\eta^*}$ may take place.

The determination of the optimal bias value η^* seemingly requires the evaluation of the expression $J(f^{\eta})$ for all values of η in the unit interval [0,1]. This is a non-trivial task even in the simplest of situations when $K=1$ and the policies f^0 and f^1 are static priority assignments as was the case in the model discussed by Nain and Ross [14] . As already pointed out there, for $0<\eta<1$, the two competing queues can be interpreted (under the probability measure P^{η} associated with the policy f^{η}) as a two processor-sharing system of the type studied by Fayolle and Iasnogorodski [7] . The computation of the probability generating function in equilibrium reduces to the solution of a Riemann-Hilbert problem, from which the optimal bias η^* could in principle be determined numerically as a function of the arrival statistics $q(\cdot)$ and of the rates vector μ. Moreover, there has been no success in extending these methods to the higher dimensional case $K\geq 2$. In short, even if the statistical model parameters were available to the decision-maker, the evaluation of η^* seems to constitute a formidable computational undertaking [14] .

To circumvent these difficulties, it seems natural to find alternate implementations of the randomized policy g identified through the analysis. To fix ideas, suppose both policies f^0 and f^1 to be *implementable.* The very definition of the policy f^{η^*} naturally suggests schemes where a sequence of [0,1]-valued RV $\{\eta(n)\}_1^{\infty}$, acting as estimates for the optimal bias value η^*, are substituted for it in the definition (3.8). The adaptive non-idling policy α can be formally defined through the mappings $\{\alpha_n\}_1^{\infty}$ with

$$\alpha_n(H(n)) := f^{\eta(n)}(X(n)). \qquad n=1,2,...(3.10)$$

Needless to say, the estimates $\{\eta(n)\}_1^{\infty}$ need to be generated, possibly via some *recursive* algorithm, so as to ensure that the corresponding policy α meets the constraint (3.6).

An example of such a scheme was proposed by Shwartz and Makowski [21] for the two competing queues situation treated by Nain and Ross [14] on the basis of well-known ideas from the theory of Stochastic Approximations, of which the Robbins-Monro scheme is the archetypical example [16] . The key idea being to solve *on-line* the constraint equation (3.6), the proposed scheme generates a sequence of bias values $\{\eta(n)\}_1^{\infty}$ through the recursion

$$\eta(n+1) = \Big[\ \eta(n) - a_n \Big(V - c(X(n+1)) \Big) \Big]_0^1 \qquad n=1,2,...(3.11)$$

with $\eta(0)$ 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 satisfies the conditions

$$0< a_n \to 0, \quad \sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} |a_{n+1} - a_n| < \infty. \tag{3.12}$$

In [20] the authors show that for the problem at hand, the policy α is optimal in that

$J(\alpha)=J(g)$.

A general convergence result

The two examples discussed above can both be accommodated into the following general framework: Let g be a Markov stationary policy in Π held fixed hereafter, and consider an admissible policy α in Π to be an implementation of it. The question of interest here can be formulated as one of finding *natural* conditions under which $J(\alpha) = J(g)$.

To that end, it will be convenient to say that an admissible policy α in Π satisfies the *convergence condition (C) (with respect to g)* if

(C): The RV's $\{\alpha_n(H(n)) - g(X(n))\}_1^\infty$ converge to 0 in probability under P^α, i.e., for every $\epsilon>0$,

$$\lim_{n\uparrow\infty} P^\alpha\left[\,|\alpha_n(k;H(n)) - g(k;X(n))| > \epsilon,\ 0\le k\le K\right] = 0.$$

This paper is devoted to the study of the performance properties of admissible policies α satisfying the convergence condition (C). To introduce the necessary hypotheses, it is convenient to define the mapping $Z: \mathbb{N}^{K+1} \to \mathbb{R}^+$ given by

$$Z(x) := \sum_{k=0}^{K} \frac{x_k}{\mu_k} \tag{3.14}$$

for all x in $\mathbb{N}^{K+1}$. The main result will be derived under the following technical conditions (H1)-(H4), where

(H1): The RV's $\{Z(X(n))\}_1^\infty$ are *uniformly integrable* under the probability measure P^g;

(H2): The RV's $\{Z(X(n))\}_1^\infty$ are *uniformly integrable* under the probability measure P^α;

(H3): The *growth* condition

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

holds;

(H4): The RV's $\{c(X(n))\}_1^\infty$ are *uniformly integrable* under the probability measure P^g, *and*

(H5): The RV's $\{c(X(n))\}_1^\infty$ are *uniformly integrable* under the probability measure P^α.

Theorem 3.1. *Under the foregoing assumptions (R1)-(R3), whenever the conditions (H1)-(H5) are enforced, and the policy α satisfies the convergence condition (C) with respect to the non-idling policy g, the convergence*

$$J(\alpha) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} c(X(i)) = J(g) \tag{3.15}$$

takes place in $L^1(\Omega, \mathbb{F}, P^\alpha)$, *whence*

$$J(\alpha) = \lim_{n \to \infty} \frac{1}{n} E^\alpha \sum_{i=1}^{n} c(X(i)) = J(g). \tag{3.16}$$

Proof: The result follows readily from Theorem 7.1. □

At this stage, the reader may wonder as to how easy it is to verify the conditions (H1)-(H5) from the basic data of the problem. Sufficient conditions for establishing (H1)-(H5) are now given in the form of the additional requirements (R4) and (R5) on the data of the problem, namely

(R4): There exists some constant $\gamma \geq 1$ such that for every policy π in Π, the moment conditions

$$E^\pi \left[\sum_{k=0}^{K} | \Xi_k |^\gamma \right] = \sum_{x \in \mathbb{N}^{K+1}} \left[\sum_{k=0}^{K} | x_k |^\gamma \right] q_\Xi(x) < \infty$$

and

$$E^\pi \left[\sum_{k=0}^{K} | A_k(n) |^\gamma \right] = \sum_{a \in \mathbb{N}^{K+1}} \left[\sum_{k=0}^{K} | a_k |^\gamma \right] q(a) < \infty \qquad n=1,2,..$$

hold true.

Moreover, the mapping c is assumed to satisfy the following growth condition (R5), where

(R5): There exists constants $\delta > 0$ and $L > 0$ in $\mathbb{R}$ such that

$$| c(x) | \leq L(1 + | x |^\delta)$$

for all $x \neq 0$ in $\mathbb{N}^{K+1}$.

Theorem 3.2. *Assume the policies* g *and* α *to be non-idling. Under the foregoing assumptions (R1)-(R5), assumptions (H1)-(H5) are satisfied whenever* γ *is such that*

$$\max\{3, 1 + \delta(1+\epsilon)\} \leq \gamma \tag{3.17}$$

for some $\epsilon > 0$.

Proof: The bound

$$0 \leq Z(x) \leq \left(\min_{0 \leq k \leq K} \mu_k \right)^{-1} | x |, \tag{3.18}$$

is obviously valid for all x in $\mathbb{N}^{K+1}$, whereas Theorem 6.1. gives

$$\sup_{n} E^{\pi}\left[\, |\, X(n)\,|^{\gamma-1} \right] < \infty \tag{3.19}$$

for *any* non-idling policy π in Π, since condition (3.17) implies $\gamma - 1 \geq 2$. The validity of the assumptions (H1)-(H3) is now immediate upon combining remarks (3.18)-(3.19). To obtain (H4)-(H5), it suffices to note that (3.17) implies (6.4) and to apply Corollary 6.1.1. □

An operational version of the main convergence result of Theorem 3.1 can now be given.

Theorem 3.1bis. *Assume the policy g to be non-idling. Under the foregoing assumptions (R1)-(R5) with (3.17), any non-idling policy α in Π which satisfies the convergence condition (C) with respect to g has the property that*

$$J(\alpha) = \lim_{n \uparrow \infty} \frac{1}{n} E^{\alpha} \sum_{i=1}^{n} c(X(i)) = J(g). \tag{3.20}$$

Proof: The result follows readily from combining Theorems 3.1 and 3.2. □

4. PASSAGE TIMES TO THE EMPTY STATE:

Throughout this section, let π be a fixed *non*-idling (not necessarily Markov) policy in Π. The results given below extend to randomized policies some of the results obtained by Baras, Dorsey and Makowski ([1] , Section 4) for non-randomized policies only. Here too the discussion focuses on generating *exponential* or Wald martingales which are rich enough to yield statistical information on various passage times for the queue size sequence $\{X(n)\}_1^{\infty}$. The quantities of interest include first passage times to the empty state (Theorem 4.1) and first exit times from the empty state (Lemma 4.2) in order to study the busy and idle periods of such a system.

At this stage, it is convenient to introduce the sequence $\{V(n)\}_1^{\infty}$ of $\{0,1\}^{K+1}$-valued RV's which are defined componentwise by

$$V_k(n) := I[X_k(n) \neq 0] U_k(n), \; 0 \leq k \leq K, \qquad n=1,2,... \tag{4.1}$$

Note that with this notation, the relation (2.9) (valid for the non-idling policy π) readily translates into the identity

$$\sum_{k=0}^{K} V_k(n) = I[X(n) \neq 0] \qquad P^{\pi}\text{-}a.s \qquad n=1,2,... \tag{4.2}$$

The next proposition is essential for generating a $(P^{\pi}, \mathbb{F}_n)$-martingale of interest: it presents a variation of a basic relation already given by Baras, Dorsey and Makowski in [1] . For ease of notation, let $\{\mathbf{F}_n^*\}_1^{\infty}$ denote the filtration on the sample space Ω defined by

$$\mathbb{F}_n^* := \mathbb{F}_n \vee \sigma\{U(n)\} = \sigma\{H(n), U(n)\}. \qquad n=1,2... (4.3)$$

Lemma 4.1. *Under the foregoing assumptions (R1)-(R4), for every* z *in* $(0,1]^{K+1}$, *the equality*

$$E^\pi\left[\prod_{k=0}^{K} z_k^{X_k(n+1)} \mid \mathbb{F}_n^*\right] = a(z)\prod_{k=0}^{K} b_k(z_k)^{V_k(n)} \prod_{k=0}^{K} z_k^{X_k(n)} \qquad n=1,2,... (4.4)$$

holds true, with

$$a(z) := E^\pi[\prod_{k=0}^{K} z_k^{A_k(n)}] \qquad (4.5a)$$

and

$$b_k(z_k) := E^\pi[z_k^{-B_k(n)}] = \frac{\mu_k}{z_k} + (1-\mu_k), \quad 0 \le k \le K. \qquad (4.5b)$$

Note that the right handsides of (4.5) depend neither on the policy π nor on the time index n owing to the assumption (R2).

Proof: Upon substitution of the the system dynamics (2.2) into the left handside of (4.4), easy calculations lead to the relation

$$E^\pi\left[\prod_{k=0}^{K} z_k^{X_k(n+1)} \mid \mathbb{F}_n^*\right] \qquad n=1,2....$$

$$= \prod_{k=0}^{K} z_k^{X_k(n)} E^\pi\left[\prod_{k=0}^{K} z_k^{A_k(n)-v_k B_k(n)} \mid \mathbb{F}_n^*\right]_{v=V(n)} \qquad (4.6)$$

where the $\mathbb{F}_n^*$-measurability of the RV's $X(n)$ and $V(n)$ has been used.

Owing to the assumption (R2) imposed on the probability measure P^π, the mutually independent RV's $A(n)$ and $B(n)$ are also seen to be independent of the σ-field $\mathbb{F}_n^*$. Consequently, for all v in $\{0,1\}^{K-1}$, the relation

$$E^\pi\left[\prod_{k=0}^{K} z_k^{A_k(n)-v_k B_k(n)} \mid \mathbb{F}_n^*\right]$$

$$= E^\pi[\prod_{k=0}^{K} z_k^{A_k(n)}] \prod_{k=0}^{K}\left(E^\pi[z_k^{-B_k(n)}]\right)^{v_k}. \qquad n=1,2,... (4.7)$$

holds true, with the mutual independence of the components of the RV $B(n)$ being used to get the product form of the second factor. Substitution of (4.7) into (4.6) readily yields (4.4) with the notation (4.5). □

In view of (4.2), the relation (4.4) is expected to simplify whenever *all* $K+1$ factors $b_k(z_k)$, $0 \leq k \leq K$ are equal. To that end, note that each one of the mappings $z_k \rightarrow b_k(z_k)$, $0 \leq k \leq K$, *decreases monotonically* from $+\infty$ to 1 on the interval (0,1]. Hence, for each b in the interval [1, $+\infty$), there exists exactly one element in $(0,1]^{K+1}$ with the property that

$$b_k(z_k) = \frac{\mu_k}{z_k} + (1-\mu_k) = b \quad . \quad 0 \leq k \leq K. \tag{4.8}$$

In fact, a simple calculation shows that the components of this element, denoted throughout by $z(b)$, are given by

$$z_k(b) = \frac{\mu_k}{\mu_k - b - 1}, \quad 0 \leq k \leq K. \tag{4.9}$$

Let $\{L(n)\}_1^\infty$ denote the $\mathbb{N}$-valued RV's that count the time (expressed in slots) spent in the empty state, i. e.,

$$L(n+1) := \sum_{i=1}^{n} I[\, X(i) = 0\,] \qquad n = 1,2,... \tag{4.10}$$

with $L(1) := 0$. For every b in [1, $+\infty$), the $\mathbb{R}^+$-valued RV's $\{M(n,b)\}_1^\infty$ are defined by

$$M(n,b) = \frac{\prod_{k=0}^{K} z_k(b)^{X_k(n)}}{r(b)^n}\, b^{L(n)} \tag{4.11}$$

where for notational convenience,

$$r(b) := a(z(b)) \cdot b \tag{4.12}$$

Proposition 4.1. *Under the foregoing assumptions (R1)-(R4), the RV's* $\{M(n,b)\}_1^\infty$ *form an integrable positive* $(P^z, \mathbb{F}_n)$*-martingale for every* $b \geq 1$.

Proof: The integrability of the RV's $\{M(n,b)\}_1^\infty$ is readily established from the easy bounds

$$0 < M(n,b) \leq \frac{b^{L(n)}}{r(b)^n} \leq \left[\frac{b}{r(b)}\right]^n . \tag{4.13}$$

valid for all n=1,2,... and $b \geq 1$.

For every b in the interval [1, $+\infty$), the relations $b_k(z_k(b)) = b$, $0 \leq k \leq K$ hold true by the very definition (4.9) of $z(b)$, and it is now plain that

$$\prod_{k=0}^{K} b_k (z_k (b))^{V_k(n)} = b^{\sum_{k=0}^{K} V_k(n)} = b^{1-I[X(n)=0]} \qquad n=1,2,\ldots(4.14)$$

where the last equality is obtained with the help of (4.2). Substitution of (4.14) into (4.4) (with $z=z(b)$) implies

$$E^{\pi}\left[\prod_{k=0}^{K} z_k (b)^{X_k(n+1)} \mid \mathbb{F}_n^{*}\right] = a(z(b)) b^{1-I[X(n)=0]} \prod_{k=0}^{K} z_k (b)^{X_k(n)} \tag{4.15}$$

and the RV $X(n)$ being $\mathbb{F}_n$- measurable with $\mathbb{F}_n \subseteq \mathbb{F}_n^{*}$, an elementary smoothing argument now yields

$$E^{\pi}\left[\prod_{k=0}^{K} z_k (b)^{X_k(n-1)} \mid \mathbb{F}_n\right] = r(b) \frac{\prod_{k=0}^{K} z_k (b)^{X_k(n)}}{b^{I[X(n)=0]}} \qquad n=1,2,\ldots(4.16)$$

The martingale property of the sequence $\{M(n,b)\}_1^{\infty}$ is now immediate from (4.16). □

The particular structure of the martingales $\{M(n,b)\}_1^{\infty}$, $b \geq 1$, is now exploited to obtain information on the first passage times to the empty state. To that end, if σ is any arbitrary $\mathbb{F}_n$-stopping time, let $\nu(\sigma)$ be the $\mathbb{N}$-valued RV defined by

$$\nu(\sigma) := \inf \{n \geq 1 : X(\sigma+n)=0\} \quad \text{if } \sigma<\infty, \tag{4.17}$$

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 $\mathbb{F}_n$-stopping time.

Theorem 4.1. *Under the foregoing assumptions (R1)-(R4), the conservation law*

$$E^{\pi}\left[I[\sigma<\infty, \tau(\sigma)<\infty] \frac{1}{r(b)^{\nu(\sigma)}} \mid \mathbb{F}_{\sigma}\right] = I[\sigma<\infty] \prod_{k=0}^{K} z_k (b)^{X_k(\sigma)} b^{-I[X(\sigma)=0]} \tag{4.18}$$

holds true P^{π}–a.s. for all $b \geq 1$.

Proof: See Appendix.

An immediate consequence of this result is stated in the following

Corollary 4.1.1. *Under the assumptions of Theorem 4.1., the relation*

$$P^{\pi}[\sigma<\infty, \nu(\sigma)<\infty \mid \mathbb{F}_{\sigma}] = I[\sigma<\infty] \qquad P^{\pi}\text{–a.s.} \tag{4.19}$$

holds true, and in particular, if $\sigma<\infty$ P^{π}–a.s. , then necessarily $\nu(\sigma)<\infty$ P^{π}–a.s.

Proof: The events $[\sigma<\infty, \tau(\sigma)<\infty]$ and $[\sigma<\infty, \nu(\sigma)<\infty]$ coincide, and the result (4.19) follows readily by letting the variable b go to 1 monotonically in (4.18) and using the Monotone Convergence Theorem for conditional expectations. The second part of the corollary is now immediate. □

As the forthcoming discussion will show, the conservation law given in Theorem 4.1 can be used to study the structure of the busy cycles of the chain $\{X(n)\}_1^\infty$. To that end, for any $\mathbb{F}_n$-stopping time σ, define the $\mathbb{N}$-valued RV $\beta(\sigma)$ by

$$\beta(\sigma) := \inf\{ n \geq 0 : A(\sigma+n) \neq 0 \} \quad \text{if } \sigma<\infty, \tag{4.20}$$

with the convention that $\beta(\sigma)=\infty$ whenever this set is empty or when $\sigma=\infty$; the RV $\gamma(\sigma) := \sigma + \beta(\sigma)$ is clearly an $\mathbb{F}_n$-stopping time. Several useful properties of the RV's $\beta(\sigma)$, $\gamma(\sigma)$ and $A(\gamma(\sigma))$ are now given for easy reference.

Lemma 4.2 *Under the foregoing assumptions (R1)-(R4), for every $r>1$ and every z in $(0,1]^{K+1}$, the relation*

$$E^\pi\left[I[\sigma<\infty, \beta(\sigma)<\infty] r^{-\beta(\sigma)} \prod_{k=0}^{K} z_k^{A_k(\gamma(\sigma))} \mid \mathbb{F}_\sigma \right] = \frac{r[a(z) - q(0)]}{r - q(0)} I[\sigma<\infty] \tag{4.21}$$

holds true P^π-a.s., and consequently

$$P^\pi[\sigma<\infty, \beta(\sigma)=l \mid \mathbb{F}_\sigma] = I[\sigma<\infty] q(0)^l [1 - q(0)]. \qquad l=0,1,\ldots \tag{4.22}$$

In the particular case when $\sigma<\infty$ P^π*-a.s*, Lemma 4.2 thus implies that the RV $\beta(\sigma)$ is also P^π*-a.s.* finite, has a *geometric* distribution with parameter $q(0)$ and is *independent* of the RV $A(\gamma(\sigma))$; moreover these two RV's are (jointly) *independent* from the σ-field $\mathbb{F}_\sigma$.

Proof: It is plain from the definition of the RV $\beta(\sigma)$ that for all $l=0,1,\ldots$,

$$[\beta(\sigma) = l] = [A(\sigma+j)=0, 0\leq j<l; A(\sigma+l)\neq 0]. \tag{4.23}$$

For $l\neq 0$ in $\mathbb{N}$, a smoothing argument using the inclusion $\mathbb{F}_\sigma \subseteq \mathbb{F}_{\sigma+(l-1)}$ leads to the chain of equalities

$$E^\pi\left[I[\sigma<\infty, \beta(\sigma)=l] \prod_{k=0}^{K} z_k^{A_k(\gamma(\sigma))} \mid \mathbb{F}_\sigma \right]$$

$$=E^\pi\left[I[\sigma<\infty] \left(\prod_{0\leq j<l} I[A(\sigma+j)=0] \right) I[A(\sigma+l)\neq 0] \prod_{k=0}^{K} z_k^{A_k(\sigma+l)} \mid \mathbb{F}_\sigma \right]$$

$$=E^\pi\left[I[\sigma<\infty] \left(\prod_{0\leq j<l} I[A(\sigma+j)=0] \right) E^\pi\left[I[A(\sigma+l)\neq 0] \prod_{k=0}^{K} z_k^{A_k(\sigma+l)} \mid \mathbb{F}_{\sigma+(l-1)} \right] \mid \mathbb{F}_\sigma \right] \tag{4.24}$$

Under P^π, the RV $A(\sigma+l)$ is distributed according to the common distribution $q(\cdot)$ of the i.i.d sequence $\{A(n)\}_1^\infty$ and is clearly *independent* of the σ-field $\mathbb{F}_{\sigma+(l-1)}$. These facts are readily established from the properties (P1)-(P4), which also imply

$$P^{\pi}[A(\sigma+n)=0] = q(0). \qquad n=0,1,\ldots \quad (4.25)$$

It is now clear that

$$E^{\pi}\left[I[\sigma<\infty] I[A(\sigma+l)\neq 0] \prod_{k=0}^{K} z_k^{A_k(\sigma+l)} \mid \mathbb{F}_{\sigma+(l-1)} \right] = I[\sigma<\infty][a(z) - q(0)] \quad (4.26)$$

Substitution of (4.26) into (4.24) easily implies that

$$\begin{aligned} &E^{\pi}\left[I[\sigma<\infty, \beta(\sigma)=l] \prod_{k=0}^{K} z_k^{A_k(\gamma(\sigma))} \mid \mathbb{F}_{\sigma} \right] \\ &= I[\sigma<\infty]\left(\prod_{0\leq j<l} P^{\pi}[A(\sigma+j)=0] \right) [a(z) - q(0)] \end{aligned} \quad (4.27)$$

by making use of the $\mathbb{F}_{\sigma}$-measurability of the event $[\sigma<\infty]$ and by noting that the RV's $\{A(\sigma+k), 0\leq k<l\}$ are jointly *independent* (under P^{π}) from the σ-field $\mathbb{F}_{\sigma}$ as a consequence of (P1)-(P4). The relation

$$E^{\pi}[I[\sigma<\infty, \beta(\sigma)=l] \prod_{k=0}^{K} z_k^{A_k(\gamma(\sigma))} \mid \mathbb{F}_{\sigma}] = I[\sigma<\infty] q(0)^{l} [a(z) - q(0)] \quad (4.28)$$

obtains by direct inspection upon substituting (4.25) into (4.27). The reader will readily check by arguments using earlier remarks that (4.28) also holds for $l=0$, whence the conclusions (4.21) and (4.22) are now immediate by elementary calculations. □

5. PROPERTIES OF BUSY CYCLES:

This section is devoted to the study of the busy cycles when the system is operated under a fixed non-idling policy π. To that end, consider the following collections of $\mathbb{N}$-valued RV's, namely $\{\tau_n\}_1^{\infty}$, $\{\overline{\tau}_n\}_1^{\infty}$ and $\{\nu_n\}_1^{\infty}$, whose definitions and interpretations are now presented. First, pose

$$\tau_1 := \nu(1) I[\Xi\neq 0] + I[\Xi=0] \quad (5.1)$$

and recursively define

$$\overline{\tau}_n := \gamma(\tau_n) - 1 = \tau_n + \beta(\tau_n) + 1 \qquad n=1,2,\ldots \quad (5.2a)$$

$$\nu_n := \nu(\overline{\tau}_n) \qquad n=1,2,\ldots \quad (5.2b)$$

and

$$\tau_{n+1} := \tau(\overline{\tau}_n) = \overline{\tau}_n + \nu_n \qquad n=1,2,\ldots \quad (5.2c)$$

where the definitions (4.17) and (4.20) are used. Note that τ_n is the n-th time epoch at which the queueing system empties, and $\overline{\tau}_n$ represents the first time after τ_n that the system is again *not* empty. Since the interval $[\tau_n, \tau_{n+1})$ can be viewed as the n-th *busy cycle*, it is natural to consider the $\mathbb{N}$-valued RV's $\{\theta_n\}_1^\infty$ defined by

$$\theta_{n+1} := \tau_{n+1} - \tau_n = 1+\beta(\tau_n)+\nu(\overline{\tau}_n) \qquad n=1,2,\ldots(5.3)$$

with $\theta_1 := \tau_1$. It is clear that $1+\beta(\tau_n)$ and $\nu(\overline{\tau}_n)$ are the lengths of the n-th *idle* period and n-th *busy* period, respectively, whereas θ_{n+1} is exactly the length of the n-th busy cycle.

The RV's $\{\tau_n\}_1^\infty$ and $\{\overline{\tau}_n\}_1^\infty$ are $\mathbb{F}_n$-stopping times with the property that for all $n=1,2,\ldots,$

$$X(\tau_n) = 0 \;\; on \;\; the \;\; event \;\; [\tau_n < \infty] \tag{5.4a}$$

and

$$X(\overline{\tau}_n) = A(\gamma(\tau_n)) \;\; on \;\; the \;\; event \;\; [\tau_n < \infty, \beta(\tau_n) < \infty]. \tag{5.4b}$$

Moreover, owing to Corollary 4.1.1 and to the remarks following Lemma 4.2, it is easy to see recursively that the RV's τ_n, $\beta(\tau_n)$, $\overline{\tau}_n$ and ν_n are all P^π-*a.s.* finite.

Theorem 5.1. *The RV τ_1 satisfies the relation*

$$E^\pi\left[\frac{1}{r(b)^{\tau_1}} \mid \mathbb{F}_1\right] = \prod_{k=0}^{K} z_k(b)^{\Xi_k} I[\Xi\neq 0] + \frac{1}{r(b)} I[\Xi=0] \tag{5.5}$$

for every $b \geq 1$, and its conditional moment is given by

$$E^\pi[\tau_1 \mid \mathbb{F}_1] = I[\Xi=0] + \frac{Z(\Xi)}{1-\rho} I[\Xi\neq 0]. \tag{5.6}$$

The expression (5.5) for the conditional probability generating function of the RV τ_1 clearly shows that the conditional distribution of the RV τ_1 given the σ-field $\mathbb{F}_1$ is *independent* of the *non*-idling policy π used, a property that is immediately transferred to the (unconditional) distribution of the RV τ_1.

Proof: As the very definition of τ_1 implies

$$E^\pi\left[\frac{1}{r(b)^{\tau_1}} \mid \mathbb{F}_1\right] = E^\pi\left[\frac{1}{r(b)^{\nu(1)}} \mid \mathbb{F}_1\right] I[\Xi\neq 0] + \left[\frac{1}{r(b)}\right] I[\Xi=0], \tag{5.7}$$

a direct application of Theorem 4.1 (with $\sigma\equiv 1$) then readily yields the first part of the result. The relation (5.5) can be rewritten in the equivalent form

$$E^\pi[\frac{1}{r(b)^{\tau_1}} - 1 \mid \mathbb{F}_1] = \left[\prod_{k=0}^{K} z_k(b)^{\Xi_k} - 1\right] I[\Xi\neq 0] + \left[\frac{1}{r(b)} - 1\right] I[\Xi=0] \tag{5.8}$$

and a well-known identity for geometric series thus implies

$$E^{\pi}\Big[\sum_{0\le k<\tau_1}\Big(\frac{1}{r(b)}\Big)^k \mid \mathbb{F}_1\Big] = I[\Xi=0] + \left[\frac{\prod_{k=0}^{K} z_k(b)^{\Xi_k}-1}{r(b)^{-1}-1}\right] I[\Xi\neq 0]. \tag{5.9}$$

Elementary calculations now show that

$$\left[\frac{\prod_{k=0}^{K} z_k(b)^{\Xi_k}-1}{r(b)^{-1}-1}\right]$$

$$= \frac{\exp[\sum_{k=0}^{K}\Xi_k \ln z_k(b)]-1}{b-1} \times r(b)\left(\frac{1-r(b)}{b-1}\right)^{-1} \tag{5.10}$$

with

$$\lim_{b\downarrow 1}\frac{1-r(b)}{b-1} = -1-\lim_{b\downarrow 1} b\frac{a(z(b))-1}{b-1} = -(1-\rho) \tag{5.11}$$

and $\lim_{b\downarrow 1} r(b) = 1$ by virtue of (4.5a),(4.9) and (4.12), and

$$\lim_{b\downarrow 1}\frac{\exp[\sum_{k=0}^{K}\Xi_k \ln z_k(b)]-1}{b-1} = -Z(\Xi). \tag{5.12}$$

Finally, note that

$$\lim_{b\downarrow 1} E^{\pi}\Big[\sum_{0\le k<\tau_1}\Big(\frac{1}{r(b)}\Big)^k \mid \mathbb{F}_1\Big] = E^{\pi}[\tau_1 \mid \mathbb{F}_1] \tag{5.13}$$

by the Monotone Convergence Theorem for conditional expectations, whence the moment result (5.6) follows upon letting b go to 1 monotonically in (5.10) and making use of the limits (5.11)-(5.12). □

Theorem 5.2. *Under the foregoing assumptions (R1)-(R4), the relations*

$$E^{\pi}\left[\frac{1}{r^{\beta(\tau_n)}}\times\frac{1}{r(b)^{\nu(\overline{\tau}_n)}} \mid \mathbb{F}_{\tau_n}\right] = \frac{r[a(z(b))-q(0)]}{r-q(0)} \qquad n=1,2,\ldots \tag{5.14}$$

hold true for all r and b in the interval $[1,+\infty)$. In particular, the RV's $\beta(\tau_n)$ and $\nu(\overline{\tau}_n)$ are conditionally independent of the σ-field $\mathbb{F}_{\tau_n}$.

Proof: A direct application of Theorem 4.1 (with $\sigma=\overline{\tau}_n$) readily implies that

$$E^{\pi}\left[I[\overline{\tau}_n<\infty,\tau_{n-1}<\infty]\frac{1}{r(b)^{\nu(\overline{\tau}_n)}} \mid \mathbb{F}_{\overline{\tau}_n}\right] = I[\overline{\tau}_n<\infty]\prod_{k=0}^{K} z_k(b)^{A_k(\gamma(\tau_n))} \tag{5.15}$$

where use has been made of the property (5.4b). The σ-field inclusion $\mathbb{F}_{\tau_n}\subset\mathbb{F}_{\overline{\tau}_n}$ and an elementary smoothing argument using (5.15) now imply that

$$E^{\pi}\left[I\,[\overline{\tau}_n <\infty, \tau_{n+1}<\infty] \frac{1}{r^{\beta(\tau_n)}} \times \frac{1}{r(b)^{\nu(\overline{\tau}_n)}} \mid \mathbb{F}_{\tau_n} \right]$$

$$= E^{\pi}\left[I\,[\overline{\tau}_n <\infty] \frac{1}{r^{\beta(\tau_n)}} \prod_{k=0}^{K} z_k(b)^{A_k(\gamma(\tau_n))} \mid \mathbb{F}_{\tau_n} \right]. \qquad (5.16)$$

The relation (5.14) follows immediately from (5.16) upon applying Lemma 4.2 (with $\sigma=\tau_n$) and making use of the P^{π}-a.s finiteness of the RV's $\overline{\tau}_n$ and τ_{n+1}. □

The statistical structure of the busy cycles can now be easily obtained.

Theorem 5.3. *Under the foregoing assumptions (R1)-(R4), the RV's $\{\theta_n\}_1^{\infty}$ form a (possibly delayed) renewal sequence. More precisely, the RV's $\{\theta_n\}_2^{\infty}$ are i.i.d RV's mutually independent of the σ-field $\mathbb{F}_{\tau_1}$, with common distribution independent of the non-idling policy π and characterized by*

$$E^{\pi}\left[\frac{1}{r(b)^{\theta_{n+1}}} \mid \mathbb{F}_{\tau_n} \right] = \frac{a(z(b))-q(0)}{r(b)-q(0)} \qquad n=2,3,\ldots(5.17)$$

for all $b \geq 1$.

Proof: The result follows immediately from (5.14) (with $r=r(b)$) upon observing that the RV's $\{\theta_k,\ 1\leq k\leq n\}$ are clearly $\mathbb{F}_{\tau_n}$-measurable. □

The *independence* of the common distribution of the RV's $\{\theta_{n+1}\}_1^{\infty}$ on the non-idling policy π was already obtained through sample path arguments in [1] , where the discussion was carried out in the context of a different, though probabilistically equivalent, model of competing queues. Here, however, sharper statistical information on the length of busy cycles has been derived: Indeed, with $a_{\Xi}(\cdot)$ denoting the probability generating function of the RV Ξ as in (4.5a), the reader will readily conclude from (5.5) and (5.17) that the relations

$$E^{\pi}\left[\frac{1}{r(b)^{\theta_1}} \right] = a_{\Xi}(z(b)) + \frac{1-r(b)}{r(b)}\, q_{\Xi}(0) \qquad (5.18a)$$

and

$$E^{\pi}\left[\frac{1}{r(b)^{\theta_n}} \right] = \frac{a(z(b))-q(0)}{r(b)-q(0)} \qquad n=2,3,\ldots(5.18b)$$

hold for all $b \geq 1$, thus providing *explicit* expressions for the *probability generating functions* of the RV's $\{\theta_n\}_1^{\infty}$, taken in the variable $1/r(b)$ (in (0,1] whenever $b \geq 1$ by Lemma A.1). Consequently, information on the *moments* of the RV's $\{\theta_n\}_1^{\infty}$ is now readily available upon computing the successive right derivatives at $b=1$ of both sides in (5.18).

The interested reader will easily check that the mappings $b \to a(z(b))$, $b \to a_B(z(b))$ and $b \to r(b)$ are all *analytic* on the interval $(1, +\infty)$; moreover, existence and computation of the k-th derivative for each of the functions $b \to 1/r(b)^n$, $n=1,2,\ldots$, require existence and computation of all the derivatives of order l with $0 \leq l \leq k$ of the function $b \to a(z(b))$. With this information in mind, it is easy to extract the following result from (5.18).

Theorem 5.4. *Under the assumptions (R1)-(R4), if (R4) holds with γ integer, then the moment result*

$$E^{\pi}\left[\,|\,\theta_n\,|^{\gamma}\,\right] < \infty \qquad n=1,2,\ldots(5.19)$$

holds true, and in particular,

$$E^{\pi}\left[\,|\,\theta_n\,|\,\right] < \infty. \qquad n=1,2,\ldots(5.20)$$

In other words, the finiteness of the moments of the RV's $\{\theta_n\}_1^{\infty}$ is exactly of the same order as the one for the initial condition and the arrival process.

6. MOMENT ESTIMATES AND A REPRESENTATION OF THE COST:

The results on the length of busy cycles, obtained in the previous section, are now used to derive bounds on various moments of the sequences of RV's $\{\,|\,X(n)\,|\,\}_1^{\infty}$ and $\{\,|\,c(X(n))\,|\,\}_1^{\infty}$ under any non-idling policy π in Π. The key fact is contained in the observation that the *total* number of customers in the system at any given time n decreases by *at most one* unit in the next time slot $[n, n+1)$, and is therefore bounded above by the number of slots it takes for the queue sizes to empty for the first time after n. To formalize this idea, consider the (continuous-time) counting process $\{N(t),\, t \geq 0\}$ naturally associated with either sequence $\{\tau_n\}_0^{\infty}$ or $\{\theta_n\}_0^{\infty}$ (under the convention $\tau_0=\theta_0=0$), i.e., for all $t \geq 0$,

$$N(t) := \max\{k \geq 0:\ \tau_k \leq t\} \qquad (6.1)$$

with the ready interpretation that $N(t)$ represents the number of times the queue has returned to the empty state by time t. With this notation, the observation made earlier now translates into

$$|\,X(n)\,| \leq \tau_{N(n)+1} - n. \qquad n=1,2,\ldots(6.2)$$

By making use of this fact, it will be possible to show the following strong estimates.

Theorem 6.1. *Under the foregoing assumptions (R1)-(R4), there exists a single positive constant C such that for every non-idling policy π in Π, the moment estimate*

$$\sup_n E^{\pi}\left[\,|\,X(n)\,|^{\gamma-1}\right] \leq C < \infty \tag{6.3}$$

holds true.

Theorem 6.1, whose proof is presented below, turns out to be a special case of an intermediate result of independent interest which is discussed in Theorem 6.2. The reader will readily observe from Theorem 6.1 that whenever $\gamma>2$, the RV's $\{X(n)\}_1^\infty$ form a *uniformly integrable* sequence under P^π, and this uniformly in the policy π. Another simple consequence of Theorem 6.1 is presented in the following corollary.

Corollary 6.1.1. *Under the assumptions (R1)-(R5), whenever γ is such that*

$$1+\delta(1+\epsilon) \leq \gamma \tag{6.4}$$

for some $\epsilon>0$, the RV's $\{c(X(n))\}_1^\infty$ are uniformly integrable under the probability measure P^π associated with any non-idling policy π.

As will be clear from the proof given below, this uniform integrability is also uniform over the class of non-idling policies.

Proof: Assumption (R5) and (6.4) immediately imply that

$$|\,c(X(n))\,|^{1+\epsilon} \leq L^{1+\epsilon}2^{\epsilon}\left[1+|\,X(n)\,|^{\gamma-1}\right] \qquad n=0,1,... \tag{6.5}$$

and the result follows by a direct application of Theorem 6.1. □

As shown in Section 5, the process $\{\theta_n\}_1^\infty$ is a delayed renewal process under any non-idling policy π in Π, with statistics *independent* of the policy π. For reference, denote by $G(\cdot)$ the distribution of the RV $\theta_1(=\tau_1)$ and by $F(\cdot)$ the common distribution of the i.i.d RV's $\{\theta_n\}_2^\infty$. In general, the distributions $G(\cdot)$ and $F(\cdot)$ do not coincide. Now, for any *monotone non-decreasing* mapping $r : \mathbb{R}^+ \to \mathbb{R}^+$, define the $\mathbb{R}$-valued process $\{R(t),\ t\geq 0\}$ by

$$R(t) := r(\tau_{N(t)+1} - t) \tag{6.6}$$

for all $t\geq 0$, with corresponding expected value

$$M_G(t) := E_{\Xi}^{\pi}\left[r(\tau_{N(t)+1} - t)\right] (= E^{\pi}[R(t)]). \tag{6.7}$$

The subscripts G and Ξ in (6.7) emphasizes the fact that the system is started with an initial queue size Ξ distributed according to the distribution $q_{\Xi}(\cdot)$. If $G(\cdot)=F(\cdot)$, the sequence $\{\theta_n\}_1^\infty$ is a non-delayed *renewal* sequence and it is appropriate to pose

$$M_F(t) := E_0^{\pi}\left[r(\tau_{N(t)+1} - t)\right] \tag{6.8}$$

This corresponds to an appropriate choice of the initial condition Ξ.

The first part of this section is devoted to the derivation of a bound on the expected values $\{M_G(t),\ t \geq 0\}$ for any non-idling policy π, with a view towards generating (via (6.2)) a bound for the sequence of expected values $\{E^\pi[r(\,|X(n)|\,)]\}_1^\infty$.

Theorem 6.2. *Let π be an arbitrary non-idling policy in Π. Under the finite moment assumptions*

$$\int_0^\infty r(\theta)dG(\theta) < \infty \ ,\ \int_0^\infty r(\theta)dF(\theta) < \infty. \tag{6.9}$$

the condition

$$\int_0^\infty\int_0^\theta r(\theta-t)\ dt\ \ dF(\theta) = \int_0^\infty\int_t^\infty r(\theta-t)\ dF(\theta)\ dt\ < \infty \tag{6.10}$$

implies

$$\sup_{t\geq 0} M_G(t) = \sup_{t\geq 0} E^\pi[R(t)] < \infty. \tag{6.11}$$

Proof: Define the mappings $a_G(\cdot)$, $a_F(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ by

$$a_G(t) := \int_t^\infty r(\theta-t)\ dG(\theta),\ \ a_F(t) := \int_t^\infty r(\theta-t)\ dF(\theta) \tag{6.12}$$

for all $t \geq 0$. Since $r(\cdot)$ takes positive values and is monotone non-decreasing, the indefinite integrals entering the definition (6.12) are well defined and satisfy the obvious inequalities

$$0 \leq \int_t^\infty r(\theta-t)\ dG(\theta) \leq \int_s^\infty r(\theta-t)\ dG(\theta) \leq \int_s^\infty r(\theta-s)\ dG(\theta) \tag{6.13}$$

whenever $0 \leq s \leq t$, with a similar chain of inequalities when $G(\cdot)$ replaced by $F(\cdot)$. It is now clear from (6.9) and (6.13) that $0 \leq a_G(t) \leq a_G(0) < \infty$ for all $t \geq 0$, whence the mapping $a_G(\cdot)$ is well defined and monotone non-increasing. Similar comments are of course valid for $a_F(\cdot)$.

A standard renewal argument ([10] , pp. 183) applied to the process $\{R(t),\ t \geq 0\}$ shows that for all $t \geq 0$,

$$M_G(t) = \int_0^t M_F(t-\theta)\ dG(\theta)\ +\ \int_t^\infty r(\theta-t)\ dG(\theta) \tag{6.14}$$

and the remarks made earlier thus imply that

$$M_G(t) \leq \int_0^t M_F(t-\theta)\ dG(\theta)\ +\ \int_0^\infty r(\theta)\ dG(\theta) \tag{6.15}$$

$$\leq \sup_{0 \leq s \leq t} M_F(s) + \int_0^\infty r(\theta)\, dG(\theta). \tag{6.16}$$

This clearly shows that under the assumed condition (6.9), the result (6.11) will hold if it can be established that

$$\sup_{t \geq 0} M_F(t) < \infty. \tag{6.17}$$

With the notation (6.12), the renewal equation (6.14), this time with $G(\cdot)=F(\cdot)$, specializes to

$$M_F(t) = a_F(t) + \int_0^t M_F(t-\theta)\, dF(\theta) \tag{6.18}$$

for all $t \geq 0$. Since, as pointed out earlier, the mapping $a_F(\cdot)$ is *monotone non-increasing* and takes *non-negative* values, it is consequently *integrable* owing to (6.10), and therefore *directly Riemann integrable* ([10] , pp. 190-191). The distribution $F(\cdot)$ has support on $\mathbb{N}$ and is thus arithmetic, say with span d. These remarks validate an application of the Basic Renewal Theorem ([10] , Thm. 5.5.1., p. 191) on the renewal equation (6.18), in that the convergence

$$\lim_{n \to \infty} M_F(c+nd) = \frac{d}{\overline{\theta}} \sum_{n=0}^{\infty} a_F(c+nd) \tag{6.19}$$

takes place. Here, $\overline{\theta}$ denotes the mean of the distribution $F(\cdot)$ which is *finite* as noted at the end of Section 5.

The mapping $M_F(\cdot)$ is non-increasing on each one of the intervals $[n, n+1)$, whereas the mapping $a_F(\cdot)$ is non-increasing. These properties and (6.19) readily imply that

$$\overline{\lim_{t \uparrow \infty}}\, M_F(t) \leq \frac{d}{\overline{\theta}} \sum_{n=0}^{\infty} a_F(nd) \leq \frac{d}{\overline{\theta}} \sum_{n=0}^{\infty} a_F(n)$$

$$\leq \frac{d}{\overline{\theta}} \left[a_F(0) + \int_0^\infty \int_t^\infty r(\theta - t)\, dF(\theta)\, dt \right] < \infty \tag{6.20}$$

where the finiteness of the bound in (6.20) follows from the assumptions (6.9)-(6.10). Since (6.9)-(6.10) imply that for *each* t, $M_F(t)$ is finite, the bound (6.17) obtains and this completes the proof as pointed out earlier. □

Proof of Theorem 6.1: Start with the mapping $r(\cdot)$ given by $r(x) = x^{\gamma-1}$ for all $x \geq 0$, and note from elementary calculations that

$$\int_0^\infty \int_0^\theta r(\theta - t)\, dt\ \ dF(\theta) = \frac{1}{\gamma} \int_0^\infty \theta^\gamma\, dF(\theta). \tag{6.21}$$

It follows from Theorem 5.4 and (6.21) that the conditions (6.9)-(6.10) hold true, and a straightforward application of Theorem 6.2 thus implies (6.3). □

It is noteworthy that if the much stronger hypothesis (R4bis) is substituted to (R4), where

(R4bis): For some constants $\lambda > 0$ and $0 < D < \infty$, the bounds

$$E^\pi[e^{\lambda |B|}] < D \quad , \quad E^\pi[e^{\lambda |A(n)|}] < D$$

hold true.

then the results of Section 5 and Theorem 6.2 (with $r(x) = e^{\lambda x}$) imply the bound $E^\pi e^{\lambda |X(n)|} \leq D' < \infty$ for all $n = 1,2,...$ This can be also proved directly by an easy application of the results of Hajek [8] without referring to the results of Section 5.

This section closes with a useful representation result for the cost under any non-idling Markov stationary policy. For any such a policy g, the sequence $\{X(n)\}_1^\infty$ is a *homogeneous* Markov chain over the state-space $\mathbb{N}^{K-1}$ under the probability measure P^g induced by the policy g. All states communicate with each other under the assumed stability condition $\rho < 1$ since $0 < \mu_k \leq 1$, whence the chain is *irreducible*. From this property, it is clear that the chain is also *aperiodic* for the empty state is aperiodic. Moreover, the condition $\gamma \geq 1$ in the assumption (R4) gives the finite mean property (5.20) which 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 ([10] , Thm. 3.1.3., pp. 85) that the Markov chain $\{X(n)\}_1^\infty$ admits under P^g a *unique invariant* measure, which is denoted throughout by $\mathbb{P}^g$ with corresponding expectation operator $\mathbb{E}^g$.

Theorem 6.3. *Under the foregoing assumptions (R1)-(R4), if the RV's $\{c(X(n))\}_1^\infty$ are uniformly integrable under the probability measure P^g associated with the non-idling Markov stationary policy g, then the following convergence results hold:*

(i): With X denoting a generic $\mathbb{N}^{K-1}$-valued RV,

$$\lim_{n \to \infty} E^g c(X(n)) = \mathbb{E}^g c(X) \tag{6.22a}$$

independently of the initial state distribution, and

(ii):

$$J(g) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} c(X(i)) \tag{6.22b}$$

where the convergence takes place P^g-a.s. and in $L^1(\Omega, \mathbb{F}, P^g)$.

It should be clear to the reader that (6.22a)-(6.22b) implies

$$J(g) = \lim_{n \to \infty} \frac{1}{n} E^g \sum_{i=1}^{n} c(X(i)) = \mathbb{E}^g c(X) \tag{6.23}$$

whenever the RV's $\{c(X(n))\}_1^\infty$ are uniformly integrable under the probability measure

P^g.

Proof: (i): For each $B > 0$, introduce the so-called truncation of c at level B as the bounded mapping $c_B : \mathbb{N}^{K+1} \to \mathbb{R}$ defined by $c_B(x) := (c(x) \vee -B) \wedge B$ for all x in $\mathbb{N}^{K+1}$. With this notation, the elementary inequality

$$\begin{aligned} | E^g [c(X(n))] - \mathbb{E}^g [c(X)] | \ \leq \ & | E^g [c(X(n))] - E^g [c_B(X(n))] | \\ & + | E^g [c_B(X(n))] - \mathbb{E}^g [c_B(X)] | \\ & + | \mathbb{E}^g [c_B(X)] - \mathbb{E}^g [c(X)] | \qquad n=1,2,\ldots \end{aligned} \tag{6.24}$$

is clearly valid valid for every $B > 0$. The argument for establishing (6.22a) thus reduces to showing that each one of the difference terms can be made arbitrarily small as n goes to infinity:

Since the chain is ergodic under P^g, it is well known ([10] , Thm. 3.1.3, pp. 85), that $\lim_{n \to \infty} P^g [X(n) = x] = \mathbb{P}^g [X = x]$ for all x in $\mathbb{N}^{K+1}$, and the convergence

$$\lim_{n \to \infty} E^g \, b(X(n)) = \mathbb{E}^g \, b(X) \tag{6.25}$$

thus takes place for any *bounded* mapping $b := \mathbb{N}^{K+1} \to \mathbb{R}$. The assumed uniform integrability now implies that for *all* $B > 0$,

$$\sup_n E^g [| c_B(X(n)) |] \leq \sup_n E^g [| c(X(n)) |] < \infty \tag{6.26}$$

whence

$$\mathbb{E}^g \ | c_B(X) | \ = \lim_{n \uparrow \infty} E^g [| c_B(X(n)) |] \leq \sup_n E^g [| c(X(n)) |] < \infty \tag{6.27}$$

with the equality following from (6.25). The Monotone Convergence Theorem, used on the positive and negative parts of $c_B(X)$, readily yields the conclusion that

$$\lim_{B \uparrow \infty} \mathbb{E}^g \ | c_B(X) | \ = \mathbb{E}^g \ | c(X) | \ \leq \ \sup_n E^g [| c(X(n)) |] < \infty, \tag{6.28a}$$

and

$$\lim_{B \uparrow \infty} \mathbb{E}^g \, c_B(X) = \mathbb{E}^g \, c(X), \tag{6.28b}$$

Consequently, $\lim_{B \to \infty} | \mathbb{E}^g c(X) - \mathbb{E}^g c_B(X) | = 0$, i.e., for every $\epsilon > 0$, there exists $B_\epsilon > 0$ such that for $B > B_\epsilon$,

$$\left| \mathbb{E}^g \, c(X) - \mathbb{E}^g \, c_B(X) \right| \leq \epsilon. \tag{6.29}$$

The convergence (6.25) can be expressed by saying that for every $B > 0$ and every $\epsilon > 0$, there exists $n(\epsilon, B)$ in $\mathbb{N}$ such that whenever $n > n(\epsilon, B)$

$$| E^g [c_B(X(n))] - \mathbb{E}^g [c_B(X)] | < \epsilon \tag{6.30}$$

Finally, the very definition of c_B implies

$$| E^g [c(X(n))] - E^g [c_B(X(n))] | \leq E^g \left[I[\, | c(X(n)) | > B \,] \, | c(X(n)) | \right] \quad n=1,2,... \tag{6.31}$$

and by the uniform integrability of the RV's $\{c(X(n))\}_1^\infty$, for every $\epsilon>0$, there exists $B^\epsilon>0$ such that

$$\begin{aligned} \sup_n \; | E^g [c(X(n))] - E^g [c_B(X(n))] | &\leq \\ \leq \sup_n E^g \left[I[\, | c(X(n)) | > B \,] \, | c(X(n)) | \right] &\leq \epsilon \end{aligned} \tag{6.32}$$

whenever $B > B^\epsilon$.

The reader will readily check from (6.29),(6.30) and (6.32) that

$$| E^g [c(X(n))] - \mathbb{E}^g [c(X)] | \leq 3\epsilon \tag{6.33}$$

whenever $n > n(\epsilon, B_\epsilon \vee B^\epsilon)$ and this complete the proof of (6.22a) since ϵ is arbitrary.

(ii): For ease of exposition, define the $\mathbb{R}$-valued RV's $\{Y^c(n)\}_1^\infty$ by

$$Y^c(n) := \frac{1}{n} \sum_{i=1}^{n} c(X(i)). \quad n=1,2,... \tag{6.34}$$

The uniform integrability of the RV's $\{c(X(n))\}_1^\infty$ carries over to the RV's $\{Y^c(n)\}_1^\infty$ and consequently, it suffices to establish in (6.22b) the P^g-a.s. convergence of the sequence of RV's $\{Y^c(n)\}_1^\infty$ ([6] , Thm. 4.5.4, pp. 97-98). As discussed by Chung ([5] , Thm. I.15.2, pp. 92), this convergence takes place with

$$\lim_{n \to \infty} Y^c(n) = \mathbb{E}^g \, c(X), \tag{6.35}$$

provided the condition $| \mathbb{E}^g \, c(X) | < \infty$ holds. This absolute summability condition was obtained earlier in the proof as (6.28a) and this completes the proof of part (ii). □

7. AN EXTENSION OF MANDL'S RESULT:

Throughout this section, let g denote a fixed *non-idling* Markov stationary policy in Π and let α be a second policy in Π (which is not necessarily *non-idling*). The results will be given in as generic a form as possible to emphasize the broad applicability of the methodology.

Theorem 7.1. *Under the foregoing assumptions (R1)-(R3), assume the assumptions (H1)-(H5) to be enforced. Whenever the policy α satisfies the convergence condition (C) with respect to the non-idling policy g, the convergence*

$$J(\alpha) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} c(X(i)) = J(g) = \mathbb{E}^g c(X) \tag{7.1}$$

takes place in $L^1(\Omega, \mathbb{F}, P^\alpha)$.

In order to prove Theorem 7.1, it is necessary to extend an argument given by Mandl [13] to the case of *unbounded* costs over *countable* state-spaces and *randomized* strategies. The forthcoming discussion leads via several lemmas of independent interest to a proof of the main Theorem 7.1. The first lemma provides a well-known characterization of the long-run average cost $J(g)$, which is valid for an arbitrary (not necessarily non-idling) Markov stationary policy.

Lemma 7.1. *If the mapping* $h : \mathbb{N}^{K+1} \to \mathbb{R}$ *and the constant* J *solve the equations*

$$h(x) + J = \sum_y p(x, y; g(x)) h(y) + c(x) \tag{7.2}$$

for all x in $\mathbb{N}^{K+1}$ *under the conditions*

$$E^g \mid h(X(n)) \mid < \infty \qquad n=1,2,... \tag{7.3a}$$

and

$$\lim_{n \to \infty} \frac{1}{n} E^g [h(X(n))] = 0, \tag{7.3b}$$

then necessarily

$$J = J(g) = \lim_{n \to \infty} \frac{1}{n} E^g \sum_{i=1}^{n} c(X(i)). \tag{7.4}$$

Proof: The formula (2.3) for the transition probabilities allows a rewriting of (7.2) in the form

$$h((X(i)) + J = E^g [h(X(i+1)) \mid \mathbb{F}_i] + c(X(i)). \qquad i=1,2,... \tag{7.5}$$

and a direct iteration then gives

$$E^g [h(\Xi)] + nJ = E^g [h(X(n+1))] - E^g [\sum_{i=1}^{n} c(X(i))]. \qquad n=1,2,... \tag{7.6}$$

The result now follows readily upon dividing by n in (7.6) and letting n go to infinity. □

When the cost function c is *bounded*, at least one solution pair (h, J) can be shown to exist which satisfies the conditions of Lemma 7.1. This existence result is established by a standard argument available in the monographs by Ross [18, 19] : Let the expected discounted cost function associated with the one-step cost c be denoted by C_β whenever the non-idling policy g is used over the infinite horizon and the discount factor is $\beta < 1$, i. e., for all x in $\mathbb{N}^{K+1}$, pose

$$C_\beta(x) := E^g\left[\sum_{i=1}^{\infty} \beta^i c(X(i)) \mid X(1)=x\right]. \tag{7.7}$$

In that case, it is plain that for all x in $\mathbb{N}^{K+1}$, the relation

$$C_\beta(x) = c(x)+\beta\sum_y p(x,y;g(x))C_\beta(y) \tag{7.8}$$

holds, or equivalently,

$$(1-\beta)C_\beta(0)+h_\beta(x) = c(x)+\beta\sum_y p(x,y;g(x))h_\beta(y). \tag{7.9}$$

where the definition

$$h_\beta(x) := C_\beta(x)-C_\beta(0) \tag{7.10}$$

has been posed for all x in $\mathbb{N}^{K+1}$.

This last remark is useful for establishing an intermediary result given in the next proposition. The discussion is presented under condition (H1bis), where

(H1bis): The uniform bound

$$\sup_n E^g\left[Z(X(n))\right] < \infty$$

holds. It should be clear to the reader that (H1bis) is a weaker condition than (H1).

Lemma 7.2. *Assume the mapping c to be bounded and the non-idling policy g to satisfy the condition (H1bis). Under the assumptions (R1)-(R3), with the notation and definitions given above,*

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

$$|h_\beta(x)| \leq CZ(x) \tag{7.11}$$

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

(2): *The convergence*

$$\lim_{\beta\uparrow 1}(1-\beta)C_\beta(0) = J(g) = \mathbb{E}^g c(X) \tag{7.12}$$

takes place, and

(3): *There exists a pair (h,J) which satisfies (7.2)-(7.3), given by $J=J(g)$ and*

$$h(x) := \lim_{\beta\uparrow 1} h_\beta(x). \tag{7.13a}$$

the limit being taken along a single subsequence, with the property

$$|h(x)| \leq CZ(x) \tag{7.13b}$$

for all x in $\mathbb{N}^{K+1}$.

Proof: (1): Fix $x \neq 0$ in $\mathbb{N}^{K+1}$. With the notation (5.1), standard arguments yield

$$C_\beta(x)-C_\beta(0)=E^g\left[\sum_{i=1}^{\tau_1-1}\beta^i c(X(i)) \mid X(1)=x\right]-C_\beta(0)E^g\left[1-\beta^{\tau_1} \mid X(1)=x\right]. \tag{7.14}$$

It is clear that

$$\left| E^g\left[\sum_{i=1}^{\tau_1-1}\beta^i c(X(i)) \mid X(1)=x\right]\right| \leq |c| E^g[\tau_1 \mid X(1)=x] \tag{7.15}$$

whereas the easy bound

$$|C_\beta(0)| \leq \frac{|c|}{1-\beta} \tag{7.16}$$

immediately implies

$$\left|C_\beta(0)E^g\left[1-\beta^{\tau_1} \mid X(1)=x\right]\right| \leq |c| E^g\left[\frac{1-\beta^{\tau_1}}{1-\beta} \mid X(1)=x\right]$$

$$\leq |c| E^g[\tau_1 \mid X(1)=x] \tag{7.17}$$

by standard properties of geometric series. Use of (7.15) and (7.17) on (7.14) readily leads to

$$|h_\beta(x)| \leq 2|c| E^g[\tau_1 \mid X(1)=x] \leq \frac{2|c|}{1-\rho}Z(x). \tag{7.18}$$

where the last inequality follows from (5.6), and (7.10) obtains with $C := \frac{2|c|}{1-\rho}$.

(2): Since c is bounded, Theorem 6.3 applies and the result (7.12) readily follows from (6.23), the definition of $J(g)$ and the version of the Tauberian Theorem stated in Prop. 4-7 of ([9] , pp. 173).

(3): Owing to (7.11), the mapping $\beta \to h_\beta(x)$ is bounded for all x in $\mathbb{N}^{K+1}$. 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 $h(x)$ for *all* x in $\mathbb{N}^{K+1}$. The mapping h clearly satisfies (7.13b) and therefore enjoys the properties (7.3) under the uniform bound (H1bis). A simple bounding argument, that uses the form of the dynamics (2.2) and the enforced assumptions (R1)-(R3), easily implies that

$$0 \leq \sum_y p(x,y;g(x))Z(y) \leq Z(x)+\rho < \infty \tag{7.19}$$

for all x in $\mathbb{N}^{K+1}$. Dominated convergence, coupled to (7.11), (7.13b) and (7.19), now gives

$$\lim_{n\to\infty}\beta_n \sum_y p(x,y;g(x))h_{\beta_n}(y) = \sum_y p(x,y;g(x))h(y) \tag{7.20}$$

for all x in $\mathbb{N}^{K+1}$. Upon taking the limit in (7.9), these remarks and (7.20) readily imply that the pair $(h, J(g))$ indeed solves (7.2)-(7.3). □

Throughout this section, let h denote the mapping $\mathbb{N}^{K+1} \to \mathbb{R}$ whose existence was established in Lemma 7.2 when c is a bounded mapping. The sequences $\{\Phi(n)\}_1^\infty$ and $\{Y(n)\}_1^\infty$ of $\mathbb{R}$-valued RV's are defined to be

$$\Phi(n) = E^\alpha[h(X(n+1)) \mid \mathbb{F}_n] - E^g[h(X(n+1)) \mid \mathbb{F}_n] \qquad n=1,2,... (7.21)$$

and

$$Y(n) = h(X(n+1)) - E^\alpha[h(X(n+1)) \mid \mathbb{F}_n] \qquad n=1,2,... (7.22)$$

with $Y(1) = h(X(1)) - E^\alpha[h(X(1))]$. Note that these definitions are well posed under the assumptions (H1bis)-(H2) owing to the properties enjoyed by the mapping h.

Lemma 7.3. *Assume the mapping c to be bounded and the assumptions (H1)-(H2) to be enforced. Under the foregoing assumptions (R1)-(R3), whenever the policy α satisfies the convergence condition (C) with respect to the non-idling policy g, the convergence*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \Phi(i) = 0 \qquad (7.23)$$

takes place in $L^1(\Omega, \mathbb{F}, P^\alpha)$.

Proof: Let f_k be the Markov stationary policy in Π which always gives service attention to the k-th queue, i.e., for all x in $\mathbb{N}^{K+1}$,

$$f_k(l,x) = \delta(k,l), \; 0 \le l,k \le K. \qquad (7.24)$$

With this notation, it is plain from (2.3)-(2.5) that

$$\Phi(n) = \sum_{k=0}^K [\alpha_n(k, H(n)) - g(k, X(n))] E^{f_k}\left[h(X(n+1)) \mid \mathbb{F}_n\right] \qquad (7.25)$$

for all $n=1,2,...$ For each $0 \le k \le K$, observe from (7.13b) and the form of the one-step transition probabilities (2.3)-(2.5) that

$$\begin{aligned} \left| E^{f_k}\left[h(X(n+1)) \mid \mathbb{F}_n\right] \right| &\le CE^{f_k}\left[Z(X(n+1)) \mid \mathbb{F}_n\right] \\ &\le CE^{f_k}\left[Z(X(n)) + Z(A(n)) \mid \mathbb{F}_n\right] \\ &\le C\Big(Z(X(n) + \rho\Big) \qquad n=1,2,... (7.26) \end{aligned}$$

If the $\mathbb{R}$-valued RV's $\{\Delta(n)\}_1^\infty$ are defined by

$$\Delta(n) := \sum_{k=0}^K | \alpha_i(k, H(n)) - g(k, X(n)) |, \qquad n=1,2,... (7.27)$$

it is now immediate from (7.26) that

$$| \Phi(n) | \le C\Big(Z(X(n) + \rho\Big)\Delta(n) \qquad n=1,2,...(7.28)$$

and consequently,

$$\frac{1}{n}\sum_{i=1}^{n} | \Phi(i) | \le \frac{C}{n}\sum_{i=1}^{n} Z(X(i))\Delta(i) + \frac{C\rho}{n}\sum_{i=1}^{n}\Delta(i). \tag{7.29}$$

Under the condition (C), the RV's $\{\Delta(n)\}_1^\infty$ converge to 0 in probability under P^α. i. e., for every $\epsilon>0$ and $\delta>0$. there exists an integer $n(\epsilon,\delta)$ in $\mathbb{N}$ with the property that whenever $n > n(\epsilon,\delta)$ in $\mathbb{N}$,

$$P^\alpha[\Delta(n) > \epsilon] < \delta. \tag{7.30}$$

Cesaro convergence of the RV's $\{\Delta(n)\}_1^\infty$ to 0 also takes place in probability under P^α, and the RV's $\{\Delta(n)\}_1^\infty$ being uniformly bounded by $2K$, a well-known result on the convergence of uniformly integrable RV's ([6] , Thm. 4.5.4, pp. 97-98) gives the convergence

$$\lim_{n\uparrow\infty}\frac{C\rho}{n}\sum_{i=1}^{n}\Delta(i) = 0. \tag{7.31}$$

both in probability under P^α and in $L^1(\Omega,\mathbb{F},P^\alpha)$.

Moreover, the uniform integrability assumption (H2) is equivalent to the uniform bound

$$Z_\alpha := \sup_n E^\alpha[Z(X(n))] < \infty \tag{7.32}$$

and to the fact that for every $\eta>0$ there exists some $\delta(\eta)>0$ such that

$$\sup_n E^\alpha[Z(X(n))I(A)] < \eta \tag{7.33}$$

for any event A in $\mathbb{F}$ with $P^\alpha(A) < \delta(\eta)$.

Now fix $\epsilon>0$ and $\eta>0$. Upon combining (7.29) and (7.32), the reader will now readily check that for all $n > n(\epsilon.\delta(\eta))$ in $\mathbb{N}$, $P^\alpha[\Delta(n)>\epsilon]<\delta(\eta)$ and

$$E^\alpha\Big[Z(X(n))I[\Delta(n)>\epsilon]\Big] \le \eta, \tag{7.34}$$

whence

$$E^\alpha\Big[Z(X(n))\Delta(n)\Big] \le \epsilon E^\alpha\Big[Z(X(n))\Big] + \eta \le \epsilon Z_\alpha - \eta. \tag{7.35}$$

It is now straightforward to see that for $n > n(\epsilon,\delta(\eta))$ in $\mathbb{N}$,

$$E^\alpha\left[\frac{1}{n}\sum_{i=1}^{n} Z(X(i))\Delta(i)\right]$$

$$= E^\alpha\left[\frac{1}{n}\sum_{i=1}^{n(\epsilon,\delta(\eta))} Z(X(i))\Delta(i)\right] + E^\alpha\left[\frac{1}{n}\sum_{n(\epsilon,\delta(\eta))<i\le n} Z(X(i))\Delta(i)\right]$$

$$\leq 2KZ_\alpha \frac{n(\epsilon,\delta(\eta))}{n} + \left(\epsilon Z_\alpha + \eta\right)\frac{n - n(\epsilon,\delta(\eta))}{n}. \tag{7.36}$$

Let n go to infinity in (7.36) and observe that

$$\overline{\lim_{n\uparrow\infty}} E^\alpha\left[\frac{1}{n}\sum_{i=1}^{n} Z(X(i))\Delta(i)\right] \leq \epsilon Z_\alpha + \eta, \tag{7.37}$$

whence

$$\overline{\lim_{n\uparrow\infty}} E^\alpha\left[\frac{1}{n}\sum_{i=1}^{n} Z(X(i))\Delta(i)\right] = 0 \tag{7.38}$$

since both ϵ and η are arbitrary. The proof of Lemma 7.3. is now completed upon using (7.31) and (7.38) on (7.29). □

Lemma 7.4. *Assume the mapping c to be bounded and the assumptions (H2) and (H3) to be enforced. Under the foregoing assumptions (R1)-(R3), the convergence*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} Y(i) = 0 \tag{7.39}$$

takes place P^α–a.s. and in $L^1(\Omega,\mathbb{F},P^\alpha)$.

Proof: The sequence $\{Y(n)\}_1^\infty$ forms a $(P^\alpha,\mathbb{F}_n)$-*martingale difference* sequence. Under (H3), the estimate

$$E^\alpha\left(\sum_{n=1}^{\infty}\frac{Y^2(n)}{n^2}\right) < 4C^2 E^\alpha\left(\sum_{n=1}^{\infty}\frac{|Z(X(n))|^2}{n^2}\right) < \infty \tag{7.40}$$

readily follows from Jensen's inequality and (7.13b). A martingale version of the Law of Large Numbers ([13] , Theorem 3), the so-called Stability Theorem, thus applies to give the convergence (7.39) in the P^α–*a.s.* sense. Assumption (H2), when coupled to the estimate (7.13b), immediately implies the uniform integrability of the RV's $\{h(X(n))\}_1^\infty$ under the probability measure P^α, whence the convergence (7.39) also takes place in $L^1(\Omega,\mathbb{F}.P^\alpha)$ owing to standard results on the convergence of uniformly integrable RV's. □

Theorem 7.1. is now shown to hold under the more restrictive assumption that the mapping c is *bounded.*

Theorem 7.2. *Assume the mapping c to be bounded and the assumptions (H1)-(H3) to be enforced. Under the foregoing assumptions (R1)-(R3), whenever the policy α satisfies the condition (C) with respect to the non-idling policy g , the convergence*

$$J(\alpha) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} c(X(i)) = J(g) = \mathbb{E}^g c(X) \tag{7.41}$$

takes place in $L^1(\Omega, \mathbb{F}, P^\alpha)$.

Proof: The discussion follows closely the one given by Mandl ([13] , p. 46): In order to compare the relative effects of the policies g and α on the cost, (7.5) is rewritten in the equivalent form

$$h(X(i)) + J(g) = -\Phi(i) - Y(i) + h(X(i+1)) + c(X(i)) \qquad i=1,2\ldots(7.42)$$

by adding and substracting both RV's $E^\alpha[h(X(i+1)) \mid \mathbb{F}_i]$ and $h(X(i+1))$ on the right handside of (7.4). Iteration of (7.42) readily implies

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

$$= \frac{1}{n} \sum_{i=1}^{n} c(X(i)) - \frac{1}{n} \Big(h(X(n+1)) - h(X(1)) \Big) \qquad n=1,2,\ldots(7.43)$$

and note that

$$\lim_{n \to \infty} \frac{1}{n} E^\alpha[\, | h(X(n+1)) | \,] = \lim_{n \to \infty} \frac{1}{n} E^\alpha[\, | h(X(1)) | \,] = 0, \tag{7.44}$$

owing to either (H2) or (H3) when coupled to the estimate (7.13b). The result (7.41) is now easily obtained upon taking the limit in (7.43), with the help of (7.44), and applying Lemmas 7.2-7.4. □

Proof of Theorem 7.1: For each $B > 0$, let c_B be the mapping $\mathbb{N}^{K+1} \to \mathbb{R}$ used in the proof of Theorem 6.3. Under the assumed conditions, Theorem 7.2 implies the convergence

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} c_B(X(i)) = \mathbb{E}^g c_B(X) \tag{7.45}$$

in $L^1(\Omega, \mathbb{F}, P^\alpha)$, i.e., for every $\epsilon > 0$ and every $B > 0$, there exists $n(\epsilon, B)$ in $\mathbb{N}$ with the property that for all $n \geq n(\epsilon, B)$,

$$E^\alpha \left| \frac{1}{n} \sum_{i=1}^{n} c_B(X(i)) - \mathbb{E}^g c_B(X) \right| < \epsilon. \tag{7.46}$$

On the other hand, under assumption (H4), as pointed out in the proof of Theorem 6.3, $\mathbb{E}^g \mid c(X) \mid < \infty$ and for every $\epsilon > 0$, there exists $B_\epsilon > 0$ such that for $B \geq B_\epsilon$,

$$\left| \mathbb{E}^g c(X) - \mathbb{E}^g c_B(X) \right| \leq \epsilon. \tag{7.47}$$

It is also clear from the definition of the mapping c_B that

$$\left| \frac{1}{n} E^{\alpha} \sum_{i=1}^{n} c(X(i)) - c_B(X(i)) \right|$$

$$\leq \frac{1}{n} E^{\alpha} \sum_{i=1}^{n} \Big[\mid c(X(i)) \mid I[\mid c(X(i)) \mid > B] \Big], \tag{7.48}$$

whereas the uniform integrability condition (H5) guarantees that $\lim_{n \uparrow \infty} E^{\alpha} \mid c(X(n)) \mid I[\mid c(X(n)) \mid > B] := 0$ *uniformly* in B. This last fact carrying over to Cesaro convergence, it follows that for every $\epsilon>0$, there exists $m(\epsilon)$ in $\mathbb{N}$ such that whenever $m \geq m(\epsilon)$ in $\mathbb{N}$,

$$\left| \frac{1}{n} E^{\alpha} \sum_{i=1}^{n} c(X(i)) - c_B(X(i)) \right| \leq \epsilon \tag{7.49}$$

for *every* $B>0$.

To conclude the argument, observe that for every $n=1,2,...$ and every $B>0$,

$$E^{\alpha} \left| \frac{1}{n} \sum_{i=1}^{n} c(X(i)) - \mathbb{E}^{g} c(X) \right| \leq E^{\alpha} \left| \frac{1}{n} \sum_{i=1}^{n} c(X(i)) - c_B(X(i)) \right|$$

$$+ E^{\alpha} \left| \frac{1}{n} \sum_{i=1}^{n} c_B(X(i)) - \mathbb{E}^{g} c_B(X) \right|$$

$$+ \left| \mathbb{E}^{g} c(X) - \mathbb{E}^{g} c_B(X) \right|. \tag{7.50}$$

Now fix $\epsilon>0$ and pose $n(\epsilon) := m(\epsilon) \vee n(\epsilon, B_{\epsilon})$. It is now clear from (7.46)-(7.50) that whenever $n>n(\epsilon)$ in $\mathbb{N}$,

$$E^{\alpha} \left| \frac{1}{n} \sum_{i=1}^{n} c(X(i)) - \mathbb{E}^{g} c(X) \right| \leq 3\epsilon \tag{7.51}$$

and this completes the proof since $\epsilon>0$ is arbitrary. □

APPENDIX:

Lemma A.1 *Under the condition $\rho<1$, the inequality*

$$r(b):=a(z(b))b>1 \tag{A.1}$$

holds for all $b>1$.

Proof: With the mapping $\Phi:[1,+\infty)\times\mathbb{R}^{K+1}\rightarrow\mathbb{R}$ being defined by

$$\Phi(b,a):=\log b+\sum_{k=0}^{K}a_k\log\left(\frac{\mu_k}{b-1+\mu_k}\right) \tag{A.2}$$

for all pair (b,a) in $[1,+\infty)\times\mathbb{R}^{K+1}$, it is plain from Jensen's inequality that

$$r(b)=E^{\pi}[e^{\Phi(b,A(n))}]\geq e^{\Phi(b,\lambda)} \tag{A.3}$$

for all π in Π and all $n=1,2,...$ Simple calculations readily show that

$$\frac{d}{db}\Phi(b,a)=\frac{1}{b}[1-\sum_{k=0}^{K}a_k g_k(b)] \tag{A.4}$$

where the mappings $g_k:[1,-\infty)\rightarrow\mathbb{R}$ are defined by

$$g_k(b):=\frac{b}{b-1+\mu_k},\ 0\leq k\leq K. \tag{A.5}$$

Each one of these mappings is *monotone* decreasing on the interval $[1,+\infty)$ owing to the fact that

$$\frac{d}{db}g_k(b):=\frac{-(1-\mu_k)}{(b-1+\mu_k)^2}<0,\ 0\leq k\leq K, \tag{A.6}$$

and consequently for all $b>1$, $g_k(b)\leq g_k(1)=\frac{1}{\mu_k}$. Therefore,

$$\frac{d}{db}\Phi(b,\lambda)=\frac{1}{b}[1-\sum_{k=0}^{K}\lambda_k g_k(b)]$$

$$\geq\frac{1}{b}[1-\sum_{k=0}^{K}\lambda_k g_k(1)]=\frac{1-\rho}{b}>0 \tag{A.7}$$

and the mapping $b\rightarrow\Phi(b,\lambda)$ is thus monotone increasing with the property that $0=\Phi(1,\lambda)\leq\Phi(b,\lambda)$ on the interval $[1,+\infty)$. The conclusion (A.1) now immediately follows from the inequality (A.3). □

Proof of Theorem 4.1.: Fix b in the interval $[1,+\infty)$. For each $n=1,2,...$, the $\mathbb{F}_n$-stopping times $\sigma\wedge n$ and $\tau\wedge n$ being bounded, Doob's Optional Sampling Theorem [15], applied to the $(P^{\pi},\mathbb{F}_n)$-martingale $\{M(n,b)\}_1^{\infty}$, gives

$$E^{\pi}[M(\tau\wedge n,b)\mid \mathbb{F}_{\sigma\wedge n}] = M(\sigma\wedge n,b). \qquad n=1,2,\ldots(A.8)$$

The event $[n<\sigma]$ is $\mathbb{F}_{\sigma\wedge n}$-measurable and on it, the identity

$$M(\sigma\wedge n,b) = M(\tau\wedge n,b) = M(n,b) \qquad n=1,2,\ldots(A.9)$$

holds. The relation

$$E^{\pi}[I[n<\sigma]M(\tau\wedge n,b)\mid \mathbb{F}_{\sigma\wedge n}] = I[n<\sigma]M(\sigma\wedge n,b) \qquad n=1,2,\ldots(A.10)$$

therefore follows, owing to the $\mathbb{F}_{\sigma\wedge n}$-measurability of the RV $M(\sigma\wedge n,b)$, and the relation (A.8) thus reduces to

$$E^{\pi}[I[\sigma\leq n]M(\tau\wedge n,b)\mid \mathbb{F}_{\sigma\wedge n}] = I[\sigma\leq n]M(\sigma\wedge n,b). \qquad n=1,2,\ldots(A.11)$$

To proceed further, note that the RV's $\{M(\tau\wedge n,b)\}_1^{\infty}$ can be rewritten in the factored form

$$M(\tau\wedge n,b) = \frac{b^{L(\sigma\wedge n)}}{r(b)^{\sigma\wedge n}} \times \frac{b^{L(\tau\wedge n)-L(\sigma\wedge n)}}{r(b)^{\tau\wedge n-\sigma\wedge n}} \prod_{k=0}^{K} z_k(b)^{X_k(\tau\wedge n)} \qquad (A.12)$$

for all $n=1,2,\ldots$ Substitution of this last expression into (A.11) thus leads to the equality

$$E^{\pi}\left[I[\sigma\leq n]\frac{b^{L(\tau\wedge n)-L(\sigma\wedge n)}}{r(b)^{\tau\wedge n-\sigma\wedge n}}\prod_{k=0}^{K} z_k(b)^{X_k(\tau\wedge n)}\mid \mathbb{F}_{\sigma\wedge n}\right] \qquad (A.13)$$

$$= I[\sigma\leq n]\prod_{k=0}^{K} z_k(b)^{X_k(\sigma\wedge n)}, \qquad n=1,2\ldots.$$

after some easy simplifications that exploit the fact that the first factor on the right handside of (A.12) is $\mathbb{F}_{\sigma\wedge n}$-measurable.

Direct inspection shows that on the event $[\sigma\leq n<\tau]$, the relations

$$\tau\wedge n-\sigma\wedge n = n-\sigma\wedge n \quad \text{and} \quad L(\tau\wedge n)-L(\sigma\wedge n) = I[X(\sigma\wedge n=0)], \qquad (A.14)$$

hold for all $n=1,2,\ldots$, whence

$$0 \leq E^{\pi}\left[I[\sigma\leq n<\tau]\frac{b^{L(\tau\wedge n)-L(\sigma\wedge n)}}{r(b)^{\tau\wedge n-\sigma\wedge n}}\prod_{k=0}^{K} z_k(b)^{X_k(\tau\wedge n)}\mid \mathbb{F}_{\sigma\wedge n}\right]$$

$$= \frac{b^{I[X(\sigma\wedge n)=0]}}{r(b)^n} E^{\pi}[I[\sigma\leq n<\tau]r(b)^{\sigma\wedge n}\mid \mathbb{F}_{\sigma\wedge n}] \qquad (A.15)$$

$$\leq \frac{br(b)^{\sigma\wedge n}}{r(b)^n} I[\sigma\leq n] \leq \frac{br(b)^{\sigma}}{r(b)^n} \qquad n=1,2,\ldots(A.16)$$

where the passage from (A.15) to (A.16) made use of the fact that the RV's $\sigma\wedge n$ and $I[\sigma\leq n]$ are both $\mathbb{F}_{\sigma\wedge n}$-measurable. By Lemma A.1, $r(b)>1$ whenever $b>1$ under the assumed condition $\rho<1$, and the bounds (A.16) immediately imply that on $[\sigma<\infty]$,

$$\lim_{n\to\infty} E^{\pi}\left[I[\sigma\le n<\tau] \frac{b^{L(\tau\wedge n)-L(\sigma\wedge n)}}{r(b)^{\tau\wedge n-\sigma\wedge n}} \prod_{k=0}^{K} z_k(b)^{X_k(\tau\wedge n)} \mid \mathbb{F}_{\sigma\wedge n} \right] = 0 \quad P^{\pi}\text{-}a.s. \tag{A.17}$$

On the other hand, on the event $[\tau\le n]$,

$$\tau\wedge n-\sigma\wedge n=\tau-\sigma=\nu \ , \ L(\tau\wedge n)-L(\sigma\wedge n)=I[X(\sigma\wedge n)=0], \tag{A.18a}$$

while

$$X(\tau\wedge n)=0, \qquad n=1,2,\ldots \tag{A.18b}$$

and therefore

$$E^{\pi}\left[I[\tau\le n] \frac{b^{L(\tau\wedge n)-L(\sigma\wedge n)}}{r(b)^{\tau\wedge n-\sigma\wedge n}} \prod_{k=0}^{K} z_k(b)^{X_k(\tau\wedge n)} \mid \mathbb{F}_{\sigma\wedge n} \right]$$

$$= E^{\pi}[I[\tau\le n]\frac{1}{r(b)^{\nu}} \mid \mathbb{F}_{\sigma\wedge n}]\, b^{I[X(\sigma\wedge n)=0]} \tag{A.19}$$

$$= I[\sigma\le n]E^{\pi}[I[\tau\le n]\frac{1}{r(b)^{\nu}} \mid \mathbb{F}_{\sigma}]\, b^{I[X(\sigma)=0]} \tag{A.20}$$

for all $n=1,2,\ldots$ The passage from (A.19) to (A.20) is readily justified by well known properties of conditional expectations ([15] , Thm. pp.) based on the fact that the traces of the σ-fields $\mathbb{F}_{\sigma\wedge n}$ and $\mathbb{F}_{\sigma}$ coincide on the event $[\sigma\le n]$. It now follows from the Monotone Convergence Theorem for conditional expectations that

$$\lim_{n\to\infty} E^{\pi}[I[\tau\le n] \frac{b^{L(\tau\wedge n)-L(\sigma\wedge n)}}{r(b)^{\tau\wedge n-\sigma\wedge n}} \prod_{k=0}^{K} z_k(b)^{X_k(\tau\wedge n)} \mid \mathbb{F}_{\sigma\wedge n}]$$

$$= I[\sigma<\infty]E^{\pi}[I[\tau<\infty]\frac{1}{r(b)^{\nu}} \mid \mathbb{F}_{\sigma}]\, b^{I[X(\sigma)=0]} \quad P^{\pi}\text{-}a.s. \tag{A.21}$$

Upon combining (A.17) and (A.21), the reader will now check that

$$\lim_{n\to\infty} E^{\pi}\left[I[\sigma\le n] \frac{b^{L(\tau\wedge n)-L(\sigma\wedge n)}}{r(b)^{\tau\wedge n-\sigma\wedge n}} \prod_{k=0}^{K} z_k(b)^{X_k(\tau\wedge n)} \mid \mathbb{F}_{\sigma\wedge n} \right]$$

$$= I[\sigma<\infty]E^{\pi}[I[\tau<\infty]\frac{1}{r(b)^{\nu}} \mid \mathbb{F}_{\sigma}]\, b^{I[X(\sigma)=0]} \quad P^{\pi}\text{-}a.s. \tag{A.22}$$

whereas

$$\lim_{n\to\infty} I[\sigma\le n] \prod_{k=0}^{K} z_k(b)^{X_k(\sigma\wedge n)} = I[\sigma<\infty] \prod_{k=0}^{K} z_k(b)^{X_k(\sigma)}. \tag{A.23}$$

The conclusion (4.18) is now obtained by letting n go to infinity in (A.13) and by using the limits (A.22)-(A.23). □

REFERENCES:

[1] 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).

[2] 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).

[3] J. S. Baras, D. -J. Ma, and A. M. Makowski, "K competing queues with geometric service requirements and linear costs: the mu-c rule is always optimal," *Systems and Control Letters* vol. 6(3), pp. 173-180 (August 1985).

[4] C. Buyyukkoc, P. P. Varaiya, and J. Walrand, "The cmu -rule revisited," *Adv. Appl. Prob.* vol. 17 , pp. 234-235 (March 1985).

[5] K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, Second Edition, Springer-Verlag, New York (1967).

[6] K. L. Chung, *A course in probability theory*, Second Edition, Academic Press, New York (1974).

[7] 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).

[8] B. Hajek, "Hitting times and occupation time bounds implied by drift analysis with applications," *Adv. Appl. Prob.* vol. 14, pp. 501-525 (September 1982).

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

[10] S. Karlin and H. Taylor, *A First Course in Stochastic Processes*, Academic Press, New York (1974).

[11] 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).

[12] A. M. Makowski and A. Shwartz, "Implementation issues for Markov decision processes.," in *Proceedings of a workshop on Stochastic Differential Systems, Institute of Mathematics and its applications, Universilty of Minnesota*, ed. W. Fleming and P.-L. Lions, Springer Verlag Lecture Notes is Control and Information Sciences (1986).

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

[14] P. Nain and K. W. Ross, "Optimal priority assignment with hard constraint," *IEEE Trans. Auto. Control* vol. AC-31, no. 10, pp. 883-888 (October 1986).

[15] J. Neveu, *Discrete Parameter Martingales*, North-Holland, Amsterdam, Holland (1975).

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

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

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

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

[20] A. Shwartz and A. M. Makowski, *Adaptive policies for a system of competing queues II: Implementable schemes for optimal server allocation.*, Systems Research Report, In preparation, Systems Research Center, University of Maryland at College Park (1986).

[21] A. Shwartz and A. M. Makowski, "An optimal adaptive scheme for two competing queues with constraints," pp. 515-532 in *Proceedings of the 7th International Conference on Analysis and Optimization of Systems*, ed. A. Bensoussan and J. -L. Lions, Springer Verlag Lecture Notes in Control and Information Sciences, Antibes, France (June 1986).