MARYLAND

TECHNICAL RESEARCH REPORT

SRC TR 87-169

Optimality Results for a Simple Flow Control Problem

by

D. Ma and A. M. Makowski

SYSTEMS RESEARCH CENTER

UNIVERSITY OF MARYLAND

COLLEGE PARK, MARYLAND 20742

SRC Library
PLEASE DO NOT REMOVE
Thank You

Contributed paper to the 26th IEEE Conference on Decision and Control,
Los Angeles, California, December 1987

OPTIMALITY RESULTS FOR A SIMPLE FLOW CONTROL PROBLEM

by

Dye-Jyun Ma[1] and Armand M. Makowski[2]

Electrical Engineering Department and Systems Research Center
University of Maryland, College Park, Maryland 20742

ABSTRACT

This paper presents a problem of optimal flow control for discrete-time $M|M|1$ queues, where the decision-maker seeks to maximize the throughput subject to a bound on the average queue size. The problem is cast as a constrained Markov decision process and solved via Lagrangian arguments. The optimal strategy is shown to be a threshold policy which saturates the constraint. The method of analysis proceeds through the discounted version of the Lagrangian problems whose value functions are shown to be integer-concave. Dynamic Programming and stochastic comparison ideas constitute the main ingredients of the solution.

[1] The work of this author was supported partially through NSF Grant ECS-83-51836 and partially through NSF Grant NSFD CDR-85-00108.

[2] The work of this author was supported partially through ONR Grant N00014-84-K-0614 and partially through a grant from AT&T Bell Laboratories.

1. Introduction

Consider a *synchronous* communication channel between two entities, a transmitter and a receiver, which are both equipped with buffers of infinite capacity. Information is formatted in packets and time is slotted so that the duration of a time slot coincides with the transmission time of a packet. Packet transmissions are initiated at the beginning of a slot. The channel is assumed noisy in that a packet transmission may not be successful with probability $1 - \mu$ in which case retransmission is attempted in the next slot. This scenario is repeated until successful transmission occurs, at which time the packet is deleted from the transmitter's buffer. Packets arrive at the transmitter one at a time according to a Bernoulli sequence with rate λ, i.e., λ is the probability that a packet will arrive in any time slot, and the transmission failures are assumed independent from slot to slot, and independent of the arrival process.

As this system may experience congestion, it is desirable to take certain actions in order to guarantee an expected performance level. One possible approach consists in restricting access to the communication system, i.e., new packets which are about to enter the transmitter's buffer may be denied entrance on the basis of information reflecting system congestion. This is often referred to as *flow control* and should be done on the basis of some performance criterion [3]. Here, an approach similar to the one of Lazar [5] is adopted in that a flow control strategy is sought that maximizes the channel throughput subject to a constraint on the long-run average number of packets in the system.

Under the statistical assumptions given earlier, the uncontrolled system can be modelled as a discrete-time $M|M|1$ queue, and the problem of finding good flow control schemes can be cast as a Markov decision process (MDP) with *constraint.* Analysis shows that this constrained flow control problem admits a solution within the class of *threshold* policies which are parametrized by an integer-valued threshold level L $(= 0, 1, \ldots)$ and an acceptance probability η $(0 \leq \eta \leq 1)$. A threshold policy (L, η) has a simple structure in that at the beginning of each time slot, a new packet is accepted (resp. rejected) if the buffer content is strictly below L (resp. above L), while if there are *exactly* L packets in the buffer, this new packet is accepted (resp. rejected) with probability η (resp. $1 - \eta$).

Throughout the years, various problems of flow control (or control of arrivals) have been studied in the context of simple queueing systems, and a good discussion of such work can be found in the survey paper of Stidham [13]. It should be pointed out that previous papers

dealt exclusively with continuous-time models, and that the *concavity* of the value function for the single node situation could be obtained fairly easily through standard arguments. Here, establishing the concavity of the value functions of interest turns out to be a much more cumbersome task and constitutes the key technical contribution of this paper. This difficulty can possibly be explained by the fact that multiple transitions can be realized, a phenomenon which precludes use of the homogeneization technique for the discrete-time situation [4]. *Dynamic Programming* and *stochastic comparison* ideas constitute the main technical ingredients of the solution.

Amongst the models covered in Stidham's survey paper, only the work of Lazar [5] formulates the problem as a *constrained* problem. However, the approach taken here is different from the one used by Lazar in that he considers a *closed* system from the onset with a *fixed* number of packets, while the work discussed here assumes an open system. Of course, both approaches lead to similar results, as expected.

The paper is organized as follows. The model is described in Section 2 and the constrained flow control problem is posed in Section 3, where the optimality results are summarized and the necessary Lagrangian are briefly outlined. Section 4 is devoted to the study of the discounted version of the Lagrangian problems, for which threshold policies are identified to be optimal. The properties of threshold policies are discussed in Section 5, and used in Section 6 to find the solution to the long-run version of the Lagrangian problems. A useful comparison result is given in the Appendix.

A word on the notation: The set of real numbers is denoted by $\mathbb{R}$, while $\mathbb{N}$ denotes the set of all non-negative integers. For any x in $\mathbb{R}$, it is convenient to pose $\bar{x} = 1 - x$, and the characteristic function of any set E is denoted simply by $1[E]$.

2. Model

In order to formally define a flow control model for discrete-time $M|M|1$ systems, start with the sample space $\Omega := \mathbb{N} \times \left(\{0,1\}^3\right)^\infty$, and recursively define the information spaces $\{\mathbb{H}_n\}_0^\infty$ by $\mathbb{H}_0 := \mathbb{N}$ and $\mathbb{H}_{n+1} := \mathbb{H}_n \times \{0,1\}^3$ for all $n = 0, 1, \ldots$.

An element ω of Ω is viewed as a sequence $(x, \omega_0, \omega_1, \ldots)$ with x in $\mathbb{N}$ and ω_n in $\{0,1\}^3$ for all $n = 0, 1, \ldots$. Each block component ω_n is written in the form (u_n, a_n, b_n), with u_n, a_n and b_n being all elements in $\{0,1\}$. An element h_n in $\mathbb{H}_n$ is uniquely associated with the

sample ω by $h_n := (x, \omega_0, \ldots, \omega_{n-1})$ with $h_0 := x$.

Let the sample $\omega = (x, \omega_0, \omega_1, \ldots)$ be realized. The initial queue size is set at x. During each time slot $[n, n+1)$, $a_n = 1$ (resp. a_n=0) indicates that a customer (resp. no customer) has arrived into the queue, $b_n = 1$ (resp. b_n=0) encodes a successful (resp. unsuccessful) completion of service in that slot, whereas control action u_n is selected at the beginning of the time slot $[n, n+1)$, with u_n=1 (resp. u_n=0) for admitting (resp. rejecting) the incoming customer during that slot. If x_n denotes the queue size at the beginning of the slot $[n, n+1)$, its successive values are determined through the recursion

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

with $x_0 := x$.

The coordinate mappings $\Xi, \{U(n)\}_0^\infty, \{A(n)\}_0^\infty$ and $\{B(n)\}_0^\infty$ are defined on the sample space Ω by posing $\Xi(\omega) := x$, $U(n, \omega) := u_n$, $A(n, \omega) := a_n$ and $B(n, \omega) := b_n$ for $n = 0, 1, \ldots$ and ω in Ω, with the information mappings $\{H(n)\}_0^\infty$ being given by $H(n, \omega) := (x, \omega_0, \omega_1, \ldots, \omega_{n-1}) := h_n$.

For each $n = 0, 1, \ldots$, let $\mathbb{F}_n$ be the σ-field generated by the mapping $H(n)$ on the sample space Ω. Clearly, $\mathbb{F}_n \subset \mathbb{F}_{n+1}$, and with standard notation, $\mathbb{F} := \vee_{n=0}^\infty \mathbb{F}_n$ is simply the σ-field on Ω generated by the mappings Ξ and $\{U(n), A(n), B(n)\}_0^\infty$. Thus, on the space $(\Omega, \mathbb{F})$, the mappings $\Xi, \{U(n)\}_0^\infty$, $\{A(n)\}_0^\infty$, $\{B(n)\}_0^\infty$ and $\{H(n)\}_0^\infty$ are all random variables (RV) taking values in $\mathbb{N}$, $\{0,1\}$, $\{0,1\}$, $\{0,1\}$ and $\mathbb{H}_n$, respectively. The queue sizes $\{X(n)\}_0^\infty$ are $\mathbb{N}$-valued RV's recursively defined by

$$X(n+1) = X(n) + U(n)A(n) - 1[X(n) \neq 0] B(n) \qquad n = 0, 1, \ldots (2.2)$$

with $X(0) := \Xi$. Each RV $X(n)$ is clearly $\mathbb{F}_n$-measurable.

Since randomization is allowed, an admissible policy π is defined as any collection $\{\pi_n\}_0^\infty$ of mappings $\pi_n : \mathbb{H}_n \to [0, 1]$, with the interpretation that the potential arrival during the slot $[n, n+1)$ is admitted (resp. rejected) with probability $\pi_n(h_n)$ (resp. $1 - \pi_n(h_n)$) whenever the information h_n is available to the decision-maker. In the sequel, denote the collection of all such admissible policies by $\mathcal{P}$.

Let $q(\bullet)$ be a probability distribution on $\mathbb{N}$, and let λ and μ be fixed constants in $(0, 1)$. Given any policy π in $\mathcal{P}$, there exists an unique probability measure P^π on $\mathbb{F}$, with expectation operator E^π, satisfying the requirements (R1)-(R3) below, where

(R1): For all x in $\mathbb{N}$,

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

(R2): For all a and b in $\{0,1\}$,

$$P^\pi[A(n) = a, B(n) = b | \mathbb{F}_n \vee \sigma\{U(n)\}] := (a\lambda + \bar{a}\bar{\lambda})(b\mu + \bar{b}\bar{\mu}) \qquad n = 0, 1, \ldots$$

(R3):

$$P^\pi[U(n) = 1 | \mathbb{F}_n] := \pi_n(H(n)). \qquad n = 0, 1, \ldots$$

This notation is specialized to P_x^π and E_x^π, respectively, when $q(\bullet)$ is the point mass distribution at x in $\mathbb{N}$; it is plain that $P^\pi[A|X(0) = x] = P_x^\pi[A]$ for every A in $\mathbb{F}$. It readily follows from (R1)-(R3) that under each probability measure P^π,

(P1): The $\mathbb{N}$-valued RV Ξ is independent of the sequences of RV's $\{A(n)\}_0^\infty$ and $\{B(n)\}_0^\infty$,

(P2): The sequences $\{A(n)\}_0^\infty$ and $\{B(n)\}_0^\infty$ of $\{0,1\}$-valued RV's are mutually independent Bernoulli sequences with parameters λ and μ, respectively, and

(P3): The transition probabilities take the form

$$P^\pi[X(n+1) = y | \mathbb{F}_n] = p[X(n), y; \pi_n(H(n))] \qquad n = 0, 1, \ldots \quad (2.3)$$

where

$$p[x, y; \eta] := \eta Q^1(x, y) + \bar{\eta} Q^0(x, y) \quad (2.4)$$

with

$$Q^i(x, y) := P^\pi[x + iA(n) - 1(x \neq 0)B(n) = y], \quad i = 0, 1 \quad (2.5)$$

for all x and y in $\mathbb{N}$, and all η in [0,1].

The right-hand sides of (2.5) depend neither on n nor on the policy π owing to the assumptions (R1)-(R3) made earlier. Throughout this paper, the RV Ξ is assumed to have finite mean, i.e., $\sum_{x=0}^\infty xq(x) < \infty$.

Several subclasses of policies in $\mathcal{P}$ will be of interest in the sequel.

A policy π in $\mathcal{P}$ is said to be a *Markov* policy if there exists a family $\{g_n\}_0^\infty$ of mappings $g_n: \mathbb{N} \to [0, 1]$ such that $\pi_n(H(n)) = g_n(X(n))$ $P^\pi - a.s.$ for all $n = 0, 1, \ldots$ In the event $g_n = g$ for all $n = 0, 1, \ldots$, the Markov policy π is called *stationary* and can be identified with the mapping g itself.

A policy π in $\mathcal{P}$ is said to be a *pure* (or *non-randomized*) policy if there exists a family $\{f_n\}_0^\infty$ of mappings $f_n: \mathbb{H}_n \to \{0,1\}$ such that $\pi_n(H(n)) = f_n(H(n))$ $P^\pi - a.s.$ for all $n = 0, 1, \ldots$ A *pure Markov stationary* policy π is thus fully characterized by a single mapping $f: \mathbb{N} \to \{0,1\}$.

A stationary policy g is said to be of *threshold* type if there exists a pair (L, η), with L in $\mathbb{N}$ and η in $[0,1]$, such that

$$g(x) = \begin{cases} 1 & \text{if } x < L; \\ \eta & \text{if } x = L; \\ 0 & \text{if } x > L. \end{cases} \tag{2.6}$$

Such a *threshold* policy is denoted by (L, η), and by extension, the Markov stationary policy g that admits every single customer, i.e., $g(x) = 1$ for all x in $\mathbb{N}$, is conveniently denoted by $(\infty, 1)$.

3. The optimal control problems

For any admissible policy π in $\mathcal{P}$, pose

$$T(\pi) := \liminf_n \frac{1}{n+1} E^\pi \sum_{t=0}^{n} \mu 1[X(t) \neq 0] \tag{3.1}$$

and

$$N(\pi) := \limsup_n \frac{1}{n+1} E^\pi \sum_{t=0}^{n} X(t). \tag{3.2}$$

These quantities $T(\pi)$ and $N(\pi)$ are readily interpreted as the *throughput* and the long-run average *queue size*, respectively, when the policy π is used.

Given $V > 0$, the problem (P_V) of interest is the *constrained* optimization problem

$$(P_V): \quad \text{Maximize } T(\pi) \text{ over } \mathcal{P}_V$$

where

$$\mathcal{P}_V := \{\pi \in \mathcal{P} : N(\pi) \leq V\}. \tag{3.3}$$

If the constraint is satisfied when admitting every single customer, i.e., $N((\infty, 1)) \leq V$, then $\mathcal{P}_V = \mathcal{P}$ and the constrained optimization problem (P_V) reduces to an unconstrained optimization problem with *trivial* solution $(\infty, 1)$ as shown in [6, Thm. 3.1] by simple stochastic

comparison arguments. On the other hand, if $N((\infty,1)) > V$, the solution to the constrained problem (P_V) is no longer trivial and it is the main objective of this paper to identify its structure. The main result is summarized in

Theorem 3.1 *If $N((\infty,1)) > V$, then there exists a threshold policy (L^*,η^*) which solves problem (P_V) with $N((L^*,\eta^*)) = V$.*

The proof of Theorem 3.1 is outlined in Section 6. The solution method for these constrained optimization problems uses Lagrangian arguments similar to the ones given in [2,10]. Here, the appropriate Lagrangian functional is defined for any admissible policy π in $\mathcal{P}$ to be

$$J^\gamma(\pi) := \liminf_n \frac{1}{n+1} E^\pi \sum_{t=0}^{n} \mu 1[X(t) \neq 0] - \gamma X(t) \tag{3.4}$$

with $\gamma > 0$ denoting the Lagrange multiplier. The corresponding *Lagrangian* problem (LP^γ) is then the *unconstrained* problem

$$(LP^\gamma): \quad \text{Maximize } J^\gamma(\pi) \text{ over } \mathcal{P}.$$

Under (P1)-(P3), each unconstrained problem $(LP^\gamma), \gamma > 0$, can be viewed as a Markov decision problem under the long-run average cost criterion, with state process $\{X(n)\}_0^\infty$, cost per stage $c^\gamma: \mathbb{N} \to \mathbb{R}$ given by

$$c^\gamma(x) := \mu 1[x \neq 0] - \gamma x \tag{3.5}$$

for all x in $\mathbb{N}$, and information pattern $\{H(n)\}_0^\infty$. Although the information pattern $H(n)$ is richer than the standard state feedback information pattern $\{\Xi, U(k), X(k+1), 0 \leq k < n\}$, it is easy to see that the value functions for the corresponding discounted problems coincide.

If $\gamma \geq \mu$, then $c^\gamma(x) \leq 0$ for all x in $\mathbb{N}$, and an elementary argument now shows that $J^\gamma((0,0)) = 0 \geq J^\gamma(\pi)$ for all π in $\mathcal{P}$, whence the threshold policy $(0,0)$ trivially solves the Lagrangian problem (LP^γ). Therefore, only the case $\gamma < \mu$ needs to be investigated and this is done in the next section by considering the appropriate discounted problems.

4. The discounted problems

Let $\mu > \gamma > 0$ and $0 < \beta < 1$ be held fixed throughout this section. The expected *β-discounted* Lagrangian cost $J_\beta^\gamma(\pi)$ associated with an admissible policy π in $\mathcal{P}$ is defined by

$$J_\beta^\gamma(\pi) := E^\pi \sum_{t=0}^{\infty} \beta^t c^\gamma(X(t)), \tag{4.1}$$

and the corresponding discounted optimization problem (LP_β^γ) is then simply

$$(LP_\beta^\gamma): \quad \text{Maximize } J_\beta^\gamma(\pi) \text{ over } \mathcal{P}.$$

Since at most one arrival can be admitted in each time slot, the pathwise bound $X(n) \leq \Xi + n$ holds for all $n = 0, 1, \ldots$ and yields the estimate

$$|J_\beta^\gamma(\pi)| \leq \frac{\mu + \gamma E^\pi[\Xi]}{1-\beta} + \frac{\gamma\beta}{(1-\beta)^2} < \infty. \tag{4.2}$$

The bound (4.2) being independent of the policy π in $\mathcal{P}$, the quantity $J_\beta^\gamma(\pi)$ is thus well-defined and *uniformly* bounded over $\mathcal{P}$.

As customary with the Dynamic Programming methodology, the *β-discounted cost-to-go* associated with any policy π in $\mathcal{P}$ is the mapping $J_\beta^{\gamma,\pi}: \mathbb{N} \to \mathbb{R}$ defined by

$$J_\beta^{\gamma,\pi}(x) := E_x^\pi[\sum_{t=0}^{\infty} \beta^t c^\gamma(X(t))] \tag{4.3a}$$

for all x in $\mathbb{N}$, while the corresponding *value function* $V_\beta^\gamma: \mathbb{N} \to \mathbb{R}$ is given by

$$V_\beta^\gamma(x) := \sup_{\pi \in \mathcal{P}} J_\beta^{\gamma,\pi}(x). \tag{4.3b}$$

Let the RV's A and B be generic elements in $\{A(n)\}_0^\infty$ and $\{B(n)\}_0^\infty$, respectively, and for all x in $\mathbb{N}$, define the $\mathbb{N}$-valued RV's $A^0(x)$ and $A^1(x)$ by

$$A^i(x) = x + iA - 1[x \neq 0]B, \quad i = 0, 1. \tag{4.4}$$

For any mapping $f: \mathbb{N} \to \mathbb{R}$, define the mapping $T_\beta^\gamma f: \mathbb{N} \to \mathbb{R}$ by

$$(T_\beta^\gamma f)(x) = c^\gamma(x) + \beta \max_{0 \leq \eta \leq 1} \left\{ \eta E[f(A^1(x))] + \bar{\eta} E[f(A^0(x))] \right\} \tag{4.5}$$

for all x in $\mathbb{N}$. Here, for each $i = 0, 1$, $E[f(A^i(x))] := E^\pi[f(A^i(x))]$ for all π in P owing to (2.3)-(2.5), with

$$E[f(A^1(x))] = \begin{cases} \lambda f(1) + \bar{\lambda} f(0) & \text{if } x = 0; \\ \bar{\lambda}\mu f(x-1) + (\lambda\mu + \bar{\lambda}\bar{\mu}) f(x) + \lambda\bar{\mu} f(x+1) & \text{if } x \geq 1 \end{cases} \tag{4.6a}$$

and

$$E[f(A^0(x))] = \begin{cases} f(0) & \text{if } x = 0; \\ \mu f(x-1) + \bar{\mu} f(x) & \text{if } x \geq 1. \end{cases} \tag{4.6b}$$

For future reference, for any mapping $f: \mathbb{N} \to \mathbb{R}$, pose

$$\nabla f(x) = \begin{cases} f(1) - f(0) & \text{if } x = 0; \\ \mu(f(x) - f(x-1)) + \bar{\mu}(f(x+1) - f(x)) & \text{if } x \geq 1 \end{cases} \tag{4.7}$$

and observe that

$$E[f(A^1(x))] - E[f(A^0(x))] = \lambda \nabla f(x) \tag{4.8}$$

for all x in $\mathbb{N}$.

The backward induction of Dynamic Programming produces the sequence $\{V_\beta^n\}_0^\infty$ of mappings $V_\beta^n: \mathbb{N} \to \mathbb{R}$ through the recursion

$$V_\beta^{n+1} = T_\beta^\gamma V_\beta^n \qquad n = 0, 1, \ldots \tag{4.9}$$

with $V_\beta^0 := c^\gamma$. The cost c^γ being bounded above by $\mu - \gamma$, the discounted problem (LP_β^γ) is covered by Assumption P of Bertsekas [1, pp. 251]. The following theorem is now readily obtained by specializing results from Section 6.4 of Bertsekas [1].

Theorem 4.1 *The value function V_β^γ satisfies the Dynamic Programming equation*

$$V_\beta^\gamma = T_\beta^\gamma V_\beta^\gamma \tag{4.10}$$

and is obtained as the pointwise limit

$$\lim_n V_\beta^n(x) = V_\beta^\gamma(x) \tag{4.11}$$

for all x in $\mathbb{N}$. Moreover, the Markov stationary policy g^* in P defined by

$$g^*(x) = \begin{cases} 1 & \text{if } \nabla V_\beta^\gamma(x) > 0; \\ \text{arbitrary in } [0,1] & \text{if } \nabla V_\beta^\gamma(x) = 0; \\ 0 & \text{if } \nabla V_\beta^\gamma(x) < 0\ . \end{cases} \tag{4.12}$$

is optimal for problem (LP_β^γ).

The value iteration method implicit in Theorem 4.1 constitutes a powerful tool for further characterizing the structure of the optimal policy. Some insights are already obtained through Lemma 4.2 below. Pose $L^\gamma := \max\{l \in \mathbb{N}:\ \mu - \gamma l > 0\}$, and observe that $L^\gamma \geq 1$.

Lemma 4.2 *For the discounted problem* (LP_β^γ), *an optimal Markov stationary policy* g^* *can always be chosen so that* $g^*(x) = 0$ *for all* $x > L^\gamma$.

Proof. Define the $\mathbb{F}_n$-stopping time τ by

$$\tau := \begin{cases} \inf\{\ k \geq 0 :\ X(k) = L^\gamma\} & \text{if the set is non-empty;} \\ \infty & \text{otherwise} \end{cases} \tag{4.13}$$

with the obvious interpretation that τ is the first passage time into the state L^γ. For any admissible policy π in $\mathcal{P}$, pose $B_\beta^\pi(x) := E_x^\pi[\beta^\tau]$ and

$$I_\beta^{\gamma,\pi}(x) := E_x^\pi[\sum_{t=0}^{\tau-1} \beta^t c^\gamma(X(t))] = E_x^\pi[\sum_{t=0}^{\infty} \beta^t 1[\tau > t] c^\gamma(X(t))]$$

for all x in $\mathbb{N}$. Since an optimal stationary policy exists by Theorem 4.1, a simple argument based on the strong Markov property readily shows that the value function V_β^γ satisfies the relation

$$V_\beta^\gamma(x) = \max_g \{I_\beta^{\gamma,g}(x) + V_\beta^\gamma(L^\gamma) B_\beta^g(x)\} \tag{4.14}$$

for all x in $\mathbb{N}$, with the maximization being taken over all *stationary* policies g in $\mathcal{P}$.

Now, from any arbitrary stationary policy g in $\mathcal{P}$, construct from it a new policy $\tilde{g}$ which generates actions according to

$$\tilde{g}(x) := \begin{cases} g(x) & \text{if } 0 \leq x \leq L^\gamma\ ; \\ 0 & \text{if } x > L^\gamma. \end{cases} \tag{4.15}$$

Since $V_\beta^\gamma(L^\gamma) \geq J_\beta^{\gamma,(0,0)}(L^\gamma) \geq c^\gamma(L^\gamma) > 0$, Lemma 4.2 will now be established by showing that $I_\beta^{\gamma,g}(x) \leq I_\beta^{\gamma,\tilde{g}}(x)$ and $B_\beta^g(x) \leq B_\beta^{\tilde{g}}(x)$ for all x in $\mathbb{N}$.

If $0 \leq x \leq L^\gamma$, the probability measures P_x^g and $P_x^{\tilde{g}}$ coincide on the σ-field $\mathbb{F}_\tau$, whence $I_\beta^{\gamma,g}(x) = I_\beta^{\gamma,\tilde{g}}(x)$ and $B_\beta^g(x) = B_\beta^{\tilde{g}}(x)$. For $x > L^\gamma$, pose $Z(t) = 1[\tau > t]X(t)$ for all $t = 0, 1, \ldots$.

The reader will now check from the very definition of $\tilde{g}$ and from the proof of Theorem A.2 in the Appendix that

$$(\{Z(t)\}_0^\infty, P_x^{\tilde{g}}) \leq_{st} (\{Z(t)\}_0^\infty, P_x^g). \tag{4.16}$$

Moreover, for $t = 0, 1, \ldots$, the relations $1[\tau > t] = f(Z(1), \ldots, Z(t))$ and $1[\tau > t]c^\gamma(X(t)) = -h(Z(1), \ldots, Z(t))$ take place both P_x^g-a.s. and $P_x^{\tilde{g}}$-a.s., with monotone *non-decreasing* mappings $f, h: \mathbb{N}^t \to \mathbb{R}$. It is now immediate from (4.16) that $(\tau, P_x^{\tilde{g}}) \leq_{st} (\tau, P_x^g)$, or equivalently, that $(\beta^\tau, P_x^g) \leq_{st} (\beta^\tau, P_x^{\tilde{g}})$ and that

$$(1[\tau > t]c^\gamma(X(t)), P_x^g) \leq_{st} (1[\tau > t]c^\gamma(X(t)), P_x^{\tilde{g}})$$

for all $t = 0, 1, \ldots$ The inequalities $B_\beta^g(x) \leq B_\beta^{\tilde{g}}(x)$ and $I_\beta^{\gamma,g}(x) \leq I_\beta^{\gamma,\tilde{g}}(x)$ are now readily obtained. □

By virtue of Lemma 4.2, the Dynamic Programming equation (4.10) reduces to

$$V_\beta^\gamma(x) = c^\gamma(x) + \beta E[V_\beta^\gamma(A^0(x))] \tag{4.17}$$

for all $x > L^\gamma$, and easy algebraic manipulations yield

$$V_\beta^\gamma(L^\gamma + 1) - V_\beta^\gamma(L^\gamma) < 0. \tag{4.18}$$

The value iteration method of Theorem 4.1 is now used to establish the *integer-concavity* of the value function V_β^γ by showing the concavity of each one of the iterates $\{V_\beta^n\}_0^\infty$ given by (4.9). Surprisingly enough, this turns out to be a non-trivial task as several situations need to be discussed separately. The difficulty seems to stem from the fact that *multiple* transitions are possible here owing to the discrete nature of the time parameter. This is in contrast with the continuous-time version of this problem for which concavity of the value function is more readily obtained through some of the arguments of [13].

The next result shows in what sense integer-concavity is preserved under the backward induction of Dynamic Programming. It will be convenient to say that a mapping $f: \mathbb{N} \to \mathbb{R}$ satisfies the property (Ai), $i = 1, \ldots, 4$, if

(A1): f is integer-concave with $0 \leq f(1) - f(0) \leq \mu - \gamma$,

(A2): f is integer-concave with $\mu - \gamma \leq f(1) - f(0) \leq 1$,

(A3): f is integer-concave with $f(2) - f(1) \leq -\gamma$,

(A4): f is integer-concave with $f(2) - f(1) \geq -\gamma$.

Theorem 4.3 (i) *Suppose* $\lambda + \mu \leq 1$. *If* f *satisfies* (A2), *so does* $T_\beta^\gamma f$. (ii) *Suppose* $\lambda + \mu > 1$ *and* $\mu^2 < \gamma$. *If* f *satisfies* (A1) *and* (A3), *so does* $T_\beta^\gamma f$. (iii) *Suppose* $\lambda + \mu > 1$ *and* $\frac{\gamma}{\mu} \leq \bar{\lambda} \leq \mu$. *If* f *satisfies* (A2) *and* (A4), *so does* $T_\beta^\gamma f$. (iv) *Suppose* $\lambda + \mu > 1$ *and* $\bar{\lambda} < \frac{\gamma}{\mu} \leq \mu$. *If* f *satisfies* (A1) *with* $\nabla f(1) < 0$, *then* $T_\beta^\gamma f$ *satisfies* (A1) *and* (A3). *If* f *satisfies* (A1) *and* (A3) *with* $\nabla f(1) \geq 0$, *then* $T_\beta^\gamma f$ *satisfies* (A1) *and* (A4). *If* f *satisfies either* (A1) *and* (A4) *with* $\nabla f(1) \geq 0$, *or* (A2) *and* (A4) (*whence* $\nabla f(1) \geq 0$ *necessarily*), *then* $T_\beta^\gamma f$ *satisfies either* (A1) *and* (A4), *or* (A2) *and* (A4).

A complete discussion of this key result can be found in [6, Appendix II]. It is noteworthy that concavity *alone* does *not* propagate under the induction argument and that additional *growth* conditions are needed.

Theorem 4.4 *The value function* V_β^γ *is integer-concave, with*

$$\frac{\beta\lambda}{1-\beta\bar{\lambda}} V_\beta^\gamma(1) = V_\beta^\gamma(0) < V_\beta^\gamma(1) \quad \text{and} \quad V_\beta^\gamma(L^\gamma + 1) < V_\beta^\gamma(L^\gamma). \tag{4.19}$$

Proof. The argument is standard and inductively uses Theorem 4.3 on the successive iterates $\{V_\beta^n\}_0^\infty$. This is made possible by observing that the 0th iterate V_β^0 is the concave mapping c^γ which satisfies $c^\gamma(1) - c^\gamma(0) = \mu - \gamma$ and $c^\gamma(2) - c^\gamma(1) = -\gamma$. The reader will now check that each one of the four situations discussed in Theorem 4.3 applies to yield the integer-concavity of V_β^n with $0 \leq V_\beta^n(1) - V_\beta^n(0) \leq 1$ for all $n = 1, 2, \ldots$.

In the limit, by Theorem 4.1, the value function V_β^γ is thus integer-concave with $0 \leq V_\beta^\gamma(1) - V_\beta^\gamma(0) \leq 1$. Consequently, $\nabla V_\beta^\gamma(0) \geq 0$ and $g^*(0) = 1$ is an optimal action by (4.12), whence $V_\beta^\gamma(0) = \beta[\lambda V_\beta^\gamma(1) + \bar{\lambda} V_\beta^\gamma(0)]$ by virtue of the Dynamic Programming equation (4.10). The first part of (4.19) now follows from the fact that $V_\beta^\gamma(1) \geq J_\beta^{\gamma,(0,0)}(1) \geq \mu - \gamma > 0$, while the second part is nothing but (4.18). □

Theorem 4.5 *For every* $0 < \beta < 1$, *there exists a threshold policy* $(L_\beta^\gamma, 1)$ *which solves problem* (LP_β^γ), *where the optimal threshold value* L_β^γ *satisfies the relation* $0 \leq L_\beta^\gamma \leq L^\gamma$.

Proof. Integer-concavity of V_β^γ implies ∇V_β^γ to be monotone decreasing, whence the quantity ∇V_β^γ changes sign at most once, from positive to negative, as a consequence of (4.19). Therefore, there exists a level L_β^γ, with $0 \leq L_\beta^\gamma \leq L^\gamma$, such that $\nabla V_\beta^\gamma(x) \geq 0$ for $0 \leq x \leq L_\beta^\gamma$ and

$\nabla V_\beta^\gamma(x) < 0$ for all $x > L_\beta^\gamma$. The threshold policy $(L_\beta^\gamma, 1)$ is then clearly optimal by Theorem 4.1. □

5. Properties of threshold policies

For each threshold policy (L, η) with L in $\mathbb{N}$ and $0 \leq \eta \leq 1$, the sequence $\{X(n)\}_0^\infty$ is a time-homogeneous Markov chain with state space $\mathbb{N}$ under the probability measure $P^{(L,\eta)}$. This chain has a single ergodic set, namely, $\{0, 1, \ldots, L\}$ if $\eta = 0$ or $\{0, 1, \ldots, L+1\}$ if $0 < \eta \leq 1$, with all the other states being transient. Consequently, the Markov chain $\{X(n)\}_0^\infty$ admits under $P^{(L,\eta)}$ a unique invariant measure, which is denoted by $\mathbb{P}^{(L,\eta)}$ with corresponding expectation operator $\mathbb{E}^{(L,\eta)}$. As pointed out in [6, Sec. 5], routine calculations yield closed-form expressions for this invariant measure $\mathbb{P}^{(L,\eta)}$.

Let X denote a generic $\mathbb{N}$-valued random variable. For any mapping $d : \mathbb{N} \to \mathbb{R}$, the quantity $\mathbb{E}^{(L,\eta)} d(X)$ is always *finite* since $\mathbb{P}^{(L,\eta)}$ has finite support. Furthermore, the first passage time to the set of ergodic states being a.s. finite under the threshold policy (L, η), the queue sizes $\{X(n)\}_0^\infty$ satisfy the inequality

$$X(n) \leq \Xi \vee (L+1) \quad P^{(L,\eta)} - a.s. \qquad n = 0, 1, \ldots \quad (5.1)$$

The next result is now immediate from standard properties of Markov chains.

Lemma 5.1 *If the mapping $d: \mathbb{N} \to \mathbb{R}$ is monotone and the RV $d(\Xi)$ is integrable, then the RV's $\{d(X(n))\}_0^\infty$ are uniformly integrable under $P^{(L,\eta)}$, and the convergence*

$$\lim_n \frac{1}{n+1} \sum_{t=0}^{n} d(X(t)) = \mathbb{E}^{(L,\eta)} d(X) \qquad P^{(L,\eta)} - a.s. \quad (5.2)$$

takes place, independently of the initial distribution, both $P^{(L,\eta)} - a.s.$ and in $L^1(\Omega, \mathbb{F}, P^{(L,\eta)})$.

The next lemma will be useful in proving the main result of Section 6. The obtained results are also applicable to various situations discussed in the companion papers [7,8].

Lemma 5.2 *For any mapping $d: \mathbb{N} \to \mathbb{R}$ and any threshold policy (L, η), there always exist a scalar J and a mapping $h: \mathbb{N} \to \mathbb{R}$ such that*

$$h(x) + J = d(x) + (L,\eta)(x) E[h(A^1(x))] + \overline{(L,\eta)(x)} E[h(A^0(x))] \quad (5.3)$$

for all x in $\mathbb{N}$. The quantity J is given by

$$J = \lim_n \frac{1}{n+1} E_x^{(L,\eta)}[\sum_{t=0}^{n} d(X(t))] = \mathbb{E}^{(L,\eta)} d(X), \tag{5.4}$$

whereas the mapping $h: \mathbb{N} \to \mathbb{R}$ is unique up to an additive constant, and under the constraint $h(L) = 0$, is given by

$$h(x) = E_x^{(L,\eta)}[\sum_{t=0}^{\tau-1} d(X(t))] - E_x^{(L,\eta)}[\tau] J \tag{5.5}$$

for all x in $\mathbb{N}$. Here τ is the $\mathbb{F}_n$-stopping time defined by (4.13) but with L instead of L^γ.

The proof of Lemma 5.2 is by now standard and is omitted for sake of brevity. The interested reader is invited to consult the monograph by Ross [11] or the work by Shwartz and Makowski [12] for a typical discussion. Crucial to the argument is the fact that the chain $\{X(n)\}_0^\infty$ reduces to a *finite-state* Markov chain under $P_x^{(L,\eta)}$ for every x in $\mathbb{N}$.

It is clear from (5.4) and Lemma 5.1 that the equality $J^\gamma((L,\eta)) = T((L,\eta)) - \gamma N((L,\eta))$ holds, with $T((L,\eta)) = \mu \mathbb{P}^{(L,\eta)}[X \neq 0]$ and $N((L,\eta)) = \mathbb{E}^{(L,\eta)} X$ for every L in $\mathbb{N}$ and $0 \leq \eta \leq 1$. The following properties are obtained by routine inspection.

Lemma 5.3 *For every L in $\mathbb{N}$, the mappings $\eta \to T((L,\eta))$ and $\eta \to N((L,\eta))$ are continuously differentiable and strictly monotone increasing on the interval $[0,1]$, whence the quantities $T((L,0))$ and $N((L,0))$ increase as L increases.*

Lemma 5.4 *For each $\gamma > 0$, the mapping $L \to J^\gamma((L,0))$ is unimodal, with the global maximum being achieved at at most two adjacent levels.*

6. The long-run average problems

The value L^γ of Theorem 4.5 is independent of the policy π and of the discount factor β, and this makes it possible to solve the long-run average problem (LP^γ) by standard Tauberian arguments applied to the discounted problem (LP_β^γ). This result, proved in [6, Thm. 6.1], is now summarized.

Theorem 6.1 *For each $\gamma > 0$, there always exists a threshold policy $(L_\gamma^*, \eta_\gamma^*)$, with $0 \leq L_\gamma^* \leq L^\gamma$, which solves the long-run average problem (LP^γ) and yields the optimal cost J^γ as $J^\gamma = \mathbb{E}^{(L_\gamma^*, \eta_\gamma^*)} c^\gamma(X)$. If $\mu - \gamma \leq 0$,* i.e., *$L^\gamma = 0$, then necessarily $\eta_\gamma^* = 0$ and $J^\gamma = 0$, while if $\mu - \gamma > 0$,* i.e., *$L^\gamma > 0$, then η_γ^* can always be chosen to be* 1.

For each $\gamma > 0$, the search for an optimal policy can therefore be restricted to the class of all *pure* threshold policies with threshold below level $L^\gamma + 1$, and the optimal cost J^γ of problem (LP^γ) can simply be written as

$$J^\gamma = \max_{0 \le L \le L^\gamma + 1} J^\gamma((L,0)) = \max_{L \in \mathbb{N}} I\!E^{(L,0)} c^\gamma(X). \tag{6.1}$$

The results on threshold policies obtained in Section 5 can be used to identify the optimal threshold policy through (6.1). This idea is now exploited to produce a stronger result which is essential to solving the constrained problem (P_V).

Theorem 6.2 *For each threshold value L in $\mathbb{N}$, there always exists $\gamma(L) > 0$, with $L^{\gamma(L)} \ge L$, so that any admissible policy π in $\mathcal{P}$ given by*

$$\pi_n(H(n)) = \begin{cases} 1 & \text{if } X(n) < L; \\ \text{arbitrary in } [0,1] & \text{if } X(n) = L; \\ 0 & \text{if } X(n) > L \end{cases} \tag{6.2}$$

solves the long-run average problem $(LP^{\gamma(L)})$.

The following result will be useful in the proof of Theorem 6.2.

Lemma 6.3 *Let the pair (h, J) obtained in Lemma 5.2. If the sequence $\{d(X(n))\}_0^\infty$ is uniformly integrable under $P^{(L,\eta)}$, then the convergence*

$$\lim_n \frac{1}{n+1} \left\{ E^{(L,\eta)}[h(X(n+1))] - E^{(L,\eta)}[h(\Xi)] \right\} = 0 \tag{6.3}$$

takes place.

A proof of Theorem 6.2. For each L in $\mathbb{N}$, pose

$$\gamma(L) := \frac{T((L,1)) - T((L,0))}{N((L,1)) - N((L,0))}. \tag{6.4}$$

It is plain that $\gamma(L) > 0$ owing to Lemma 5.3. and that (6.4) is equivalent to $J^{\gamma(L)}((L,1)) = J^{\gamma(L)}((L,0))$. It then follows from Lemma 5.4 and (6.1) that

$$J^{\gamma(L)}((L,1)) = J^{\gamma(L)}((L,0)) = \max_{l \in \mathbb{N}} J^{\gamma(L)}((l,0)) = J^{\gamma(L)}, \tag{6.5}$$

and both policies $(L,1)$ and $(L,0)$ solve problem $(LP^{\gamma(L)})$. To prove Theorem 6.2, i.e., that any policy π of the form (6.2) is optimal for problem $(LP^{\gamma(L)})$, it only remains to show that $J^{\gamma(L)}(\pi) = J^{\gamma(L)}$. Note that the policies $(L,1)$ and $(L,0)$ are clearly among these policies.

Take the mapping $d = c^{\gamma(L)}$ in Lemma 5.2, and let (h_i, J_i) be the corresponding solution to the Poisson equation (5.4) associated with the threshold policy (L,i), $i = 0,1$, when $h_1(L) = h_0(L) = 0$. The relation (6.5) yields $J_1 = J_0 = J^{\gamma(L)}$, whereas direct inspection of (5.5) reveals $h_1(x) = h_0(x)$ for all $x \neq L$ by the very definition of the stopping time τ. The condition $h_1(L) = h_0(L) = 0$ immediately implies that $h_1 \equiv h_0 =: h$. Consequently, the Poisson equations associated with the two policies $(L,1)$ and $(L,0)$ must coincide, with $E[h(A^1(L))] = E[h(A^0(L))]$, and must be of the form

$$h(x) + J^{\gamma(L)} = c^{\gamma(L)}(x) + p(x)E[h(A^1(x))] + \overline{p(x)}E[h(A^0(x))] \tag{6.6}$$

for all x in $\mathbb{N}$, where

$$p(x) = \begin{cases} 1 & \text{if } 0 \leq x < L; \\ \text{arbitrary in } [0,1] & \text{if } x = L; \\ 0 & \text{if } x > L. \end{cases}$$

Under the policy π in $\mathcal{P}$ defined by (6.2), a standard argument based on (6.6) leads to the relation

$$J^{\gamma(L)} = \frac{1}{n+1}E^{\pi}\sum_{t=0}^{n} c^{\gamma(L)}(X(t)) + \frac{1}{n+1}\Big\{E^{\pi}[h(X(n+1))] - E^{\pi}[h(\Xi)]\Big\} \quad n = 0,1,\ldots \tag{6.7}$$

By the very form (6.2) assumed for π and the definition of τ, it is easy to see that for any $0 \leq \eta \leq 1$, both probability measures $P^{(L,\eta)}$ and P^{π} coincide on $\mathcal{F}_{\tau}$. As a result,

$$\lim_n \frac{1}{n+1}\Big\{E^{\pi}[h(X(n+1))] - E^{\pi}[h(\Xi)]\Big\} = 0 \tag{6.8}$$

upon invoking Lemma 6.3. The relation $J^{\gamma(L)}(\pi) = J^{\gamma(L)}$ is now obtained by taking the limit in (6.7) and making use of (6.8). □

A proof of Theorem 3.1. It should be clear from Theorems 5.1 and 6.2 that any threshold policy (L,η), with η arbitrary in [0,1], yields (3.1)-(3.2) as limits and solves problem (LP^{γ})

for some $\gamma(L) > 0$. As discussed in [9], in order to solve problem (P_V), it only remains to find a threshold policy that saturates the constraint.

Since $N((0,0)) = 0$, Lemma 5.3 readily implies the existence and uniqueness of the pair (L^*, η^*) such that $N((L^*, \eta^*)) = V$ if $N((\infty, 1)) > V$. The optimal threshold and bias values L^* and η^* are uniquely defined by solving $\mathbb{E}^{(L,\eta)} X = V$, $0 \leq \eta \leq 1$ and $L = 0, 1, \ldots$. □

References

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

[2] F. J. Beutler and K. W. Ross, "Time-average optimal constrained semi-Markov decision processes," *Adv. Appl. Prob.*, Vol. 18, pp. 341-359 (1986).

[3] M. Gerla and L. Kleinrock, "Flow control: A comparative survey," *IEEE Trans. Commun.*, Vol. COM-28, pp. 553-574 (1980).

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

[5] A. A. Lazar, "The throughput time delay function of an $M|M|1$ queue," *IEEE Trans. Info. Theory*, Vol. IT-29, pp. 914-918 (1983).

[6] D.-J. Ma and A. M. Makowski, "A simple problem of flow control I: Optimality results," *IEEE Trans. Auto. Control*, submitted (1987).

[7] D.-J. Ma and A. M. Makowski, "A simple problem of flow control II: Implementation of threshold policies via Stochastic Approximations," *IEEE Trans. Auto. Control*, submitted (1987).

[8] D.-J. Ma and A. M. Makowski, "Parameter estimation under threshold policies for a simple flow control problem," these Proceedings.

[9] D.-J. Ma, A. M. Makowski and A. Shwartz, "Estimation and optimal control for constrained Markov chains," *Proceedings of the 25th IEEE Conference on Decision and Control*, Athens, Greece, pp. 994-999 (1986).

[10] K. W. Ross, *Constrained Markov Decision Processes With Queueing Applications*, Ph.D. Thesis, CICE Program, University of Michigan, Ann Arbor, Michigan (1985).

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

[12] A. Shwartz and A. M. Makowski, "Comparing policies in Markov decision processes: Mandl's lemma revisited," *Mathematics of Operations Research*, submitted (1987).

[13] S. Stidham, Jr., "Optimal control of admission to a queueing system," *IEEE Trans. Auto. Control*, Vol. AC-30, pp. 705-713 (1985).

[14] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models*, John Wiley & Sons, New York (1983).

Appendix

A comparison result

The notion of *stochastic ordering* is useful for comparing the performance of various policies. This is achieved through a stochastic comparison result for the underlying queue size process. The reader is invited to consult the monograph by Stoyan [14] for further information on stochastic orderings.

Let P^1 and P^2 be two probability measures defined on $\mathcal{F}$, with expectation operators E^1 and E^2, respectively. If Y is an $\mathbb{R}^n$-valued RV defined on $(\Omega, \mathcal{F})$, then the RV (Y, P^2) is said to be *stochastically larger* than (Y, P^1) if and only if $E^1[f(Y)] \leq E^2[f(Y)]$ for all *increasing* functions $f: \mathbb{R}^n \to \mathbb{R}$ for which these expectations exist; this is customarily denoted by $(Y, P^1) \leq_{st} (Y, P^2)$.

This notion extends naturally to *sequences* of $\mathbb{R}$-valued RV's defined on $(\Omega, \mathcal{F})$. The sequence $(\{Y(t)\}_0^\infty, P^2)$ is said to be *stochastically larger* than $(\{Y(t)\}_0^\infty, P^1)$ if and only if

$$((Y(0), Y(1), \ldots, Y(n)), P^1) \leq_{st} ((Y(0), Y(1), \ldots, Y(n)), P^2) \qquad n = 0, 1, \ldots \quad (A.1)$$

This is denoted simply by $(\{Y(t)\}_0^\infty, P^1) \leq_{st} (\{Y(t)\}_0^\infty, P^2)$ and amounts to

$$E^1[f(Y(0), Y(1), \ldots, Y(n))] \leq E^2[f(Y(0), Y(1), \ldots, Y(n))] \qquad n = 0, 1, \ldots \quad (A.2)$$

for all increasing functions $f: \mathbb{R}^{n+1} \to \mathbb{R}$ for which the expectations exist [14, Thm. 4.1.2, pp. 61].

While the relation $(A.2)$ is usually hard to verify directly in practice, sufficient conditions are available in the literature. One such condition, due to Veinott [14, pp. 29], is given below

for easy reference. Throughout the discussion, the $\mathbb{R}^{n+1}$-valued RV $(Y(0),\ldots,Y(n))$ and the element $(y_0,\ldots,y_n)$ of $\mathbb{R}^{n+1}$ are denoted by $Y^{(n)}$ and $y^{(n)}$, respectively.

Lemma A.1 *Let $\{Y(t)\}_0^\infty$ be a sequence of $\mathbb{R}$-valued RV's on $(\Omega,\mathbb{F})$. If*

$$(Y(0),P^1) \leq_{st} (Y(0),P^2) \qquad (A.3a)$$

and for every a in $\mathbb{R}$,

$$P^1[Y(n+1)>a|Y^{(n)}=x^{(n)}] \leq P^2[Y(n+1)>a|Y^{(n)}=y^{(n)}] \qquad n=0,1,\ldots (A.3b)$$

whenever $x^{(n)} \leq y^{(n)}$ componentwise in $\mathbb{N}^{n+1}$, then $(\{Y(t)\}_0^\infty,P^1) \leq_{st} (\{Y(t)\}_0^\infty,P^2)$.

Here, this result is used as follows. For every admissible policy π in $\mathcal{P}$, introduce the sequence $\{\hat{\pi}_n\}_0^\infty$ of mappings $\hat{\pi}_n\colon \mathbb{N}^{n+1} \to [0,1]$ defined by

$$\hat{\pi}_n(x^{(n)}) := P^\pi[U(n)=1|X^{(n)}=x^{(n)}] = E^\pi[\pi_n(H(n))|X^{(n)}=x^{(n)}] \qquad n=0,1,\ldots (A.4)$$

for all $x^{(n)}$ in $\mathbb{N}^{n+1}$.

Theorem A.2 *Consider two admissible policies π^1 and π^2 in $\mathcal{P}$. If the relations*

$$\hat{\pi}_n^1(x^{(n)}) \leq \hat{\pi}_n^2(y^{(n)}) \qquad n=0,1,\ldots (A.5a)$$

hold for all $x^{(n)} \leq y^{(n)}$ with $x_n=y_n$, and if

$$\lambda\hat{\pi}_n^1(x^{(n-1)},0) \leq \bar{\mu}+\lambda\mu\hat{\pi}_n^2(y^{(n-1)},1) \qquad n=0,1,\ldots (A.5b)$$

holds for all $x^{(n-1)} \leq y^{(n-1)}$, then $(\{X(t)\}_0^\infty,P^{\pi^1}) \leq_{st} (\{X(t)\}_0^\infty,P^{\pi^2})$.

Proof. Since the probability distribution of Ξ is independent of the policy, the relation $(A.3a)$ trivially holds. It suffices to show that the conditions $(A.5)$ imply $(A.3b)$.

Routine calculations first imply via (2.3)-(2.5) that for every policy π in $\mathcal{P}$,

$$P^\pi[X(n+1)>a|X^{(n)}=x^{(n)}] = \begin{cases} 1 & \text{if } x_n > a+1; \\ \bar{\mu}+\lambda\mu\hat{\pi}_n(x^{(n)}) & \text{if } x_n = a+1; \\ \lambda\bar{\mu}\hat{\pi}_n(x^{(n)}) & \text{if } x_n = a > 0; \\ \lambda\hat{\pi}_n(x^{(n)}) & \text{if } x_n = a = 0; \\ 0, & \text{if } 0 \leq x_n < a \end{cases} \qquad n=0,1,\ldots (A.6)$$

for all a in $\mathbb{N}$ and $x^{(n)}$ in $\mathbb{N}^{n+1}$. If $a>0$, $0 \leq \lambda\bar{\mu}\hat{\pi}_n(x^{(n-1)},a) \leq \bar{\mu}+\lambda\mu\hat{\pi}_n(y^{(n-1)},a+1) \leq 1$ for *all* $x^{(n-1)}$ and $y^{(n-1)}$ in $\mathbb{R}^n$, and $(A.3b)$ thus holds whenever $x^{(n)} \leq y^{(n)}$ with $0<x_n<y_n$

for any two *arbitrary* policies in $\mathcal{P}$. For the policies π^1 and π^2 considered here, it is now plain that $(A.5)$ and $(A.6)$ combine to yield $(A.3b)$ whenever $x^{(n)} \leq y^{(n)}$, and the conclusion now follows from Lemma A.1. □

Theorem A.3 *Consider two admissible policies π^1 and π^2 in $\mathcal{P}$. If there exists a sequence $\{f_n\}_0^\infty$ of mappings $f_n: \mathbb{N} \to [0,1]$ such that*

$$\pi_n^1(h_n) \leq f_n(x_n) \leq \pi_n^2(h_n) \qquad n = 0, 1, \ldots (A.7a)$$

for all h_n in $\mathbb{H}_n$, with

$$\lambda f_n(0) \leq \bar{\mu} + \lambda\mu f_n(1), \qquad n = 0, 1, \ldots (A.7b)$$

then $(\{X(t)\}_0^\infty, P^{\pi^1}) \leq_{st} (\{X(t)\}_0^\infty, P^{\pi^2})$.

Proof. It is plain from $(A.4)$ and $(A.7)$ that for all $x^{(n)}$ and $y^{(n)}$ in $\mathbb{N}^{n+1}$, $\hat{\pi}_n^1(x^{(n)}) \leq f_n(x_n)$, $\hat{\pi}_n^2(y^{(n)}) \geq f_n(y_n)$, $\lambda\hat{\pi}_n^1(x^{(n-1)}, 0) \leq \lambda f_n(0)$ and $\bar{\mu} + \lambda\mu\hat{\pi}_n^2(y^{(n-1)}, 1) \geq \bar{\mu} + \lambda\mu f_n(1)$. Condition $(A.5)$ is now easily justified, and the result follows from Theorem A.2. □