JB 83-02

Paper Entitled

"Two Competing Queues with Linear Costs:
The μ c-rule is Often Optimal"

From the Proceedings of

22nd IEEE Conference on Decision and Control
pp. 1173-1178, December 1983

AND

Advances in Applied Probability
pp. 186-209
1985

Adv. Appl. Prob. **17**, 186–209 (1985)
Printed in N. Ireland

TWO COMPETING QUEUES WITH LINEAR COSTS AND GEOMETRIC SERVICE REQUIREMENTS: THE μc-RULE IS OFTEN OPTIMAL

J. S. BARAS,* *University of Maryland*
A. J. DORSEY,** *IBM—Federal Systems Division*
A. M. MAKOWSKI,* *University of Maryland*

Abstract

A discrete-time model is presented for a system of two queues competing for the service attention of a single server with infinite buffer capacity. The service requirements are geometrically distributed and independent from customer to customer as well as from the arrivals. The allocation of service attention is governed by feedback policies which are based on past decisions and buffer content histories. The cost of operation per unit time is a linear function of the queue sizes. Under the model assumptions, a fixed prioritization scheme, known as the μc-rule, is shown to be optimal for the expected long-run average criterion and for the expected discounted criterion, over both finite and infinite horizons. Two different approaches are proposed for solving these problems. One is based on the dynamic programming methodology for Markov decision processes, and assumes the arrivals to be i.i.d. The other is valid under no additional assumption on the arrival stream and uses direct comparison arguments. In both cases, the sample path properties of the adopted state-space model are exploited.

DISCRETE-TIME MODEL; SINGLE SERVER; INFINITE BUFFER; FEEDBACK; SAMPLE PATHS

1. Introduction

Dynamic control of queueing systems is currently a subject of great interest, due to potential applications in the design and performance evaluation of computer systems and communication networks. Unfortunately, classical queueing-theoretic methods typically treat static and/or steady-state situations, and do not extend easily to the more complex queueing optimization models.

Received 20 April 1983; revision received 1 June 1984.

* Postal address: Electrical Engineering Department, University of Maryland, College Park, MD 20742, USA.

** Postal address: IBM—Federal Systems Division, 21 Firstfield Road, Gaithersburg, MD 20748, USA.

This work was supported in part by the Department of Energy under Grant DOE-AC01-78ET 29244, A5 and the National Science Foundation under Grant ECS-82-04451.

During the past decade, however, progress has been made on specific dynamic control problems through the use of a wide variety of techniques; examples are found in the work of Ephremides, Varaiya and Walrand [9], Hajek [10], Harrison [12], and Rosberg, Varaiya and Walrand [21].

For practical applications, it is highly desirable to isolate models for which dynamical results can be obtained under as *minimal* a set of statistical hypotheses as possible. In some sense, such studies can be viewed as contributions to the *adaptive control* of queues [2], [8] since quite often in practice, the parameter values required to specify the queueing model fully are *not* known exactly. A definite need thus exists for a better understanding of the sensitivity of control results to statistical assumptions. This paper represents an effort in this direction for the problem of dynamically controlling two queues that compete for the service attention of a single server.

Motivation for studying this problem, as well as other similar questions of dynamic priority assignment, can be found in a wide variety of application areas ranging from urban traffic control to computer modelling to multi-user communications. Typically, the search for modes of operation that resolve and/or avoid conflict in resource-sharing environments can be formalized as problems of dynamic priority assignment. The reader is referred for instance to references [24]–[26] where some prioritization schemes for multi-node packet-radio networks are described. Mention should also be made of the monograph by Kleinrock [15] for a variety of examples, and of the survey paper by Reiser [20] for a discussion of polling schemes. An application to urban traffic control is given by Baras and Dorsey in [1].

In more specific terms, the problem analyzed here is a dynamic priority assignment problem with *two* classes of customer. A natural time unit is postulated and used to divide the time axis into contiguous slots of unit length. A *single* server is in attendance in the system and the buffer area has *infinite* capacity. The service requirements, which are *geometrically* distributed with a class dependent rate, are *independent* from customer to customer and from the arrivals. Customers that enter the service facility during a time slot, join the customers of their respective class which have not yet been serviced by the end of the previous time slot and await service in the buffer. At the beginning of each time slot, the queues compete for the server's attention which is allocated to one of the queues, for the duration of the slot, on the basis of *past decisions* and *buffer content* histories. Once service attention has been given out, *at most one* service completion will occur during that slot.

The incurred cost per unit time is assumed *linear* in the queue sizes; it induces the three performance measures considered in this paper, namely the discounted costs over the finite and infinite horizons and the expected long-run average cost. The main result of this paper is that under no additional assumption on the arrival patterns, the so-called *μc-rule* is optimal for all three criteria. The μc-rule is essentially a *static prioritization* of the two classes of customers and is defined *solely* on the basis of *service* and *cost* parameters. This result is remarkably simple and quite useful in practical applications, since the often postulated Bernoulli assumption is not always justified and the performance of a queueing system is commonly assessed by the combined use of several objective functions.

The discounted finite-horizon problems are discussed first and two different arguments are provided for establishing the optimality of the μc-rule. One of the arguments follows the standard dynamic programming methodology for Markov decision processes (under the additional assumption that the arrivals are i.i.d.) and was given in the technical report [3] with different proofs, while the other approach uses direct comparison arguments. An extension of this result is given in the works of Baras, Ma and Makowski [4] and of Buyukkoc, Varaiya and Walrand [6] for systems with an arbitrary number of customer classes.

An interesting and somewhat novel feature of the work lies in the adopted model. Here, the system evolution is given by *explicit state-space dynamics*, and not through the *prescription of transition probabilities* as it is traditionally done for Markov decision problems [16]. This sample path approach brings about a simple cost transformation which is helpful in isolating the μc-rule as a reasonable candidate for optimality and in by-passing the technical difficulties associated with unbounded rewards [17]. Moreover, this cost transformation sheds additional light on the nature of the solution by making contact with a class of optimal stochastic control problems known as *arm-acquiring bandit* problems [27].

Priority assignment problems have received considerable attention over the years, as already evidenced by the thorough exposition of Jaiswal [13] on priority queues. However, the bulk of this work deals with continuous-time models, in contrast with the discrete-time situation studied here and in [1]–[4], [6], [8]. An early discussion on a related *static* priority assignment problem was given by Cox and Smith [7] for the average waiting-time criterion at steady-state. They considered a single-server queueing system fed by different classes of customers with independent Poissonian arrivals; within each class, service times were modelled by independent identically distributed random variables with a general class-dependent probability distribution function. The optimal static strategy was identified in [7] to be a *fixed work-conserving* prioritization scheme of the customer classes known as the μc-rule. Extension to this work was given by Rykov and Lembert [23] and by Kakalik [14]; they showed that the μc-rule is also optimal even amongst all feedback rules, i.e., rules that allow the controller knowledge of queue *sizes* at each decision epoch.

Harrison [11], [12] studied the model of Cox and Smith [7] with the objective of maximizing the expected net present value of service rewards received minus holding costs incurred over an infinite planning horizon and discounted at a positive rate. There again, a very special type of priority assignment, called a *modified static* policy, was shown to be optimal in the class of all feedback policies based on queue size histories. This particular rule provides a fixed prioritization of customer classes, which is explicitly computable by a finite-step algorithm [12].

The paper is organized as follows. The mathematical model of a discrete-time system with two competing queues is developed in Section 2 and the various control problems associated with it are described in Section 3. Auxiliary discounted problems with bounded cost per unit time are then introduced in Section 4 via a simple cost transformation. The optimality of the μc-rule is established in Section 5 in two simple situations, so as to add plausibility to the general result. Section 6 contains a discussion of the optimality results of this paper. The dynamic programming arguments are given in Section 7 while the direct comparison arguments are presented in Section 8.

Finally, to fix the notation, the set of all non-negative integers and the set of all real numbers are denoted by N and R, respectively.

2. The model

In this section a simple discrete-time model is formulated to capture the evolution of the system with two competing queues loosely described in the introduction.

To describe the model, an underlying probability triple $(\Omega, \mathscr{F}, P)$ is assumed that simultaneously carries an N^2-valued random variable Ξ, a sequence $\{A(t)\}_0^\infty$ of N^2-valued random variables and a sequence $\{B(t)\}_0^\infty$ of $\{0, 1\}^2$-valued random variables. As a notational convention, the first and second components of an N^2-valued random variable (or element of N^2, respectively) are always denoted by the same symbol as the random variable (or element) but subscripted by 1 and 2, respectively.

The random variable Ξ_1 represents the initial size of the first queue, the random variable $A_1(t)$ quantifies the arrivals to this queue during the time slot $[t, t+1)$ whereas $B_1(t)$ records completion of service during that time period. The random variables Ξ_2, $A_2(t)$ and $B_2(t)$ receive identical interpretations in the context of the second queue.

The assignment of service attention in the slot $[t, t+1)$ is based on knowledge of past queue sizes and control decisions. An *admissible control policy* π is thus understood to be any collection $\{\pi_t\}_0^\infty$ of mappings π_t from $N^2\times(N^2\times\{0, 1\})^t$ into $\{0, 1\}$, with the convention that the domain of π_0 is simply the Cartesian product N^2. The collection of all such admissible policies is denoted by $\mathcal{P}$.

For any admissible policy π in $\mathcal{P}$, the sequences of random variables $\{X^\pi(t)\}_0^\infty$ and $\{U^\pi(t)\}_0^\infty$ represent the queue sizes and control sequences associated with (or generated by) the admissible policy π. They take values in N^2 and $\{0, 1\}$, respectively, and satisfy the following state dynamics: set $X^\pi(0) := \Xi$, $U^\pi(0) := \pi_0(X^\pi(0))$ and for all t in N, define

$$
\begin{aligned}
X_1^\pi(t+1) &:= X_1^\pi(t) + A_1(t) - \mathbb{1}(X_1^\pi(t))U^\pi(t)B_1(t) \\
X_2^\pi(t+1) &:= X_2^\pi(t) + A_2(t) - \mathbb{1}(X_2^\pi(t))(1 - U^\pi(t))B_2(t),
\end{aligned} \tag{2.1}
$$

with

$$
U^\pi(t+1) := \pi_{t+1}(X^\pi(0), \cdots, X^\pi(t+1); U^\pi(0), \cdots, U^\pi(t)). \tag{2.2}
$$

In (2.1) and throughout the discussion, $\mathbb{1}(n)$ is defined for all n in N to be $\mathbb{1}(n) = 1$ if $n \neq 0$ and by $\mathbb{1}(n) = 0$ if $n = 0$.

Some technical assumptions are needed. They are motivated by the fact that the service requirements should be geometrically distributed and independent from customer to customer. Moreover, they serve as a means to validate a useful cost transformation forthcoming in Section 3. Thus, the following 1
assumptions are postulated and enforced throughout this paper. 2

(A1) *The sequence* $\{B(t)\}_0^\infty$ *and* $\{A(t)\}_0^\infty$, *and the random variable* Ξ *are mutually independent.*

(A2) *The sequences* $\{B_1(t)\}_0^\infty$ *and* $\{B_2(t)\}_0^\infty$ *are independent Bernoulli sequences with known parameter* μ_1 *and* μ_2, *respectively.*

To fix the terminology, an admissible policy π in $\mathcal{P}$ is said to be (of) *non-idling* or *work-conserving* (type) if both conditions

$$
\begin{aligned}
&(X_1^\pi(t) = 0 \text{ and } X_2^\pi(t) > 0) \quad \text{imply} \quad U^\pi(t) = 0 \\
&(X_2^\pi(t) = 0 \text{ and } X_1^\pi(t) > 0) \quad \text{imply} \quad U^\pi(t) = 1
\end{aligned} \tag{2.3}
$$

hold true for all t in N. An admissible policy π which is not of non-idling type is said to be of *idling* type. The constraint (2.3) on a policy π has a simple interpretation; it is imposed to avoid waste of the system resources by preventing an empty queue from receiving service attention in a slot during which the other queue contains a customer.

For obvious practical reasons, it is often necessary to consider admissible policies that depend in a memoryless fashion on the observed past. In accordance with the standard terminology [16], [22], an admissible policy π in $\mathcal{P}$ is said to be (of) *Markov* (type) if for all t in N, the corresponding mappings π_t reduce to mappings from N^2 into $\{0, 1\}$. In that case, the control sequence

$\{U^\pi(t)\}_0^\infty$ is given by $U^\pi(t) = \pi_t(X^\pi(t))$ for all t in N. Moreover, if the admissible policy π is Markov and *all* its mappings π_t are identical, the policy π is said to be *stationary* [16], [22]. By convention, the same symbol will be used throughout to denote *both* any such Markov stationary policy and the mapping from N^2 into $\{0, 1\}$ that generates it. The collection of all such Markov (or Markov stationary) policies is denoted by $\mathcal{M}$ (or $\mathcal{S}$, respectively).

3. The optimal control problems

Simple measures of performance are associated with the operation of this two-queue system by imposing an instantaneous cost proportional to queue sizes. Specifically, if c_1 and c_2 are two positive constants held fixed hereafter, the cost per slot is given as the mapping c from N^2 into R, where

$$c(x) := c_1 x_1 + c_2 x_2 \tag{3.1}$$

for every x in N^2. For every admissible policy π in $\mathcal{P}$, three measures of system performance are associated with (3.1) by setting

$$J_{\beta,t}(\pi) := E\left[\sum_{s=0}^{t} \beta^s c(X^\pi(s))\right], \tag{3.2}$$

$$J_\beta(\pi) := \lim_{t\uparrow\infty} J_{\beta,t}(\pi) = E\left[\sum_{s=0}^{\infty} \beta^s c(X^\pi(s))\right] \tag{3.3}$$

and

$$J_{av}(\pi) := \liminf_{t\uparrow\infty} \frac{1}{1+t} J_{1,t}(\pi) = \liminf_{t\uparrow\infty} E\left[\frac{1}{1+t} \sum_{s=0}^{t} c(X^\pi(s))\right], \tag{3.4}$$

where β is a discount factor in $[0, 1]$, t is in N and the random variables $\{X^\pi(s)\}_0^\infty$ are generated via the dynamics (2.1)–(2.2). The quantities $J_{\beta,t}(\pi)$ and $J_\beta(\pi)$ are the *expected β-discounted* costs associated with the admissible policy π in $\mathcal{P}$ over the *finite* horizon $[0, t]$ and the *infinite* horizon $[0, \infty)$, respectively, whereas $J_{av}(\pi)$ is the corresponding *expected long-run average* cost.

Three families of optimal control problems $(P_{\beta,t})$, (P_β) and (P_{av}) are studied in this paper. They are simultaneously defined below as problem (P), with the convention that $J(\pi)$ represents any one of the cost functions (3.2)–(3.4) and that in each case, the parameter range is the one for the corresponding cost function.

(P): *Minimize $J(\pi)$ over the class $\mathcal{P}$ of all admissible control policies π.*

The discounted problems are studied first, for they provide the key to solving the long-run average problem. Conditions under which these discounted problems are meaningfully defined are easily obtained. To this end, define the quantities $Q_{\beta,t}$ and Q_β by

$$Q_{\beta,t} := E[c(\Xi)] + \beta \sum_{s=0}^{t} \beta^s E[c(A(s))] \tag{3.5}$$

and

$$Q_\beta := \lim_{t\uparrow\infty} Q_{\beta,t} = E[c(\Xi)] + \beta \sum_{s=0}^{\infty} \beta^s E[c(A(s))] \tag{3.6}$$

for all β in $[0, 1]$ and t in N, with the convention that $Q_{\beta,-1} := E[c(\Xi)]$. For every admissible policy π in $\mathcal{P}$, the bounds

$$0 \leqq J_{\beta,t}(\pi) \leqq \frac{Q_{\beta,t-1}}{1-\beta} \tag{3.7}$$

and

$$0 \leqq J_\beta(\pi) \leqq \frac{Q_\beta}{1-\beta} \tag{3.8}$$

hold true with β in $[0, 1)$ and t in N. Indeed, the state dynamics (2.1) readily implies the inequalities

$$0 \leqq c(X^\pi(t+1)) \leqq c(\Xi) + \sum_{s=0}^{t} c(A(s)) \tag{3.9}$$

for all t in N. Substitution of (3.9) into (3.2) and (3.3) and use of the monotone convergence theorem yield (3.7) and (3.8) after routine calculations.

As a result of the bound (3.7) and of (2.1), each one of the problems $(\mathrm{P}_{\beta,t})$, β in $[0, 1)$ and t in N, is thus well defined *if and only if* the finiteness condition

$$Q_{\beta,t-1} < \infty \tag{3.10}$$

holds, since $E[c(\Xi)] \leqq J_{\beta,t}(\pi)$. Observe that if (3.10) fails to hold, then the cost functional $J_{\beta,t}(\pi)$ is infinite for every policy π in $\mathcal{P}$. Similarly, the bound (3.8) shows that problem (P_β) is well defined *if and only if* the more constraining finiteness condition

$$Q_\beta < \infty \tag{3.11}$$

holds. Similar comments can be made for the undiscounted problems $(\mathrm{P}_{1,t})$, t in N, and (P_1).

Parts of the discussion presented in this paper will be given under an additional assumption (A3).

(A3) *The random variables* $\{A(t)\}_0^\infty$ *form a sequence of independent and identically distributed random variables.*

It should be emphasized that the sequences $\{A_1(t)\}_0^\infty$ and $\{A_2(t)\}_0^\infty$ are not necessarily mutually independent, nor are the random variables $A_1(t)$ and $A_2(t)$ independent of each other at each time t.

4. A cost transformation and the corresponding auxiliary problems

Under the statistical assumptions (A1)–(A3) introduced earlier, all three problems become problems in the theory of Markov decision processes as discussed for instance in the monographs of Bertsekas [5], Kushner [16] and Ross [22], Chapter 6. However, for the problems studied here, the cost per unit time is *not* a bounded function of its arguments (since linear) and the basic approach via the functional equation(s) of dynamic programming is much more cumbersome. Although this unboundedness could be handled by a modified dynamic programming methodology due to Lippman [17], this technical difficulty motivated an alternate and simpler line of argumentation which is believed to shed some light on the problem. The approach consists of relating the discounted problems formulated in Section 3 to auxiliary problems of the same type for which the cost per unit time is now a *bounded* function of its arguments; in fact, problem (P_β) will turn out to be *equivalent* to a corresponding infinite-horizon discounted problem and Markov decision theory need only be used (when applicable) in its elementary form to yield the various conclusions of Sections 6 and 7.

This equivalence is brought about by a simple *cost transformation* given in Proposition 4.1 and is made possible by directly using the *sample path* nature of the state dynamics (2.1)–(2.2). This model description represents a departure from the more traditional approach to Markov decision problems, where the state evolution is usually specified through *state transition probabilities.*

To proceed with the discussion, a mapping $\tilde{c}$ is defined from $\{0, 1\}\times N^2$ into R by

$$\tilde{c}(u, x) := \mu_1 c_1 \mathbb{1}(x_1)u + \mu_2 c_2 \mathbb{1}(x_2)(1-u) \tag{4.1}$$

with u in $\{0, 1\}$ and x in N^2. Two quantities are associated with every admissible policy π in $\mathcal{P}$ by setting

$$\tilde{J}_{\beta,t}(\pi) := E\left[\sum_{s=0}^{t} \beta^s \tilde{c}(U^\pi(s), X^\pi(s))\right] \tag{4.2}$$

and

$$\tilde{J}_\beta(\pi) := \lim_{t\uparrow\infty} \tilde{J}_{\beta,t}(\pi) = E\left[\sum_{s=0}^{\infty} \beta^s \tilde{c}(U^\pi(s), X^\pi(s))\right] \tag{4.3}$$

where β is a discount factor in $[0, 1]$ and t is in N. The corresponding cost per unit time $\tilde{c}$ is now a *bounded* function of its arguments, with the obvious bounds

$$0 \leqq \tilde{c}(u, x) \leqq \mu_1 c_1 + \mu_2 c_2 \tag{4.4}$$

valid for every u in $\{0, 1\}$ and every x in N^2.

These auxiliary costs are related to the original cost functionals in the following manner.

Proposition 4.1. *For any discount factor β in* $[0, 1)$ *and any t in N, both relations*

$$J_{\beta,t+1}(\pi) = Q_{\beta,t} + \beta J_{\beta,t}(\pi) - \beta \tilde{J}_{\beta,t}(\pi) \tag{4.5}$$

$$J_\beta(\pi) = Q_\beta + \beta J_\beta(\pi) - \beta \tilde{J}_\beta(\pi) \tag{4.6}$$

hold true for any admissible policy π in $\mathcal{P}$, with $Q_{\beta,t}$ and Q_β given by (3.5) *and* (3.6), *respectively.*

Proof. In view of (4.4), the uniform bound

$$0 \leqq \tilde{J}_{\beta,t}(\pi) \leqq \tilde{J}_\beta(\pi) \leqq \frac{\mu_1 c_1 + \mu_2 c_2}{1-\beta} \tag{4.7}$$

is obtained for all π in $\mathcal{P}$ and all t in N, whence only (4.5) needs to be established as (4.6) will follow by a simple limiting argument.

To establish (4.5), observe from (2.1) that for every admissible policy π in $\mathcal{P}$,

$$\begin{aligned} E[c(X^\pi(s+1))] &= E[c(X^\pi(s))] + E[c(A(s))] \\ &\quad - E[c_1 \mathbb{1}(X_1^\pi(s)) U^\pi(s) B_1(s) + c_2 \mathbb{1}(X_2^\pi(s))(1 - U^\pi(s)) B_2(s)] \end{aligned} \tag{4.8}$$

as s ranges over N. Under the statistical assumptions (A1)–(A2), the random variables $B_1(s)$ and $B_2(s)$ are independent of the random variables $X^\pi(s)$ and $U^\pi(s)$ for all s in N and the equality

$$\begin{aligned} E[c_1 \mathbb{1}(X^\pi(s)) U^\pi(s) B_1(s) + c_2 \mathbb{1}(X_2^\pi(s))(1 - U^\pi(s)) B_2(s)] \\ = E[\tilde{c}(U^\pi(s), X^\pi(s))] \end{aligned} \tag{4.9}$$

follows by an easy conditioning argument. Substitution of (4.9) in (4.8) readily implies that

$$\begin{aligned} \beta^{s+1} E[c(X^\pi(s+1))] &= \beta^{s+1} E[c(X^\pi(s))] + \beta^{s+1} E[c(A(s))] \\ &\quad - \beta^{s+1} E[\tilde{c}(U^\pi(s), X^\pi(s))] \end{aligned} \tag{4.10}$$

for all s in N. Now summing (4.10) over s ranging from 0 to t and regrouping corresponding terms in the resulting expression yields the desired result (4.5) after routine calculations.

The form of (4.5) and (4.6) naturally suggests the introduction of the auxiliary control problems $(\tilde{P}_{\beta,t})$ and $(\tilde{P}_\beta)$ that correspond to the cost functionals (4.2) and (4.3). These problems are defined simultaneously below as problem $(\tilde{P})$ with the convention that $\tilde{J}(\pi)$ represents any one of these cost functions and that in each case, the parameter's range is the one for the corresponding cost function.

$(\tilde{P})$ *Maximize* $\tilde{J}(\pi)$ *over the class* $\mathcal{P}$ *of all admissible control policies* π.

The relation (4.6) easily implies the following conclusion which is now stated as a corollary.

Corollary 4.1. *For any discount factor* β *in* $[0, 1)$, *the optimization problems* (P_β) and $(\tilde{P}_\beta)$ *are equivalent in the sense that they have the same extremizers, provided the finiteness condition* (3.11) *holds.*

A similar result holds for the problems $(P_{\beta,t})$ and $(\tilde{P}_{\beta,t})$ as will be seen from the forthcoming analysis given in Section 6. However, Corollary 4.1 already provides the motivation for studying the auxiliary problems and, as it is now pointed out, significant advantages can be gained from doing so.

First, the suggestive form of the costs (4.2) and (4.3) naturally points to the so-called *μc-rule* as a reasonable candidate for optimality. For future reference, the μc-rule is defined as the Markov stationary policy μ^* given by

$$\mu^*(x) := \begin{cases} 1 - \mathbb{1}(x_2) & \text{if } \mu_1 c_1 \geqq \mu_2 c_2 \\ \mathbb{1}(x_1) & \text{if } \mu_2 c_2 \geqq \mu_1 c_1 \end{cases} \tag{4.11}$$

for all x in N^2. This is an example of *myopic* optimization since the cost (4.1) per unit time is maximized at *every* stage by (4.11), in contrast with *global* optimality as required in the formulation of both (P_β) and $(\tilde{P}_\beta)$. In the context of problem (P_β), the μc-rule can be interpreted as a fixed prioritization scheme that maximizes the *expected cost decrease per time slot.*

Second, the cost transformation implicit in Proposition 4.1 sheds some light on the control problems and the structure of their solution. Indeed, the auxiliary problems $(\tilde{P}_{\beta,t})$ and $(\tilde{P}_\beta)$ are *distributionally equivalent* to *arm-acquiring bandit* problems on which a huge literature is available; the reader is referred to the recent work of Varaiya, Walrand and Buyukkoc [27] and references therein for additional information on this class of problems. Problems $(\tilde{P}_{\beta,t})$ and $(\tilde{P}_\beta)$ would be genuine bandit problems, as defined in [27], provided the dynamics (2.1) were replaced by

$$\begin{aligned} X_1^\pi(t+1) &= X_1^\pi(t) + A_1(t) - \mathbb{1}(X_1^\pi(t))U^\pi(t)B_1(N_1^\pi(t)+1) \\ X_2^\pi(t+1) &= X_2^\pi(t) + A_2(t) - \mathbb{1}(X_2^\pi(t))(1-U^\pi(t))B_2(N_2^\pi(t)+1) \end{aligned} \tag{4.12}$$

for all t in N. Here, the random variable $N_1^\pi(t)$ (or $N_2^\pi(t)$) counts the number of slots over the horizon $[0, t)$ during which the first (or second) queue is non-empty and is given service attention according to the admissible policy π in $\mathcal{P}$. The solution of many bandit problems is by now well understood and can be characterized by a *dynamic allocation index*, of which the μc-rule is a very simple example.

5. Optimality of the μc-rule: Two simple cases

The remainder of this paper discusses the optimality of the μc-rule for the problems $(\tilde{\mathrm{P}}_{\beta,t})$ and $(\tilde{\mathrm{P}}_\beta)$ under the assumptions (A1)–(A2). To familiarize the reader with the general result, this section investigates two special versions of the problem for which this optimality is very easily established when $\mu_1 c_1 \neq \mu_2 c_2$. The discussion assumes $\mu_2 c_2 < \mu_1 c_1$, with the understanding that the symmetrical case can be treated similarly.

5.1. *Uninterrupted arrivals.* Here only, assume that the arrival process $\{A(t)\}_0^\infty$ satisfies the additional condition that $\mathrm{P}[A_1(t) \neq 0] = \mathrm{P}[A_2(t) = 0] \neq 1$ for all t in N. Under this assumption, for *any* admissible policy π in $\mathcal{P}$, each queue is empty *at most* once and this necessarily occurs at the starting time $t = 0$, owing to the dynamical equations (2.1). Hence, for all $t \neq 0$ in N

$$\tilde{c}(U^\pi(t), X^\pi(t)) = [\mu_1 c_1 - \mu_2 c_2] U^\pi(t) + \mu_2 c_2, \tag{5.1}$$

and after a simple substitution in (4.2), it follows that for all t in N,

$$\begin{aligned}\tilde{J}_{\beta,t}(\pi) = {} & E[\tilde{c}(\pi_0(\Xi), \Xi)] + [\mu_1 c_1 - \mu_2 c_2] E\left[\sum_{s=1}^{t} U^\pi(s) \beta^s\right] \\ & + \mu_2 c_2 \frac{1-\beta^t}{1-\beta} \beta,\end{aligned} \tag{5.2}$$

whenever $\beta \neq 1$. Since $\mu_1 c_1 - \mu_2 c_2$ is positive, the cost (5.2) is maximized by selecting π^* such that $U^{\pi^*}(s) = 1$ for all $1 \leqq s \leqq t$ in N *and* such that $\tilde{c}(\pi_0^*(\Xi), \Xi)$ is maximum. The μc-rule given by (4.11) meets these requirements and is therefore optimal for problems $(\tilde{\mathrm{P}}_{\beta,t})$ for *all* values of the discount factor β in $[0, 1)$, under the additional constraint of uninterrupted arrival streams, i.e., *at least one* customer joins *each* queue in any given time slot.

5.2. *First-order analysis for small discount factors.* To obtain yet another situation where the optimality of the μc-rule is readily established, consider problem $(\tilde{\mathrm{P}}_{\beta,t})$ for some t in N, with $0 < \beta < 1$. Observe that for *any* admissible policy π in $\mathcal{P}$ and for all $0 \leqq s < t$, the decomposition

$$\sum_{r=s}^{t} \beta^r \tilde{c}(U^\pi(r), X^\pi(r)) = \beta^s \tilde{c}(U^\pi(s), X^\pi(s)) + \sum_{r=s+1}^{t} \beta^r \tilde{c}(U^\pi(r), X^\pi(r)) \tag{5.3}$$

obviously holds, and the bounds

$$(5.4)\qquad 0 \leqq \sum_{r=s+1}^{t} \beta^r \tilde{c}(U^\pi(r), X^\pi(r)) \leqq (\mu_1 c_1 + \mu_2 c_2)\beta \frac{1-\beta^{t-s}}{1-\beta} \beta^s$$

are satisfied uniformly in the policy π, as a result of (4.4).

Now fix a policy π in $\mathcal{P}$ and s in N with $0 \leqq s < t$. A new policy $\tilde{\pi} = \{\tilde{\pi}_r\}_0^\infty$ may then be defined from π by setting $\tilde{\pi}_r := \pi_r$ for all $r \neq s$ in N and $\tilde{\pi}_s := \mu^*$. The policies π and $\tilde{\pi}$ agree *except* that at time s, the latter operates according to the μc-rule and produces the control action $\mu^*(X^\pi(s))$.

The contribution of $\tilde{c}(U^\pi(s), X^\pi(s))$ to the cost function $\tilde{J}_{\beta,t}(\pi)$ may be *one of at most three* values, namely $\mu_1 c_1$, $\mu_2 c_2$ and 0, depending on the value of the random variables $X^\pi(s)$ and $U^\pi(s)$. As a result, a direct inspection of (5.3) and (5.4) easily indicates, owing to the noted *uniformity*, that for β *small enough* (to be made precise in a moment) *in* $[0, 1)$, the policy $\tilde{\pi}$ incurs a better cost than policy π on the horizon $[0, t]$, i.e., $\tilde{J}_{\beta,t}(\pi) \leqq \tilde{J}_{\beta,t}(\tilde{\pi})$. This will happen if β is selected in $[0, 1)$ such that the condition

$$(5.5)\qquad (\mu_1 c_1 + \mu_2 c_2)\beta \frac{1-\beta^{t-s}}{1-\beta} \beta^s < \min\{\mu_1 c_1 - \mu_2 c_2, \mu_2 c_2\}\beta^s$$

holds. As expected, this condition is most constraining on β at $s = 0$, where the corresponding condition reads

$$(5.6)\qquad (\mu_1 c_1 + \mu_2 c_2) \frac{1-\beta^t}{1-\beta} \beta < \min\{\mu_1 c_1 - \mu_2 c_2, \mu_2 c_2\}.$$

From the discussion given earlier, the μc-rule μ^* is easily seen to be *optimal* for problem $(\tilde{\mathrm{P}}_{\beta,t})$ under the condition (5.6) on the discount factor β. This follows by an easy inductive argument from the fact that condition (5.6) implies (5.5) for all $0 \leqq s < t$ and from the myopic optimality of the μc-rule (for $s = t$) discussed in Section 4.

It should be observed that the arguments presented here rely only on the structure of the cost, and *not* on the statistical properties of the arrival streams. In other words, for β small enough, the μc-rule is *always* optimal for problems $(\tilde{\mathrm{P}}_{\beta,t})$ under *no* assumption on the arrival pattern. That this result holds for *all* values of β when the arrivals are merely *independent* of the service streams, constitutes the basic result of this paper. Extensions of this result to more than two queues were subsequently given by Baras, Ma and Makowski [4] and Buyukkoc, Varaiya and Walrand [6].

6. Optimality of the μc-rule: the results

The optimality of the μc-rule can now be established for the various control problems dicussed in earlier sections.

The finite-horizon discounted-cost problems $(\tilde{P}_{\beta,t})$ are discussed first.

Theorem 6.1. *Under the foregoing assumptions* (A1)–(A2), *the μc-rule is optimal for each one of the problems $(\tilde{P}_{\beta,t})$, β in* [0, 1] *and t in N. For β in* [0, 1), *the μc-rule is essentially the only optimal policy when $\mu_1 c_1 \neq \mu_2 c_2$, whereas every non-idling policy is optimal when $\mu_1 c_1 = \mu_2 c_2$.*

Detailed proofs of this key result are given in the next two sections where two different approaches are presented. The first derivation, discussed in Section 7, was given originally by the authors in [3]; it requires the additional assumption (A3) and uses the dynamic programming methodology for Markov decision processes. As an alternative, Section 8 presents simple comparison arguments to obtain the same result under the basic assumptions (A1)–(A2). These ideas already appeared in a preliminary version of [6] and have been used successfully to solve other questions of optimal priority assignment [18], [19].

The remainder of this section explores the implications of Theorem 6.1, as the desired optimality is first discussed for the infinite-horizon problems $(\tilde{P}_\beta)$.

Theorem 6.2. *Under the foregoing assumptions* (A1)–(A2), *the μc-rule is optimal for each one of the problems $(\tilde{P}_\beta)$, β in* [0, 1). *If $\mu_1 c_1 \neq \mu_2 c_2$, the μc-rule is essentially the only optimal policy, whereas if $\mu_1 c_1 = \mu_2 c_2$, every non-idling policy is optimal.*

Proof. As a result of Theorem 6.1, the inequality

$$\tilde{J}_{\beta,t}(\pi) \geqq \tilde{J}_{\beta,t}(\mu^*) \tag{6.1}$$

holds for every admissible policy π in $\mathcal{P}$ and t ranging over N. A simple application in (6.1) of the monotone convergence theorem shows that

$$\tilde{J}_\beta(\pi) = \lim_{t\uparrow\infty} \tilde{J}_{\beta,t}(\pi) \geqq \lim_{t\uparrow\infty} \tilde{J}_{\beta,t}(\mu^*) = \tilde{J}_\beta(\mu^*) \tag{6.2}$$

for every admissible policy π in $\mathcal{P}$, since the μc-rule μ^* is in $\mathcal{S}$; the μc-rule μ^* is thus optimal for problem $(\tilde{P}_\beta)$.

When $\mu_1 c_1 = \mu_2 c_2$, all non-idling policies are clearly optimal for problem $(\tilde{P}_\beta)$ by virtue of Theorem 6.1 and of the limiting argument giving (6.2).

To study the case $\mu_1 c_1 \neq \mu_2 c_2$, observe that for every admissible policy π

in $\mathcal{P}$,

$$\tilde{J}_\beta(\pi) = \sum_{t=0}^{\infty} \beta^t E[\tilde{c}(U^\pi(t), X^\pi(t))] \tag{6.3}$$

with β in its usual range $[0, 1)$. From the bounds (4.4), it is easy to see that the mapping $[0, 1) \to R: \beta \to \tilde{J}_\beta(\pi)$ is *monotone increasing, strictly convex* and *analytic* throughout its domain.

Now, let π be an admissible policy in $\mathcal{P}$ optimal for problem $(\tilde{\mathrm{P}}_\beta)$ at some value of β in $(0, 1)$, say β_0. The optimality of the μc-rule μ^* obtained earlier implies that $\tilde{J}_{\beta_0}(\pi) = \tilde{J}_{\beta_0}(\mu^*)$, hence

$$\tilde{J}_\beta(\pi) = \tilde{J}_\beta(\mu^*) \tag{6.4}$$

for all β in the interval $[\beta_0, 1)$, owing to the inequality (6.2) and to the strict monotonicity and convexity of the mapping $\beta \to \tilde{J}_\beta(\pi)$. By analytic continuation, (6.4) must also hold throughout the entire interval $[0, 1)$ and the equalities

$$E[\tilde{c}(U^\pi(t), X^\pi(t))] = E[\tilde{c}(U^{\mu^*}(t), X^{\mu^*}(t))] \tag{6.5}$$

immediately result, with t ranging in N. This now implies that $\tilde{J}_{\beta,t}(\pi) = \tilde{J}_{\beta,t}(\mu^*)$ for *all* t in N and β in $(0, 1)$, and the policies π and μ^* necessarily coincide by virtue of the uniqueness result stated in Theorem 6.1 for the finite-horizon problems. This shows that the μc-rule is again the only optimal policy for the infinite-horizon problems when $\mu_1 c_1 \neq \mu_2 c_2$.

Theorem 6.3. *Under the foregoing assumptions* (A1)–(A2), *the μc-rule is optimal for each one of the problems* $(\mathrm{P}_{\beta,t})$, *β in* $[0, 1]$ *and t in N.*

Proof. The discussion need only be given in the case when the finiteness assumption (3.10) holds for otherwise the result is obviously true; indeed, in that case, the cost functional $J_{\beta,t}(\pi)$ is infinite for every policy π in $\mathcal{P}$, as pointed out in Section 3.

Under this finiteness condition, the first part of Proposition 4.1 implies that for any admissible policy π in $\mathcal{P}$,

$$J_{\beta,t+1}(\pi) - J_{\beta,t+1}(\mu^*) = \beta[J_{\beta,t}(\pi) - J_{\beta,t}(\mu^*)] + \beta[\tilde{J}_{\beta,t}(\mu^*) - \tilde{J}_{\beta,t}(\pi)] \tag{6.6}$$

whenever β is in $[0, 1)$ and t in N. From (6.1), it now follows that

$$[J_{\beta,t+1}(\pi) - J_{\beta,t+1}(\mu^*)] \geqq \beta[J_{\beta,t}(\pi) - J_{\beta,t}(\mu^*)] \tag{6.7}$$

for all t in N, and the inequality

$$J_{\beta,t+1}(\pi) - J_{\beta,t+1}(\mu^*) \geqq [\beta + \cdots + \beta^{t+1}][J_{\beta,0}(\pi) - J_{\beta,0}(\mu^*)] \tag{6.8}$$

readily results by a direct iteration argument. It is now easy to see that for every policy π in $\mathcal{P}$,

$$J_{\beta,t+1}(\mu^*) \leqq J_{\beta,t+1}(\pi) \tag{6.9}$$

since $J_{\beta,0}(\pi)=E[c(\Xi)]=J_{\beta,0}(\mu^*)$, and the optimality of the μc-rule is thus established for $\beta\neq 1$.

The case $\beta=1$ is obtained by a simple continuity argument since the μc-rule μ^* does not depend on β. An application of the monotone convergence theorem on (6.9) yields

$$J_{1,t}(\mu^*)=\lim_{\beta\uparrow 1} J_{\beta,t}(\mu^*)\leqq \lim_{\beta\uparrow 1} J_{\beta,t}(\pi)=J_{1,t}(\pi) \tag{6.10}$$

for every policy π in $\mathcal{P}$.

The original infinite-horizon problems (P_β) are finally solved as follows.

Theorem 6.4. *Under the foregoing assumptions* (A1)–(A2), *the μc-rule is optimal for each one of the problems* (P_β), β *in* $[0, 1)$.

Proof. The limiting argument that gave Theorem 6.2 when applied to (6.1) also establishes the optimality of the μc-rule for the problems (P_β) when used on (6.10). An alternate argument consists of combining Theorem 6.2 with Corollary 4.1 and the closing remarks of Section 3.

The proof of Theorems 6.3 and 6.4 clearly show that for problems $(\mathrm{P}_{\beta,t})$ and (P_β) with β in $[0, 1)$, the μc-rule is again the essentially unique optimal policy when $\mu_1 c_1\neq\mu_2 c_2$ whereas any non-idling policy is optimal when $\mu_1 c_1=\mu_2 c_2$ provided the finiteness assumptions (3.10) and (3.11) hold. The optimality results concerning the long-run average-cost problem $(\mathrm{P}_{\mathrm{av}})$ are also within reach.

Theorem 6.5. *Under the foregoing assumptions* (A1)–(A2), *the μc-rule is optimal for the problem* $(\mathrm{P}_{\mathrm{av}})$.

Proof. The result is immediate from (6.10) and the definition (3.4) of the cost function $J_{\mathrm{av}}(\pi)$.

Obviously, this last result is still valid when the definition (3.4) of the long-run average-cost function is changed to

$$J_{\mathrm{av}}(\pi):=\limsup_{t\uparrow\infty}\frac{1}{1+t}J_{1,t}(\pi)=\limsup_{t\uparrow\infty} E\left[\frac{1}{1+t}\sum_{s=0}^{t} c(X^\pi(s))\right] \tag{6.11}$$

for any admissible policy π in $\mathcal{P}$.

7. A proof of Theorem 6.1: The dynamic programming argument

This section discusses the optimality result of Theorem 6.1 by the dynamic programming methodology for Markov decision processes, under the additional assumption (A3). Since only Markov control policies need to be considered in solving the problems ($\tilde{\mathrm{P}}_{\beta,t}$) [16], pp. 139–143, the first step in investigating optimality for these problems is to introduce the corresponding Markovian value functions and to characterize optimality through the celebrated principle of optimality, a version of which is given below.

Specifically, for every discount factor β in $[0, 1]$, every t in N and any Markov policy μ in $\mathcal{M}$, the corresponding *expected cost-to-go* $\tilde{J}_{\beta,t}(\mu; x)$ and the *value function* $\tilde{V}_{\beta,t}(x)$ *over the horizon* $[0, t+1)$ *starting from state* x *in* N^2 are defined as

$$\tilde{J}_{\beta,t}(\mu; x) := E\left[\sum_{s=0}^{t} \beta^s \tilde{c}(U^\mu(s), X^\mu(s)) \mid X^\mu(0) = x\right] \tag{7.1}$$

and

$$\tilde{V}_{\beta,t}(x) := \sup_{\mu \in \mathcal{M}} \tilde{J}_{\beta,t}(\mu; x). \tag{7.2}$$

Owing to the time invariance of the cost (4.1) and to the statistical assumptions (A1)–(A3), the quantity $\beta^{t-n}\tilde{V}_{\beta,n}(x)$ is readily interpreted as the optimal cost-to-go over the time horizon $[t-n, t+1)$, $0 \leqq n \leqq t$, starting in state x at time $t-n$.

To simplify the notation in this section, let A_1, A_2, B_1 and B_2 denote any set of four random variables which are *generic* members of the random sequences $\{A_1(t)\}_0^\infty$, $\{A_2(t)\}_0^\infty$, $\{B_1(t)\}_0^\infty$ and $\{B_2(t)\}_0^\infty$, respectively, in the sense that the four random variables (A_1, A_2, B_1, B_2) have the same *joint* statistics as the random variables $(A_1(t), A_2(t), B_1(t), B_2(t))$ for arbitrary t, as imposed by the assumptions (A1)–(A3). Now, introduce the corresponding *one-step transitions* as the N^2-valued random variables $T^u(x)$ defined by

$$T^u(x) := (x_1 + A_1 - \mathbb{1}(x_1)uB_1, x_2 + A_2 - \mathbb{1}(x_2)(1-u)B_2) \tag{7.3}$$

for all u in $\{0, 1\}$ and x in N^2.

As customary with dynamic programming [5], [22], attention is given to the contraction mapping naturally induced by the principle of optimality on the space $\mathcal{B}(N^2)$ of bounded mappings from N^2 into R. For each discount factor β in $[0, 1]$, the mapping $\tilde{T}_\beta$ is defined formally by the relation

$$(\tilde{T}_\beta f)(x) := \max_{u \in \{0, 1\}} \{\tilde{c}(u, x) + \beta E[f(T^u(x))]\} \tag{7.4}$$

for all f in $\mathcal{B}(N^2)$ and all x in N^2. Owing to the bounds (4.4), the value functions $\tilde{V}_{\beta,t} := \{\tilde{V}_{\beta,t}(x), x \in N^2\}$ are elements of $\mathcal{B}(N^2)$ for all β in $[0, 1]$ and all t in N.

Proposition 7.1. *For the value function* (7.2), *the principle of optimality takes the following form: for all x in N^2 and all t in N,*

$$\tilde{V}_{\beta,t+1}(x) = \max_{u\in\{0,1\}} \{\tilde{c}(u,x)+\beta E[\tilde{V}_{\beta,t}(T^u(x))]\} = (\tilde{T}_\beta \tilde{V}_{\beta,t})(x), \tag{7.5}$$

with $\tilde{V}_{\beta,-1}$ denoting the identically zero mapping θ in $\mathscr{B}(N^2)$, and any Markov policy $\mu=\{\mu_s\}_0^t$ optimal for problem $(\tilde{\mathrm{P}}_{\beta,t})$ is given by

$$\mu_s(x) = \operatorname*{Arg\,max}_{u\in\{0,1\}} \{\tilde{c}(u,x)+\beta E[\tilde{V}_{\beta,t-s}(T^u(x))]\} \tag{7.6}$$

for all x in N^2 and $0 \leqq s \leqq t$.

The proof of this proposition is standard and is left as an exercise to the interested reader. Proposition 7.1 indicates that the sequence $\{\tilde{V}_{\beta,t}\}_0^\infty$ of value functions can be generated iteratively for all t in N, by the formula

$$\tilde{V}_{\beta,t} = \tilde{T}_\beta \tilde{V}_{\beta,t-1} = \tilde{T}_\beta^{(t+1)}\theta, \tag{7.7}$$

where $\tilde{T}_\beta^{(t)}$ denotes the tth iterate of $\tilde{T}_\beta$. In view of this fact, it would thus be helpful to find structural properties of the value functions which are *invariant* under the action of $\tilde{T}_\beta$, since a careful choice of such properties could lead to a solution of the equations (7.5)–(7.6) by specifying how the various maxima are achieved.

With this in mind, a set of properties is proposed that the value function $\tilde{V}_{\beta,t}$ is expected to possess if the μc-rule were indeed optimal for the problem $(\tilde{\mathrm{P}}_{\beta,t})$; a slightly stronger set of conditions is in fact introduced so as to obtain the uniqueness result in Theorem 6.1. For the sake of clarity, the discussion is given only in the case $\mu_2 c_2 \leqq \mu_1 c_1$, a condition enforced for the remainder of this section. The symmetrical case $(\mu_1 c_1 \leqq \mu_2 c_2)$ can be treated along identical lines with obvious modifications and the details are left to the interested reader.

First, two auxiliary mappings U_β^1 and U_β^2 from $\mathscr{B}(N^2)$ into itself are introduced by setting

$$(U_\beta^1 f)(x_1,x_2) := \mu_1 c_1 + \beta E[f(\max\{0, x_1+A_1-B_1\}, x_2+A_2)] \tag{7.8}$$

$$(U_\beta^2 f)(x_1,x_2) := \mu_2 c_2 + \beta E[f(x_1+A_1, \max\{0, x_2+A_2-B_2\})] \tag{7.9}$$

for all f in $\mathscr{B}(N^2)$ and all x in N^2.

Let ε, η and δ be three non-negative quantities. An element f in $\mathscr{B}(N^2)$ is then said to satisfy:

Property (P_1^η) *if for all x_1 and x_2 in N*

$$\beta f(x_1+1, x_2) + \eta \leqq \beta f(x_1,x_2) + c_1; \tag{7.10}$$

Property (P_2^δ) *if for all* x_1 *and* x_2 *in* N

$$\beta f(x_1, x_2+1)+\delta \leqq \beta f(x_1, x_2)+c_2; \tag{7.11}$$

Property (P_3^ε) *if for all* x_1 *and* $x_2 \geqq 1$ *in* N

$$(U_\beta^2 f)(x_1, x_2)+\varepsilon \leqq (U_\beta^1 f)(x_1, x_2). \tag{7.12}$$

The conditions (7.10)–(7.12) have simple and natural interpretations when $\eta=\delta=\varepsilon=0$, if f plays the role of the value function ([3], Proposition 6.1): the conditions (7.10) and (7.11) induce the optimality of the μc-rule when exactly one of the queues is empty, whereas condition (7.12) expresses this same fact when both queues are non-empty since for every f in $\mathscr{B}(N^2)$,

$$(\tilde{T}_\beta f)(x)=\max\{(U_\beta^1 f)(x), (U_\beta^2 f)(x)\} \tag{7.13}$$

for all x in N^2 with $x_1 \geqq 1$ and $x_2 \geqq 1$.

The next result describes the propagation of these structural conditions under the action of the mapping $\tilde{T}_\beta$.

Theorem 7.2. *Under the foregoing assumptions* (A1)–(A3) *with* $\mu_2 c_2 \leqq \mu_1 c_1$, *if an element* f *in* $\mathscr{B}(N^2)$ *satisfies the properties* (P_1^0), (P_2^0) *and* (P_3^0), *then* $g_\beta := \tilde{T}_\beta f$ *satisfies the properties* (P_1^η), (P_2^δ) *and* (P_3^ε) *with* $\eta := (1-\beta)c_1$, $\delta := (1-\beta)c_2$ *and* $\varepsilon := (1-\beta)[\mu_1 c_1 - \mu_2 c_2]$.

A proof of this key invariance result is available in [3], Appendix. The dynamic programming argument for establishing the optimality of the μc-rule for problem ($\tilde{P}_{\beta,t}$) can now be given, as follows.

Proof of Theorem 6.1. Since $\mu_2 c_2 \leqq \mu_1 c_1$, the identically zero mapping θ obviously satisfies (P_1^0), (P_2^0) and (P_3^0). Theorem 7.2 thus implies via (7.7) that $\tilde{V}_{\beta,0}$ satisfies (P_1^η), (P_2^δ) and (P_3^ε) with η, δ and ε as given there, whence (P_1^0), (P_2^0) and (P_3^0) are satisfied. From Theorem 7.2 and the relation (7.7), it follows that $V_{\beta,1}$ also enjoys the properties (P_1^η), (P_2^δ) and (P_3^ε) and thus necessarily the properties (P_1^0), (P_2^0) and (P_3^0). Inductive use of this argument easily establishes that $\tilde{V}_{\beta,t}$ satisfies (P_1^η), (P_2^δ) and (P_3^ε) for *all* t in N, i.e., for all x_1 and $x_2 \geqq 1$ in N,

$$\beta \tilde{V}_{\beta,t}(x_1+1, x_2)+\eta \leqq \beta \tilde{V}_{\beta,t}(x_1, x_2)+c_1 \tag{7.14}$$

$$\beta \tilde{V}_{\beta,t}(x_1, x_2+1)+\delta \leqq \beta \tilde{V}_{\beta,t}(x_1, x_2)+c_2 \tag{7.15}$$

and

$$(U_\beta^2 \tilde{V}_{\beta,t})(x_1, x_2)+\varepsilon \leqq (U_\beta^1 \tilde{V}_{\beta,t})(x_1, x_2). \tag{7.16}$$

Now, direct inspection of (7.4) shows that for all t in N,

$$(\tilde{T}_\beta \tilde{V}_{\beta,t})(x_1, 0)=\max\{(U_\beta^1 \tilde{V}_{\beta,t})(x_1, 0), \beta E[\tilde{V}_{\beta,t}(x_1+A_1, A_2)]\} \tag{7.17}$$

whenever $x_1 \geqq 1$ in N. Observe then from (7.8) that

$$(U_\beta^1 \tilde{V}_{\beta,t})(x_1, 0) - \beta E[\tilde{V}_{\beta,t}(x_1 + A_1, A_2)]$$

$$= \mu_1 c_1 + \mu_1 \beta E[\tilde{V}_{\beta,t}(x_1 + A_1 - 1, A_2)] - \mu_1 \beta E[\tilde{V}_{\beta,t}(x_1 + A_1, A_2)] \tag{7.18}$$

$$= \mu_1 E[[c_1 + \beta \tilde{V}_{\beta,t}(x_1 + A_1 - 1, A_2)] - \beta \tilde{V}_{\beta,t}(x_1 + A_1, A_2)] \tag{7.19}$$

by a simple preconditioning argument making use of the mutual independence of the random variables B_1 and $\{A_1, A_2\}$. The inequality (7.14) immediately implies that

$$(U_\beta^1 \tilde{V}_{\beta,t})(x_1, 0) - \beta E[\tilde{V}_{\beta,t}(x_1 + A_1, A_2)] \geqq \mu_1 \eta \tag{7.20}$$

for all $x_1 \geqq 1$ in N and consequently,

$$\beta E[\tilde{V}_{\beta,t}(x_1 + A_1, A_2)] < (U_\beta^1 \tilde{V}_{\beta,t})(x_1, 0) = (\tilde{T}_\beta \tilde{V}_{\beta,t})(x_1, 0) \tag{7.21}$$

as a result of (7.17) since c_1 is assumed strictly positive.

Identical arguments, using (7.15) this time, will show that

$$(U_\beta^2 \tilde{V}_{\beta,t})(0, x_2) - \beta E[\tilde{V}_{\beta,t}(A_1, x_2 + A_2)] \geqq \mu_2 \delta \tag{7.22}$$

for all $x_2 \geqq 1$ in N, with

$$\beta E[\tilde{V}_{\beta,t}(A_1, x_2 + A_2)] < (U_\beta^2 \tilde{V}_{\beta,t})(0, x_2) = (\tilde{T}_\beta \tilde{V}_{\beta,t})(0, x_2). \tag{7.23}$$

If the assumption $\mu_2 c_2 < \mu_1 c_1$ holds and β is in $[0, 1)$, then the quantity ε is *strictly* positive and the inequality (7.16) easily implies the inequality

$$(U_\beta^2 \tilde{V}_{\beta,t})(x_1, x_2) + \varepsilon < (U_\beta^1 \tilde{V}_{\beta,t})(x_1, x_2) = (\tilde{T}_\beta \tilde{V}_{\beta,t})(x_1, x_2) \tag{7.24}$$

for all x_1 and $x_2 \geqq 1$ in N, owing to the remark (7.13). Substitution of (7.21), (7.23) and (7.24) into the dynamic programming equation (7.5) readily implies the optimality of the μc-rule for the problem $(\tilde{\mathrm{P}}_{\beta,t})$. Because of the strict inequalities appearing in (7.21), (7.23) and (7.24), it follows from (7.6) that the optimal policy is essentially unique since it is unequivocally determined on $N^2 \backslash \{(0, 0)\}$.

If on the other hand, the condition $\mu_1 c_1 = \mu_2 c_2$ holds with β in $[0, 1)$, then (7.24) has to be replaced for all t in N by

$$(U_\beta^2 \tilde{V}_{\beta,t})(x_1, x_2) = (\tilde{T}_\beta \tilde{V}_{\beta,t})(x_1, x_2) = (U_\beta^1 \tilde{V}_{\beta,t})(x_1, x_2) \tag{7.25}$$

with x_1 and $x_2 \geqq 1$ in N. The reason for this modification is found in the proof of Theorem 6.3 as given in the technical report [3], Appendix, Equations (A.32) and (A.46), to which the reader is referred for details [3], Theorem 7.1. From (7.21), (7.23) and (7.25), it is now easy to establish via Proposition 7.1 that any non-idling policy, and the μc-rule μ^* in particular, is optimal for all problems $(\tilde{\mathrm{P}}_{\beta,t})$.

8. A proof of Theorem 6.1: The comparison arguments

The discussion of Theorem 6.1 given in this section is valid under the basic model assumptions (A1)–(A2) and uses comparison arguments often invoked to solve other scheduling and priority assignment problems. Typical examples can be found in the works of Meilijson and Yechiali [18], and of Pinedo [19].

Several auxiliary concepts are now introduced to facilitate the presentation of the arguments of this section. For any admissible policy π in $\mathcal{P}$, the policy $\pi^{(s)} := \{\pi_r^{(s)}\}_0^\infty$, with s ranging in N, is defined as the *concatenation at time s* of the policy π and of the μc-rule μ^*, where

$$\pi_r^{(s)} := \begin{cases} \pi_r & \text{for} \quad 0 \leqq r < s \\ \mu^* & \text{for} \quad s \leqq r \end{cases} \tag{8.1}$$

for all r in N. Clearly, (8.1) defines an admissible policy in $\mathcal{P}$.

Now, consider problem $(\tilde{\mathrm{P}}_{\beta,t})$ for some β in $(0, 1)$ and t in N. For every s in N with $0 \leqq s \leqq t$, define the quantities

$$\tilde{J}_{\beta,t}(s; \pi) := E\left[\sum_{r=s}^{t} \beta^r \tilde{c}(U^\pi(r), X^\pi(r))\right] \tag{8.2}$$

for all admissible policies π in $\mathcal{P}$. The statement that the *μc-rule μ^* is optimal from time s onward in problem* $(\tilde{\mathrm{P}}_{\beta,t})$ for some $0 \leqq s \leqq t$, is used below and has to be understood as saying that

$$\tilde{J}_{\beta,t}(s; \pi) \leqq \tilde{J}_{\beta,t}(s; \pi^{(s)}) \tag{8.3}$$

for all *non*-idling policies π in $\mathcal{P}$.

Observe that the μc-rule μ^* is always optimal from time t onward in problem $(\tilde{\mathrm{P}}_{\beta,t})$. The aim of the next proposition is to show that this property propagates in time downward from, t.

Theorem 8.1. *Consider problem* $(\tilde{\mathrm{P}}_{\beta,t})$ *for some* β *in* $(0, 1)$ *and* t *in* N. *The μc-rule μ^* is optimal from time t onward. If the μc-rule is optimal from time* $s+1$ *onward with* $0 \leqq s \leqq t$, *then it is necessarily optimal from time s onward.*

Proof. The first part of the proposition is obvious by virtue of the myopic optimality of the μc-rule discussed in Section 4.

To establish the second part, there is no loss of generality in assuming $\mu_1 c_1 \leqq \mu_2 c_2$, a condition enforced throughout the proof. The linearity of the expression (8.2) and the assumed optimality from time $s+1$ onward of the μc-rule readily imply that *only* non-idling control policies π in $\mathcal{P}$ with $\pi_r = \mu^*$ for $s+1 \leqq r \leqq t$ need be considered in establishing (8.3).

Now, for such an admissible policy π in $\mathcal{P}$, define the event $E^\pi(s)$ to be

$$E^\pi(s) := [X_1^\pi(s) \neq 0, X_2^\pi(s) \neq 0, U^\pi(s) = 0], \tag{8.4}$$

and observe that only on this event does the policy π at time s act differently from the μc-rule μ^*. In particular, the non-idling nature of π and its behavior like the μc-rule from $s+1$ onward cause the second and first queues to be given service attention at times s and $s+1$, respectively, on the event $E^\pi(s)$. Specifically, the dynamics (2.1) yield

$$X_1^\pi(s+1)=X_1^\pi(s)+A_1(s),\ X_2^\pi(s+1)=X_2^\pi(s)+A_2(s)-B_2(s) \tag{8.5}$$

with a contribution $\mu_2 c_2 \beta^s$ to the cost $\tilde{J}_{\beta,t}(\pi)$, and

$$X_1^\pi(s+2)=X_1^\pi(s)+A_1(s)+A_1(s+1)-B_1(s+1) \tag{8.6}$$

$$X_2^\pi(s+2)=X_2^\pi(s)+A_2(s)+A_2(s+1)-B_2(s) \tag{8.7}$$

with a contribution of $\mu_1 c_1 \beta^{s+1}$ to this cost.

The basic idea behind the proof is to construct a new policy $\tilde{\pi}$ that reverses the order of service *on the event* $E^\pi(s)$ such that $\tilde{J}_{\beta,t}(\pi)\leqq\tilde{J}_{\beta,t}(\tilde{\pi})$. To this end, define the admissible policy $\tilde{\pi}:=\{\tilde{\pi}_r\}_0^\infty$ in $\mathcal{P}$ as follows. For all $r\neq s$ and $r\neq s+1$ in N, pose $\tilde{\pi}_r:=\pi_r$ and define $\tilde{\pi}_s$ and $\tilde{\pi}_{s+1}$ such that on the event $\Omega\setminus E^\pi(s)$, $U^{\tilde{\pi}}(s)=U^\pi(s)$ and $U^{\tilde{\pi}}(s+1)=U^\pi(s+1)$ while on the event $E^\pi(s)$, $U^{\tilde{\pi}}(s)=1$ and $U^{\tilde{\pi}}(s+1)=0$. It is easy to see that $\tilde{\pi}$ is non-idling and operates like the μc-rule at time s. Now on the event $E^\pi(s)$, use of $\tilde{\pi}$ gives

$$X^{\tilde{\pi}}(s+1)=X_1^\pi(s)+A_1(s)-B_1(s),\ X_2^{\tilde{\pi}}(s+1)=X_2^\pi(s)+A_2(s) \tag{8.8}$$ 1

with a contribution $\mu_1 c_1 \beta^s$ to the cost $\tilde{J}_{\beta,t}(\tilde{\pi})$, and 3

$$X^{\tilde{\pi}}(s+2)=X_1^\pi(s)+A_1(s)+A_1(s+1)-B_1(s) \tag{8.9}$$

$$X^{\tilde{\pi}}(s+2)=X_2^\pi(s)+A_2(s)+A_2(s+1)-B_2(s+1) \tag{8.10}$$

with a contribution $\mu_2 c_2 \beta^{s+1}$ to the cost.

The random variables $X^{\tilde{\pi}}(s+2)$ and $X^\pi(s+2)$ agree on the event $\Omega\setminus E^\pi(s)$, but it is clear by construction that $X^{\tilde{\pi}}(s+2)\neq X^\pi(s+2)$ in general on $E^\pi(s)$. However, under the assumptions made, the random variables $X^{\tilde{\pi}}(s+2)$ and $X^\pi(s+2)$ have the *same distribution*. As a result, since the policies π and $\tilde{\pi}$ both operate like the μc-rule from time $s+2$ onward, it follows by direct inspection, with a simple preconditioning argument, that

$$\tilde{J}_{\beta,t}(s;\tilde{\pi})-\tilde{J}_{\beta,t}(s;\pi)=\beta^s[\mu_1c_1-\mu_2c_2](1-\beta)P[E^\pi(s)]\geqq 0. \tag{8.11}$$

On the other hand, the assumed optimality of the μc-rule from time $s+1$ onward implies

$$\tilde{J}_{\beta,t}(s;\tilde{\pi})=E[\beta^s\tilde{c}(U^{\tilde{\pi}}(s),X^\pi(s))]+\tilde{J}_{\beta,t}(s+1;\tilde{\pi}) \tag{8.12}$$

$$\leqq E[\beta^s\tilde{c}(U^{\tilde{\pi}}(s),X^\pi(s))]+\tilde{J}_{\beta,t}(s+1;\tilde{\pi}^{(s+1)}) \tag{8.13}$$

since $\tilde{\pi}$ is *non*-idling. But by construction the policy $\tilde{\pi}$ operates like the

μc-rule at time s, whence $\tilde{\pi}^{(s)} := \tilde{\pi}^{(s+1)}$, and (8.11) combines with (8.13) to give

$$\tilde{J}_{\beta,t}(s;\pi) \leqq \tilde{J}_{\beta,t}(s;\tilde{\pi}) \leqq \tilde{J}_{\beta,t}(s;\tilde{\pi}^{(s)}). \tag{8.14}$$

Obviously $\pi^{(s)} = \tilde{\pi}^{(s)}$ and (8.14) thus shows that

$$\tilde{J}_{\beta,t}(s;\pi) \leqq \tilde{J}_{\beta,t}(s;\pi^{(s)}) \tag{8.15}$$

for all non-idling policies π in $\mathcal{P}$, i.e., the μc-rule μ^* is indeed optimal from time s onward.

The next result is now a direct consequence of the induction argument implicit in Theorem 8.1, when carried out to completion.

Theorem 8.2. *Consider problem* $(\mathrm{P}_{\beta,t})$ *for some* β *in* $(0, 1)$ *and* t *in* N. *The* μc-*rule* μ^* *minimizes* (4.2) *over the class of all non-idling policies in* $\mathcal{P}$.

Idling does not pay as is shown by the following result.

Theorem 8.3. *Consider problem* $(\tilde{\mathrm{P}}_{\beta,t})$ *for some* β *in* $(0, 1)$ *and* t *in* N. *For any admissible policy* π *in* $\mathcal{P}$, *there exists a non-idling policy* $\hat{\pi}$ *in* $\mathcal{P}$ *such that*

$$\tilde{J}_{\beta,t}(\pi) \leqq \tilde{J}_{\beta,t}(\hat{\pi}) \tag{8.16}$$

Proof. The argument is similar to the one used in Theorem 8.1 and will only be sketched here for the sake of brevity. The basic idea was originally used in [4] and amounts to embedding the system of two competing queues into one of *three* competing queues. This third (somewhat fictitious) queue operates in discrete time, initially contains a *single* customer and receives in each time slot exactly *one* customer. The service requirements for customers in the third queue are assumed *infinite* (i.e., geometrically distributed with parameter 0) and completion of service thus *never* occurs in this queue which *never empties.* In fact, the corresponding queue size sequence $\{X_3(t)\}_0^\infty$ is independent of the service allocation policy and given by the recursion

$$X_3(t+1) = X_3(t) + 1 \tag{8.17}$$

for all t in N, with $X_3(0) = 1$. Finally, assume that giving service attention to the third queue incurs *no* cost.

To describe the embedding, let π be any admissible policy π in $\mathcal{P}$ for the two-queues system. In the context of the three-queues system, the server operates under π by giving service attention to the original queues as before, but to the third queue whenever the former are *empty* or the policy π is *idling* in the original system. Observe that under this interpretation, *any* policy π in $\mathcal{P}$ is translated into a *non*-idling policy for the three-queues systems, with the same incurred cost $\tilde{J}_{\beta,t}(\pi)$.

The discussion can now proceed, as only the case of an idling policy π in $\mathcal{P}$

needs to be considered for otherwise (8.16) is trivially satisfied with $\hat{\pi} = \pi$. So assume that the policy π is idling at some time s, with $0 \leqq s \leqq t$. The idea is to construct a policy $\hat{\pi}^{(s)}$ for the three-queues systems that incurs a larger cost than $\tilde{J}_{\beta,t}(\pi)$. The policy $\hat{\pi}^{(s)} = \{\hat{\pi}_r^{(s)}\}_0^\infty$ is defined to agree with π at all times but $r = s$ and $r = s + 1$ *when* idling. Moreover, when idling at time s, $\hat{\pi}^{(s)}$ is defined at time s *not* to idle for the two-queues systems and at time $s + 1$ to give service attention to the third queue. It is now a simple matter, via the arguments given in Theorem 8.1, to check that the incurred cost for $\hat{\pi}$ is larger than for π. Again, an easy induction argument will yield (8.16) since the μc-rule can always be used at time t to improve the cost.

A second proof of Theorem 6.1 can finally be given.

Proof of Theorem 6.1. Theorems 8.2 and 8.3 imply that for every admissible policy π in $\mathcal{P}$,

(8.18) $$\tilde{J}_{\beta,t}(\pi) \leqq \tilde{J}_{\beta,t}(\mu^*)$$

with β in $(0, 1)$ and t in N.

This establishes Theorem 6.1 for $\beta \neq 1$, with the uniqueness result following readily from the relation (8.11) obtained in the proof of Theorem 8.1. The optimality result is obtained for the case $\beta = 1$ by use of the monotone convergence theorem in (8.18) with β going to 1 from below.

References

[1] BARAS, J. S. AND DORSEY, A. J. (1981) Stochastic control of two partially observed competing queues. *IEEE Trans. Automatic Control* **AC-26,** 1095–1106.

[2] BARAS, J. S. AND DORSEY, A. J. (1983) Adaptive control of two competing queues. In *Proc. 2nd IEEE Annual Joint Conf. Infocom*, April 1983, San Diego, CA, 427–435.

[3] BARAS, J. S., DORSEY, A. J. AND MAKOWSKI, A. M. (1983) Two competing queues with linear costs: the μc-rule is often optimal. Technical Research Report SRR-83-1, Electrical Engineering Department, University of Maryland, College Park.

[4] BARAS, J. S., MA, D.-J. AND MAKOWSKI, A. M. (1983) Competing queues with geometric service requirements and linear costs: the μc-rule is always optimal. Technical Research Report SRR-83-9, Electrical Engineering Department, University of Maryland, College Park.

[5] BERTSEKAS, D. P. (1976) *Dynamic Programming and Stochastic Control.* Academic Press, New York.

[6] BUYUKKOC, C., VARAIYA, P. AND WALRAND, J. (1985) The $c\mu$-rule revisited. *Adv. Appl. Prob.* **17,** 234–235.

[7] COX, D. R. AND SMITH, W. L. (1961) *Queues.* Chapman and Hall, London.

[8] DORSEY, A. J. (1983) Adaptive Control of a Simple Queueing System. Ph.D. Thesis, Electrical Engineering Department, University of Maryland, College Park.

[9] EPHREMIDES, A., VARAIYA, P. AND WALRAND, J. (1980) A simple dynamic routing problem. *IEEE Trans. Automatic Control* **AC-25,** 690–693.

[10] HAJEK, B. (1982) Optimal control of two interactive service stations. In *Proc. 21st IEEE Conf. Decision and Control*, December 1982, Orlando, FL, 840–845.

[11] HARRISON, J. M. (1975) A priority queue with discounted linear costs. *Operat. Res.* **23,** 260–269.

[12] HARRISON, J. M. (1975) Dynamic scheduling of a multiclass queue: discount optimality. *Operat. Res.* **23,** 270–282.

[13] JAISWAL, N. K. (1968) *Priority Queues.* Academic Press, New York.

[14] KAKALIK, J. S. (1969) Optimal dynamic operating policies for a service facility. Technical Report No. 47, Operations Research Center, MIT.

[15] KLEINROCK, L. (1976) *Queueing Systems, Vol. 2: Computer Applications.* Wiley, New York.

[16] KUSHNER, H. (1971) *Introduction to Stochastic Control.* Holt, Rinehart and Winston, New York.

[17] LIPPMAN, S. A. (1973) Semi-Markov decision processes with unbounded rewards. *Management Sci.* **19,** 717–731.

[18] MEILIJSON, I. AND YECHIALI, U. (1977) On optimal right-of-way policies at a single server station when insertion of idle time is permitted. *Stoch. Proc. Appl.* **5,** 25–32.

[19] PINEDO, M. (1983) Stochastic scheduling with release dates and due dates. *Operat. Res.* **31,** 559–572.

[20] REISER, M. (1982) Performance evaluation of data communication systems. *Proc. IEEE* **70** (12), 171–196.

[21] ROSBERG, Z., VARAIYA, P. AND WALRAND, J. (1982) Optimal control of service in tandem queues. *IEEE Trans. Automatic Control* **AC-27,** 600–610.

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

[23] RYKOV, V. V. AND LEMBERT, E. YE. (1967) Optimal dynamic priorities in single-line queueing systems. *Engineering Cybernetics* **5**(1), 21–30.

[24] SCHOLL, M. (1976) Multiplexing Techniques for Data Transmission over Packet-Switched Radio Systems. Ph.D. Thesis, Computer Science Department, UCLA.

[25] SHACHAM, N. (1983) A protocol for preferred access in packet-switching radio networks. *IEEE Trans. Communications* **COM-31,** 253–264.

[26] SIDI, M. AND SEGALL, A. (1983) Structured priority queueing systems with applications to packet-radio networks. *Performance Evaluation* **3,** 265–275.

[27] VARAIYA, P., WALRAND, J. AND BUYUKKOC, C. (1985) Extensions of the multi-armed bandit problem. *IEEE Trans. Automatic Control.* To appear.