Stability and optimization of queueing networks and their fluid models

Sean Meyn

Abstract. We describe some approaches to the synthesis of optimal policies for multiclass queueing network models based upon the close connection between stability of queueing networks and their associated fluid limit models.

It is found that there is also a close connection between optimization of the network and optimal control of the far simpler fluid network model. This observation leads to a new class of scheduling policies that perform well on the examples considered.

1. Introduction

This paper concerns the theory of optimal scheduling for multiclass queueing networks with deterministic routing, as may be used in the modeling of a semiconductor manufacturing plant. To keep notation to a minimum we assume a discrete-time, discrete-state model described by a Markov Decision Process (MDP). Criteria which ensure the *existence* of an optimal policy for an MDP when the cost is possibly unbounded have been developed in numerous papers over the past decade [**Put94, HLL95, ABF$^+$93, Bor91**]. One of the oldest techniques is to construct a solution to the average cost optimality equation by first considering the β-discounted problem with value function

$$V_\beta(x) = \min_{\boldsymbol{w}} \mathsf{E}\Big[\sum_{k=0}^{\infty} \beta^k c(\Phi^w(k)) \mid \Phi^w(0) = x\Big]$$

where c is the one-step cost on the state space X, $\boldsymbol{\Phi}^w$ denotes the resulting state process when the policy $\boldsymbol{w}$ is applied, x is the initial condition, and the minimum is with respect to all policies. One considers the difference $h_\beta = V_\beta(x) - V_\beta(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is some distinguished state, and then seeks conditions under which h_β converges as $\beta \uparrow 1$ to a solution to the average cost optimality equations, thereby giving an optimal policy. To make this approach work, in [**Sen89**] and other papers it is

1991 *Mathematics Subject Classification.* Primary 90B35, 68M20, 90B15; Secondary 93E20, 60J20.

Work supported in part by NSF Grant ECS 940372, 1994 University of Illinois Research Board Grant *Beckman* 1-6-49749, and JSEP grant N00014-90-J-1270. This work was completed with the assistance of equipment granted through the IBM Shared University Research program and managed by the Computing and Communications Services Office at the University of Illinois at Urbana-Champaign.

assumed that there is a finite-valued function M on the state space, and a constant $N > 0$ such that

$$-N \leq h_\beta(x) \leq M(x), \tag{1.1}$$

for all states x, and all β sufficiently close to unity. Assuming bounds similar to (1.1) together with some additional conditions it is also known that policy iteration and value iteration converge to yield optimal policies [**HLL95, Sen96**], and the paper [**WS87**] uses this approach in its analysis of networks. The condition (1.1) appears to be far removed from the initial problem statement. However, it is shown in [**RS93**] that many common and natural assumptions imply (1.1).

We have shown previously in [**DM95, KM96**] that the stability of an associated fluid limit model is intimately connected with c-regularity of the stochastic network model, where $c(x) = |x|$ is the one step cost under consideration. This strong stability condition is developed in [**MT93**], where it is shown that c-regular chains satisfy several strong ergodic theorems. Moreover the associated Poisson equation, which is central to the theory of optimal average cost control, always has a solution for a c-regular chain, and this solution satisfies strict upper and lower bounds which are closely related to (1.1) (see [**GM96**] or [**MT93**, Section 17.4]). Building upon these results, this paper establishes the following structural results for the network model, whenever the usual load conditions are satisfied.

- **i:** Value iteration converges to form a "regular policy" which is optimal, and non-idling.
- **ii:** Policy iteration converges "almost monotonically" to form a regular, non-idling, optimal policy, provided that the algorithm is initiated with a policy whose fluid limit model is stable.
- **iii:** When scaled by the "fluid scaling", the policy iteration algorithm for the network approximates policy iteration for the fluid process.

An interesting conclusion of this study is that when a network is congested, the best policy is approximately equal to an optimal policy for the fluid limit model.

To give a simple concrete example, consider the model analyzed in [**Haj84**] consisting of two exponential servers. The arrival process is Poisson with rate λ, and the service rates are μ_1, μ_2, respectively. Customers that arrive to the system are routed to one of the two servers in such a way that the total cost J is minimized, where

$$J = \limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathsf{E}_x[c_1\Phi_1(k) + c_2\Phi_2(k)].$$

The weightings c_1, c_2 are assumed strictly positive. It is shown in [**Haj84**] that the optimal policy is of a nonlinear threshold form: There is a non-decreasing function $s\colon \mathbb{Z}_+ \to \mathbb{Z}_+$ such that when a job arrives when the queue lengths are x_1, and x_2, then buffer one receives the job if and only if $s(x_1) \geq x_2$. The analysis of [**XC93**] implies that the function s is unbounded, as illustrated in Figure 2. Unfortunately, in general no analytic formula is available for the computation of s.

Associated with this network is a *fluid model* evolving on $\mathbb{R}^2_+$ with state process $\phi(t) = (\phi_1(t), \phi_2(t))'$ [**Che96**]. The fluid model is defined by the equations

$$\phi(t) = \lambda \begin{pmatrix} U_1(t) \\ U_2(t) \end{pmatrix} - \begin{pmatrix} \mu_1 B_1(t) \\ \mu_2 B_2(t) \end{pmatrix}.$$

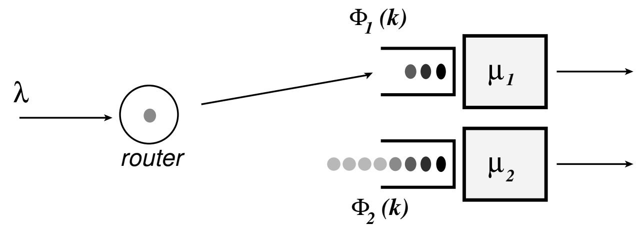

FIGURE 1. A network with controlled routing.

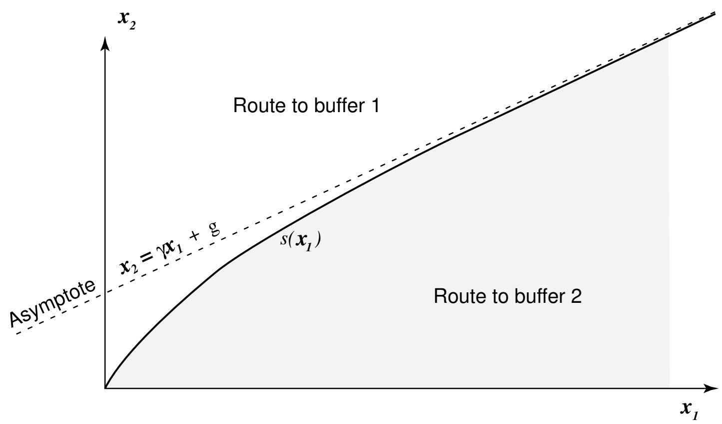

FIGURE 2. Optimal control for the two station network.

Fluid flows to the system at rate λ, whereupon it is routed to the ith buffer at rate $\frac{d}{dt}U_i(t)$. The quantity $B_i(t)$ is the cumulative busy time at buffer i. The variables $\{U_i, B_i\}$ satisfy for almost all t,

i: $U_1(t) + U_2(t) = t$;
ii: $\frac{d}{dt}B_i(t) \leq 1$, $i = 1, 2$;
iii: $\frac{d}{dt}B_i(t) = 1$ whenever $\phi_i(t) > 0$, $i = 1, 2$.

Although computation of optimal policies for stochastic networks is infeasible for all but the simplest models, optimal control of fluid models appears to be a far simpler problem [**Per93, FAR95, Wei94, DEM96**]. To minimize the total cost

$$\int_0^\infty c_1\phi_1(t) + c_2\phi_2(t)\,dt,$$

one optimal policy is of the threshold form, as illustrated in Figure 3. Hence for some "ratio-threshold" γ^*, buffer one receives full service ($\frac{d}{dt}U_1(t) = 1$) if and only if $\gamma^* x_1 \geq x_2$ [**Che96**]. Using the same ideas as in the proof of Theorem 5.2 below it may be shown that the constant γ shown in Figure 2 is equal to an optimal ratio-threshold for the fluid model. This gives the asymptote for the function s, and hence we know at least approximately the optimal policy when the network is congested.

In the remainder of the paper we develop these ideas using a re-entrant line model with an arbitrary number of buffers, and fixed routing. Below we give a network description, and in Section 3 we review previous work which establishes

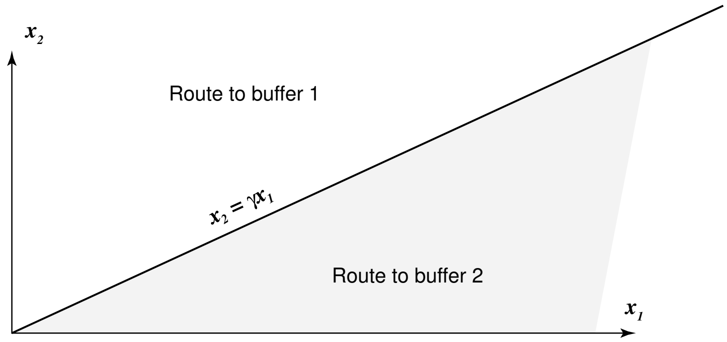

FIGURE 3. Optimal control for the fluid model.

a close connection between c-regularity of a network, and stability of a fluid limit model. These results are used in Sections 4 and 5 to prove that policy iteration and value iteration converge to form a non-idling, stationary optimal policy for the network. In Section 6 we present the results of some numerical experiments for a three-buffer re-entrant line to illustrate the way in which the optimal discrete policy converges to the optimal fluid policy, and we show in Section 7 how the policy for the fluid model may be effectively translated to form a policy for the original network of interest. The conclusion lists some topics for future research.

2. An MDP Network Model

Consider a network of the form illustrated in Figure 4, composed of d single server stations, indexed by $\sigma = 1, \dots, d$. The network is populated by K classes of customers, and an exogenous stream of customers of class 1 arrive to machine $s(1)$. A customer of class k requires service at station $s(k)$. If the service times and interarrival times are assumed to be exponentially distributed, then after a suitable time scaling and sampling of the process, the dynamics of the network can be described by the random linear system,

$$\Phi(k+1) = \Phi(k) + \sum_{i=0}^{K} I_i(k+1)[e^{i+1} - e^i]w_i^k, \tag{2.1}$$

where the state process $\mathbf{\Phi}$ evolves on $\mathsf{X} = \mathbb{Z}_+^K$ [**Lip75**]. The random variable $\Phi_i(k)$ denotes the number of class i customers in the system at time k which await service in buffer i at station $s(i)$. We assume throughout the paper that the following load conditions are satisfied

$$\rho_\sigma = \sum_{i:s(i)=\sigma} \frac{\lambda}{\mu_i} < 1, \qquad 1 \le \sigma \le d.$$

The random variables $\{I(k) : k \ge 0\}$ are i.i.d. on $\{0,1\}^{K+1}$, with $\mathsf{P}\{\sum_i I_i(k) = 1\} = 1$, and $\mathsf{E}[I_i(k)] = \mu_i$. For $1 \le i \le K$, μ_i denotes the service rate for class i customers, and for $i = 0$ we let $\mu_0 := \lambda$ denote the arrival rate of customers of class 1. For $1 \le i \le K$ we let e^i denote the ith basis vector in $\mathbb{R}^K$, and we set $e^0 = e^{K+1} := 0$. If $w_i^k I_i(k+1) = 1$, this means that at the discrete time k, a customer of class i is just completing a service, and is moving on to buffer $i+1$ or, if $i = K$, the customer then leaves the system.

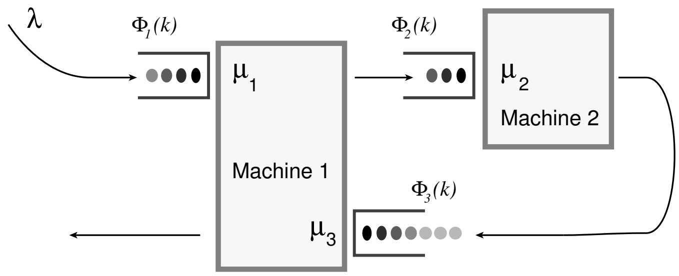

FIGURE 4. A multiclass network with $d = 2$ and $K = 3$.

The sequence $\boldsymbol{w} = \{w^k : k \geq 0\}$ is the control. Each w^k takes values in $\{0,1\}^{K+1}$, and is envisioned as a column vector $w^k = (w_0^k, w_1^k, \dots, w_K^k)'$, where we define $w_0^k \equiv 1$. In general, $\boldsymbol{w}$ will be an adapted (history dependent) stochastic process. However, we consider primarily stationary Markov policies of the form $\boldsymbol{w} = (w(\Phi(0)), w(\Phi(1)), w(\Phi(2)), \dots)$, where the state feedback law w is a mapping from X to $\{0,1\}^{K+1}$.

The set of admissible control actions $\mathcal{A}(x)$ when the state is $x \in \mathsf{X}$ is defined in an obvious manner: for $a = (a_0, \dots, a_K)' \in \mathcal{A}(x) \subset \{0,1\}^{K+1}$,

i: For any $1 \leq i \leq K$, $a_i = 0$ or 1;
ii: For any $1 \leq i \leq K$, $x_i = 0 \Rightarrow a_i = 0$;
iii: For any station σ, $0 \leq \sum_{i:s(i)=\sigma} a_i \leq 1$.

We will at times also assume

iv: For any station σ, $\sum_{i:s(i)=\sigma} a_i = 1$ whenever $\sum_{i:s(i)=\sigma} x_i > 0$.

If $a_i = 1$, then buffer i is chosen for service. Condition (ii) then imposes the physical constraint that a customer cannot be serviced at a buffer if that buffer is empty. Condition (iii) means that only one customer may be served at a given instant at a single machine σ. Condition (iv) is the *non-idling* property that a server will always work if there is work to be done.

It is evident that these specifications define an MDP whose state transition function has the simple form,

$$\begin{aligned} \mathsf{P}_a(x, x + e^{i+1} - e^i) &= \mu_i a_i, \qquad 0 \leq i \leq K. \\ \mathsf{P}_a(x, x) &= 1 - \sum_0^K \mu_i a_i. \end{aligned}$$

Since the control is bounded, a reasonable choice of cost function is $c(x) = c'x$, where $c \in \mathbb{R}^K$ is a vector with strictly positive entries. For concreteness, we take $c(x) = |x| := \sum_i x_i$. A policy $\boldsymbol{w}$ is then called optimal if the cost

$$J(x, \boldsymbol{w}) = \limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \mathsf{E}_x^w[|\Phi(k)|]$$

is minimized over all policies, for each x.

We show in Theorem 4.1 that for an optimal policy, the non-idling condition (iv) can be assumed without loss of generality. The optimal policy can also be taken to be of the stationary Markov form, $\boldsymbol{w} = (w, w, \dots)'$, where w is a function on X. One fruitful approach to computing the optimal feedback law w is through the following optimality equations:

$$\eta_* + h(x) = \min_{a\in\mathcal{A}(x)} [c(x) + P_a h\,(x)] \tag{2.2}$$

$$w(x) = \arg\min_{a\in\mathcal{A}(x)} P_a h\,(x), \qquad x \in \mathsf{X}. \tag{2.3}$$

The equality (2.2), a version of Poisson's equation, is known as the *average cost optimality equation* (ACOE), and the function h is called a *relative value function*. The second equation (2.3) defines a stationary Markov policy $\boldsymbol{w}$. If a state feedback law w, a function h, and a constant η_* exist which solve these equations, then typically the policy $\boldsymbol{w}$ is optimal (see for example [**Put94**]). In this paper we give several results which guarantee solutions to (2.2,2.3), where the function h is shown to be equivalent in some sense to a quadratic on X. These results are based primarily on the close connection between c-regularity, and the stability of a fluid limit model.

3. Regularity of the State Process and Stability of the Fluid Model

To begin an analysis of the state process $\boldsymbol{\Phi}$ we require conditions that guarantee that the state process possesses a single communicating class, so that it is ψ-irreducible, as defined in [**MT93**]. It is easily seen that for any non-idling policy there is a single communicating class: For any initial condition x, if $|x| = m$, and if the total number of buffers in the system is K, then because of the non-idling assumption (iv) it follows that the network will be empty at time mK provided that (a) no customers arrive to the network during the time interval $[0, mK]$, and (b) none of the mK services are virtual services. That is, $\sum_{i=1}^K I_i(k{+}1)w_i^k = 1$, $0 \le k < mK$. The probability of this event is bounded from below by $(1-\lambda)^{mK}\lambda^{mK}$, and hence we have for any non-idling policy $\boldsymbol{w}$,

$$P^{|x|K}_{\boldsymbol{w}}(x, \boldsymbol{\theta}) \ge s(x) := (1-\lambda)^{|x|K}\lambda^{|x|K}, \qquad \text{for all } x \in \mathsf{X}, \tag{3.1}$$

where $\boldsymbol{\theta}$ denotes the empty state $\boldsymbol{\theta} = (0, \dots, 0)' \in \mathsf{X}$. If assumption (iv) is violated then it is certainly possible that the chain may be highly non-irreducible. Consider for example the lazy server that is always idle!

We define the first entrance time and first return time to the empty state $\boldsymbol{\theta}$ by

$$\boldsymbol{\sigma}_\theta = \min(k \ge 0 : \Phi(k) = \boldsymbol{\theta}) \qquad \boldsymbol{\tau}_\theta = \min(k \ge 1 : \Phi(k) = \boldsymbol{\theta}).$$

For a particular policy $\boldsymbol{w}$, the controlled chain is called *c-regular* if for any initial condition x,

$$\mathsf{E}^{\boldsymbol{w}}_x\Big[\sum_{i=0}^{\boldsymbol{\tau}_\theta - 1} c(\Phi(i))\Big] < \infty.$$

A c-regular chain always possesses a unique invariant probability π such that

$$\pi(c) := \int c(x)\,\pi(dx) < \infty.$$

We call a stationary Markov policy $\boldsymbol{w}$ *regular* if the associated controlled chain is c-regular. The following result is a consequence of the f-norm ergodic theorem of [**MT93**, Chapter 14].

THEOREM 3.1. *For any regular stationary Markov policy* $\boldsymbol{w}$*, the controlled state process* $\boldsymbol{\Phi}$ *satisfies* $\int |x|\,\pi(dx) < \infty$*, and for each initial condition*

$$\textbf{\textit{(i)}} \qquad J(x,\boldsymbol{w}) = \lim_{k\to\infty} \mathsf{E}_x[|\Phi(k)|] \;=\; \int |x|\,\pi(dx).$$

$$\textbf{\textit{(ii)}} \qquad \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} |\Phi(k)| \;=\; \int |x|\,\pi(dx), \qquad a.s.\ [\mathsf{P}_x].$$

□

It is now known that stability of (2.1) in terms of c-regularity is closely connected with the stability of an associated fluid limit model [**Dai95, KM96, DM95, Mey95a**]. Suppose that a stationary Markov policy $\boldsymbol{w}$ is given so that the controlled chain $\boldsymbol{\Phi}$ is a Markov chain with stationary transition probabilities. For each initial condition $\Phi(0) = x \neq \boldsymbol{\theta}$, we construct a continuous time process $\phi^x(t)$ as follows. If $|x|t$ is an integer, set

$$\phi^x(t) = \frac{1}{|x|}\Phi(|x|t).$$

For all other $t \geq 0$, define $\phi^x(t)$ by linear interpolation, so that it is continuous and piecewise linear in t. Note that $|\phi^x(0)| = 1$, and that ϕ^x is Lipschitz continuous. The collection of all "fluid limits" is defined by

$$\mathcal{L} := \bigcap_{n=1}^{\infty} \overline{\{\phi^x : |x| > n\}}$$

where the overbar denotes weak closure. The process ϕ evolves on the state space $\mathbb{R}_+^K$. We shall also call $\mathcal{L}$ the *fluid limit model*, in contrast to the *fluid model* which is defined as the set of all solutions to the differential equations

$$\frac{d}{dt}\phi_i(t) = \sum_{i=0}^{K} \mu_i[e^{i+1} - e^i]u_i(t), \qquad \text{a.e. } t \in \mathbb{R}_+, \tag{3.2}$$

where the function $u(t)$ is analogous to the discrete control, and satisfies similar constraints [**CM91**].

The fluid limit model $\mathcal{L}$ is called L_p*-stable* if

$$\lim_{t\to\infty} \sup_{\phi\in\mathcal{L}} \mathsf{E}[|\phi(t)|^p] = 0.$$

There are now numerous papers which derive stability conditions for specific fluid network models (see for instance [**DV96, DBT95, CZ96, DW96, DM95, KK94**]). In particular, the recent paper [**CZ96**] develops a unified approach to stability which is used to show that LBFS and FBFS (last buffer-first served and first buffer-first served) are L_p-stable for any p. It is shown in [**KM96**] that L_2-stability of the fluid limit model is equivalent to a form of c-regularity for the network. The following result strengthens this equivalence by linking L_2 stability to boundedness of the total cost for the fluid limit model.

THEOREM 3.2. *The following stability criteria are equivalent for the network under any nonidling, stationary Markov policy* $\boldsymbol{w}$*.*

i: *The drift condition holds*

$$P_w V\,(x) := \mathsf{E}_x^{\boldsymbol{w}}[V(\Phi(k+1)) \mid \Phi(k) = x] \leq V(x) - |x| + \kappa, \qquad x \in \mathsf{X}, \tag{3.3}$$

where $\kappa \in \mathbb{R}_+$, and the function $V\colon \mathsf{X} \to \mathbb{R}_+$ is equivalent to a quadratic in the sense that, for some $\gamma > 0$,

$$1 + \gamma |x|^2 \le V(x) \le 1 + \gamma^{-1}|x|^2, \qquad x \in \mathsf{X}.$$

ii: *For some quadratic function V,*

$$\mathsf{E}_x^{\boldsymbol{w}}\Big[\sum_{n=0}^{\boldsymbol{\sigma}_\theta} |\Phi(n)|\Big] \quad \le \quad V(x), \qquad x \in \mathsf{X}.$$

iii: *For some quadratic function V and some $\gamma < \infty$,*

$$\sum_{n=1}^{N} \mathsf{E}_x^{\boldsymbol{w}}[|\Phi(n)|] \quad \le \quad V(x) + \gamma N, \quad \text{for all } x \text{ and } N \ge 1.$$

iv: *The fluid limit model $\mathcal{L}$ is L_2-stable.*

v: *The total cost for the fluid limit $\mathcal{L}$ is uniformly bounded in the sense that*

$$\sup_{\phi \in \mathcal{L}} \mathsf{E}\Big[\int_0^\infty |\phi(\tau)|\, d\tau\Big] < \infty.$$

Proof. The equivalence of the first four statements has been established in [**KM96**]. To complete the proof we first show that (iii) implies (v), and then prove that (v) implies (iv).

For any policy $\boldsymbol{w}$ and any time $T > 0$ we have the approximation

$$\frac{1}{|x|}\mathsf{E}_x^{\boldsymbol{w}}\Big[\sum_{k=0}^{\lfloor T|x|\rfloor} \frac{|\Phi(k)|}{|x|}\Big] = \mathsf{E}_x^{\boldsymbol{w}}\Big[\int_0^T |\phi^x(\tau)|\, d\tau\Big] + o(1), \tag{3.4}$$

where the term $o(1)$ vanishes as $x \to \infty$. Using (iii) it then follows that for any $\phi \in \mathcal{L}$,

$$\mathsf{E}^{\boldsymbol{w}}\Big[\int_0^\infty |\phi(\tau)\, d\tau\Big] \le \limsup_{T\to\infty} \limsup_{|x|\to\infty} \mathsf{E}_x^{\boldsymbol{w}}\Big[\int_0^T |\phi^x(\tau)|\, d\tau\Big] \le \sup_{x \ne \theta} \frac{V(x)}{|x|^2} < \infty.$$

Taking the supremum over all ϕ proves the implication (iii) $\Rightarrow$ (v).

Assume now that (v) holds. To deduce (iv) we will apply bounds on the conditional expection of functionals of the process $\boldsymbol{\Phi}$ and the fluid limit ϕ. Define for $s \ge 0$, $k \ge 0$, the σ-algebras

$$\mathcal{G}_s = \sigma\{\phi(r) : r \le s\} \qquad \mathcal{F}_k = \sigma\{\Phi(i) : i \le k\}.$$

These bounds concern mainly the super martingale $M(s)$ defined as

$$M(s) = \mathsf{E}^{\boldsymbol{w}}\Big[\int_s^\infty |\phi(\tau)|\, d\tau \mid \mathcal{G}_s\Big], \qquad s \ge 0. \tag{3.5}$$

By Lipschitz continuity of the model we have the lower bound

$$\tfrac{1}{2}|\phi(s)|^2 \le M(s), \qquad s \ge 0. \tag{3.6}$$

By assumption there is a constant b_0 such that

$$M(0) = \mathsf{E}^{\boldsymbol{w}}\Big[\int_0^\infty |\phi(\tau)|\, d\tau\Big] \le b_0 \tag{3.7}$$

for any $\phi \in \mathcal{L}$, and by dominated convergence $\mathsf{E}^{\boldsymbol{w}}[M(s)] \to 0$ as $s \to \infty$. Hence, from the lower bound (3.6),

$$\mathsf{E}^{\boldsymbol{w}}[|\phi(s)|^2] \to 0, s \to \infty, \qquad \text{for any } \phi \in \mathcal{L}. \tag{3.8}$$

To establish (iv) it is only necessary to show that this convergence is uniform over all $\phi \in \mathcal{L}$. For this we require a uniform upper bound on $M(s)$.

For any $s \geq 0$, $t > 0$ we have the generalizations of (3.4):

$$\frac{1}{|x|^2} \sum_{k=\lfloor s|x| \rfloor}^{\lfloor (s+t)|x| \rfloor} |\Phi(k)| = \int_s^{s+t} |\phi^x(\tau)|\, d\tau + o(1). \tag{3.9}$$

$$\frac{\Phi(\lfloor s|x| \rfloor)}{|x|} = \phi^x(s) + o(1). \tag{3.10}$$

To maintain bounded random variables we fix $\varepsilon > 0$ and define a continuous approximation to the indicator function $\mathbb{1}_{(0,\infty)}$ as follows:

$$\delta_\varepsilon(z) = \begin{cases} 1, & z \leq \varepsilon; \\ 0, & z \geq \varepsilon/2; \\ -1 + 2z/\varepsilon, & \varepsilon/2 \leq z \leq \varepsilon; \end{cases}, \qquad z \in \mathbb{R}_+.$$

Letting $y = \Phi(\lfloor s|x| \rfloor)$, from (3.9) we have

$$\begin{aligned} \mathsf{E}_x^{\boldsymbol{w}}\Big[\int_s^{s+t} |\phi^x(\tau)|\, d\tau \mid \mathcal{F}_{\lfloor |x| s \rfloor}\Big]\delta_\varepsilon(\phi^x(s)) + o(1) \\ = |\phi^x(s)|^2 \delta_\varepsilon(\phi^x(s)) \frac{1}{|y|^2} \mathsf{E}^{\boldsymbol{w}}\Big[\sum_{k=\lfloor s|x| \rfloor}^{\lfloor (s+t)|x| \rfloor} |\Phi(k)| \mid \mathcal{F}_{\lfloor |x| s \rfloor}\Big] \\ \leq |\phi^x(s)|^2 \delta_\varepsilon(\phi^x(s)) \frac{1}{|y|^2} \mathsf{E}^{\boldsymbol{w}}\Big[\sum_{k=0}^{\lfloor |y| (2t/\varepsilon) \rfloor} |\Phi(\lfloor s|x| \rfloor + k)| \mid \mathcal{F}_{\lfloor |x| s \rfloor}\Big] \end{aligned} \tag{3.11}$$

where we have used the fact that $|y| > \varepsilon |x|/2 + o(1)$ whenever $\delta_\varepsilon(\phi^x(s)) > 0$.

Let $\varepsilon_1 > 0$ be a second small constant. Using (3.7) and (3.4), we may find a constant $n = n(\varepsilon, t) > 0$ such that whenever $|y| \geq n$,

$$\frac{1}{|y|^2} \mathsf{E}_y^{\boldsymbol{w}}\Big[\sum_{k=0}^{|y|(2t/\varepsilon)} |\Phi(k)|\Big] \leq b_0 + \varepsilon_1$$

Assuming that $|x| \geq (2t/\varepsilon)n$, we then have $|y| \geq n + o(1)$ whenever $\delta_\varepsilon(\phi^x(s)) > 0$, so that by the Markov property,

$$\begin{aligned} \delta_\varepsilon(\phi^x(s)) \frac{1}{|y|^2} \mathsf{E}_x^{\boldsymbol{w}}\Big[\sum_{k=0}^{\lfloor |y| (2t/\varepsilon) \rfloor} |\Phi(\lfloor s|x| \rfloor) + k)| \mid \mathcal{F}_{\lfloor |x| s \rfloor}\Big] \\ = \delta_\varepsilon(\phi^x(s)) \frac{1}{|y|^2} \mathsf{E}_y^{\boldsymbol{w}}\Big[\sum_{k=0}^{\lfloor |y| (2t/\varepsilon) \rfloor} |\Phi(k)|\Big] \\ \leq \delta_\varepsilon(\phi^x(s))(b_0 + \varepsilon_1). \end{aligned} \tag{3.12}$$

Combining (3.11) and (3.12) then gives for $|x| \geq (2t/\varepsilon)n$,

$$\mathsf{E}_x^{\boldsymbol{w}}\Big[\int_s^{s+t} |\phi^x(\tau)|\, d\tau \mid \mathcal{F}_{\lfloor |x| s \rfloor}\Big]\delta_\varepsilon(\phi^x(s)) \leq |\phi^x(s)|^2 \delta_\varepsilon(\phi^x(s))(b_0 + \varepsilon_1) + o(1).$$

In the proof of Thereom 16 of [**KM96**] it is shown using standard conditioning arguments that bounds of this form on continuous functionals of ϕ^x pass to the limit as $x \to \infty$ to form a bound for ϕ. Hence for any $\phi \in \mathcal{L}$,

$$\mathsf{E}^{\boldsymbol{w}}\Big[\int_s^{s+t} |\phi(\tau)\, d\tau \mid \mathcal{G}_s\Big]\delta_\varepsilon(\phi(s)) \le |\phi(s)|^2\delta_\varepsilon(\phi(s))(b_0+\varepsilon_1).$$

By first letting $\varepsilon_1 \downarrow 0$, then letting $\varepsilon \downarrow 0$, and finally letting $t \uparrow \infty$ we arrive at the bound

$$M(s)\mathbb{1}_{(0,\infty)}(\phi(s)) \le b_0|\phi(s)|^2.$$

To remove the indicator function, note first that since $\phi(s)$ is a continuous function of s, the bound above implies that

$$M(s) \le b_0|\phi(s)|^2 \qquad \text{whenever } M(s) > 0.$$

Since this bound is vacuous whenever $M(s) = 0$, we obtain the desired upper bound

$$M(s) \le b_0|\phi(s)|^2, \qquad s \ge 0. \tag{3.13}$$

We can now show that (v) $\Rightarrow$ (iv) by contradiction. Suppose that $\mathcal{L}$ is not L_2 stable. Then there exist $\varepsilon > 0$, $\{\phi_n\} \subset \mathcal{L}$, and a sequence of times $t_n \to \infty$ such that $\mathsf{E}^{\boldsymbol{w}}[|\phi_n(t_n)|^2] \ge \varepsilon$ for all $n > 0$. Let $\{x_i^n : i, n \ge 0\}$ be a set of states such that $\phi^{x_i^n} \xrightarrow{\text{w}} \phi_n$ as $i \to \infty$ for each n. Denoting by M_n the super martingale corresponding to ϕ_n we thus have $\mathsf{E}^{\boldsymbol{w}}[M_n(t_n)] \ge \frac{1}{2}\varepsilon$, and since M_n is decreasing,

$$\mathsf{E}^{\boldsymbol{w}}[|\phi_n(s)|^2] \ge \frac{1}{b_0}\mathsf{E}^{\boldsymbol{w}}[M_n(t_n)] \ge \frac{1}{2b_0}\varepsilon, \qquad 0 \le s \le t_n.$$

From weak convergence it follows that we may choose a new sequence $\{y^n\}$ among the $\{x_i^n\}$ such that

$$\mathsf{E}^{\boldsymbol{w}}_{y^n}[|\phi^{y^n}(s)|^2] \ge \frac{1}{4b_0}\varepsilon, \qquad 0 \le s \le n.$$

Assumming without loss of generality that $\phi^{y^n} \xrightarrow{\text{w}} \phi$ for some $\phi \in \mathcal{L}$ we conclude that for *all* s,

$$\mathsf{E}^{\boldsymbol{w}}[|\phi(s)|^2] \ge \frac{1}{4b_0}\varepsilon.$$

This contradicts (3.8) and thereby shows that (v) $\Rightarrow$ (iv) as claimed. $\square$

We now refine this result by showing that in fact $\phi(t) = \boldsymbol{\theta}$ for all t sufficiently large if the fluid limit $\mathcal{L}$ is L_2 stable. Let $\boldsymbol{\tau}_0$ denote the first time that this occurs:

$$\boldsymbol{\tau}_0 := \min(s : \phi(s) = \boldsymbol{\theta}).$$

As motivation, first consider the special case of a stable M/M/1 queue where the service rate μ is greater than the arrival rate λ. The fluid limit model is L_2 stable in this special case since $\mathcal{L} = \{\phi\}$ is a singleton with $\phi(t) = (1-(\mu-\lambda)t) \vee 0$. The fluid limit model thus satisfies

i: *The total cost for the fluid limit model is uniformly bounded*:

$$b_0 := \int_0^\infty |\phi(\tau)\, d\tau = \frac{1}{2}\frac{1}{\mu-\lambda} < \infty.$$

ii: *The hitting time to $\boldsymbol{\theta}$ is bounded*: $\boldsymbol{\tau}_0 = 2b_0$, where b_0 is as in (i).

Exact calculations are not possible in general, but we can generalize (i) and (ii) as bounds:

THEOREM 3.3. *Suppose that the fluid limit model is L_2-stable for the nonidling, stationary Markov policy $\boldsymbol{w}$. Then*

i: *The total cost for the fluid limit model is uniformly bounded:*

$$b_0 := \sup_{\phi\in\mathcal{L}} \mathsf{E}^{\boldsymbol{w}}\Big[\int_0^\infty |\phi(\tau)\, d\tau\Big] < \infty$$

ii: *The mean hitting time to $\boldsymbol{\theta}$ is uniformly bounded:*

$$\sup_{\phi\in\mathcal{L}} \mathsf{E}^{\boldsymbol{w}}[\boldsymbol{\tau}_0(\phi)] \le 2b_0,$$

where b_0 is defined in (i).

iii: *The origin $\boldsymbol{\theta}$ is absorbing in the sense that with probability one $\phi(t) = \boldsymbol{\theta}$ for any $\phi \in \mathcal{L}$ and all $t \ge \boldsymbol{\tau}_0$.*

PROOF. The bound (i) has already been demonstrated in Theorem 3.2.

It is shown in equation (3.13) of the proof of Theorem 3.2 that when $\mathcal{L}$ is L_2 stable,

$$M(t) \le b_0|\phi(t)|^2, \qquad t \ge 0,$$

where the constant b_0 is defined in (i). Using this bound and the definition of M it follows that if $\phi(t) = \boldsymbol{\theta}$ for some t, then $\phi(s) = \boldsymbol{\theta}$ for all $s > t$. This establishes (iii).

To obtain the bound (ii) we construct a new super martingale by taking the square root: $V(t) := \sqrt{M(t)}$. The previous bound on $M(t)$ gives

$$V(t) \le \sqrt{b_0}|\phi(t)| \tag{3.14}$$

By concavity of the square root we have, whenever $V(s) \ne 0$,

$$\mathsf{E}^{\boldsymbol{w}}[V(s+t) \mid \mathcal{G}_s] \le V(s) - \frac{1}{2V(s)}\mathsf{E}^{\boldsymbol{w}}\Big[\int_s^{s+t} |\phi(\tau)|\, d\tau \mid \mathcal{G}_s\Big]. \tag{3.15}$$

To bound the negative term, note that for all t,

$$|\phi(s+t)| \ge |\phi(s)| - t. \tag{3.16}$$

This follows from the fact that $|\frac{d}{dt}\phi(t)| \le 1$. Letting $0 \le \varepsilon \le \frac{1}{2}$, the bounds (3.14), (3.15), and (3.16) together imply that whenever $|\phi(s)| \ge \varepsilon$ and $t < 2\varepsilon^2$,

$$\begin{aligned} \mathsf{E}^{\boldsymbol{w}}[V(s+t) \mid \mathcal{G}_s] &\le V(s) - \left(\frac{t|\phi(s)| - \frac{1}{2}t^2}{2\sqrt{b_0}|\phi(s)|}\right) \\ &\le V(s) - \left(\frac{t|\phi(s)| - \frac{1}{2}t(2\varepsilon|\phi(s)|)}{2\sqrt{b_0}|\phi(s)|}\right) \\ &\le V(s) - \left(\frac{1-\varepsilon}{2\sqrt{b_0}}\right)t. \end{aligned} \tag{3.17}$$

For $k \ge 0$ let $T_k = 2k\varepsilon^2$, and let $\boldsymbol{\tau}_\varepsilon$ denote the hitting time

$$\boldsymbol{\tau}_\varepsilon := \min(s : |\phi(s)| \le \varepsilon).$$

Then from (3.17) we have for any k,

$$\begin{aligned} \mathsf{E}_x^{\boldsymbol{w}}[V(T_{k+1})\mathbb{1}(\boldsymbol{\tau}_\varepsilon \ge T_{k+1}) \mid \mathcal{G}_{T_k}] &\le \mathsf{E}_x^{\boldsymbol{w}}[V(T_{k+1})\mathbb{1}(\boldsymbol{\tau}_\varepsilon \ge T_k) \mid \mathcal{G}_{T_k}] \\ &\le V(T_k)\mathbb{1}(\boldsymbol{\tau}_\varepsilon \ge T_k) - \left(\frac{1-\varepsilon}{2\sqrt{b_0}}\right)2\varepsilon^2\mathbb{1}(\boldsymbol{\tau}_\varepsilon \ge T_k) \end{aligned}$$

Summing over $k = 0$ to n, and using the smoothing property of the conditional expectation shows that for any n,

$$\mathsf{E}_x^{\boldsymbol{w}}\Big[\sum_{k=0}^{n} 2\varepsilon^2 \mathbb{1}(\boldsymbol{\tau}_\varepsilon \geq T_k)\Big] \leq \frac{2\sqrt{b_0}}{1-\varepsilon}\mathsf{E}[V(0)\mathbb{1}(\boldsymbol{\tau}_\varepsilon \geq T_0)] \leq \frac{2b_0}{1-\varepsilon},$$

where in the second inequality we have used (3.14). For any $\varepsilon > 0$ we have the staircase bound

$$\sum_{k=0}^{\infty} 2\varepsilon^2 \mathbb{1}(\boldsymbol{\tau}_\varepsilon \geq T_k) = \sum_{k=0}^{\infty} 2\varepsilon^2 \mathbb{1}(\boldsymbol{\tau}_\varepsilon \geq 2\varepsilon^2 k) \geq \boldsymbol{\tau}_\varepsilon.$$

Hence by the Monotone Convergence Theorem,

$$\mathsf{E}_x^{\boldsymbol{w}}[\boldsymbol{\tau}_\varepsilon] \leq \frac{2b_0}{1-\varepsilon}.$$

Letting $\varepsilon \downarrow 0$ and applying the Monotone Convergence Theorem once more gives the bound $\mathsf{E}_x^{\boldsymbol{w}}[\boldsymbol{\tau}_0] \leq 2b_0$. Since $\phi \in \mathcal{L}$ is arbitrary, this establishes (ii). ☐

The exponential assumption is not crucial in any of these stability results. In [**DM95**] the case of general distributions is developed, and analogous regularity results are obtained when a fluid limit model is L_2-stable. Given these structural results we can now prove that an optimal policy exists which is non-idling.

4. Convergence of Value Iteration

Value iteration is perhaps the most common approach to constructing an optimal policy. The idea is to consider the finite time problem with value function

$$V_n(x) = \min_{\boldsymbol{w}} \mathsf{E}_x^{\boldsymbol{w}}\Big[\sum_{k=0}^{n} |\Phi(k)|\Big].$$

We let $\boldsymbol{v}^n$ denote a policy which attains this minimum - It may be assumed without loss of generality that there is a sequence of state feedback functions $v^k : \mathsf{X} \to \{0,1\}^{K+1}$, $k \geq 0$, such that for any n, the policy $\boldsymbol{v}^n$ is a Markov policy which may be expressed

$$\boldsymbol{v}^n = (v^{n-1}, \dots, v^0)'.$$

The value iteration algorithm is then defined inductively as follows. If the value function V_n is given, the action $v^n(x)$ is defined as

$$v^n(x) = \mathop{\arg\min}_{a \in \mathcal{A}(x)} P_a V_n\,(x).$$

For each n the following dynamic programming equation is satisfied,

$$V_{n+1}(x) = |x| + P_{v^n(x)} V_n(x) = |x| + \min_{a \in \mathcal{A}(x)} P_a V_n\,(x),$$

which then makes it possible to compute the next function v^{n+1}. Conditions under which the algorithm converges are developed in [**Sen96, CF95**] for general MDPs under various conditions on the process. The simultaneous Lyapunov function condition of [**CF95**] is unfortunately not suited to network models of the form considered here. We instead obtain bounds related to those assumed in [**Sen96**] to establish convergence. To begin, we require the following weak monotonicity properties for the value functions.

LEMMA 4.1. *For the network model with one step cost $c(x) = |x|$, the value functions $\{V_n\}$ satisfy for each n,*

i: *For any $1 \le m_0 \le K$, and any $x \in \mathsf{X}$ for which $x_{m_0} > 0$,*

$$V_n(x) \ge V_n(x + e^{m_0+1} - e^{m_0});$$

ii: *For any x,*

$$V_n(x) \ge V_n(\boldsymbol{\theta}).$$

PROOF. To see (i), let $\boldsymbol{v}^n = (v^{n-1}, \dots, v^0)$ be an optimal n-stage policy, and consider any sample path of the process of indicator variables $\{I(1), \dots, I(i), \dots\}$. Let $a^k = v^{n-1-k}(\Phi(k))$, $k = 0, \dots, n-1$ denote the kth action taken along this fixed sample path when the initial condition is x, and define the stopping time $\boldsymbol{\tau}$ as the first time ℓ such that $a^\ell_{m_0} = 1$.

For the initial condition $x - e^{m_0+1} + e^{m_0}$, define the new action sequence $\bar{\boldsymbol{a}}$ as $\bar{a}^\ell_i = a^\ell_i \mathbb{1}(\ell \ne \boldsymbol{\tau})$, $1 \le i \le K$. The new policy idles at time $\boldsymbol{\tau}$, and since we are examining a finite time window, the action sequence is unchanged if $\boldsymbol{\tau} > n$. It is immediate that the accumulated cost $\sum_0^n |\Phi(k)|$ cannot increase with the initial condition x replaced by $x - e^{m_0+1} + e^{m_0}$, when the sequence $\boldsymbol{a}$ is simultaneously replaced by $\bar{\boldsymbol{a}}$. Letting $\bar{\boldsymbol{v}}$ denote the history dependent policy defined through this transformation, we have

$$V_n(x) = \mathsf{E}^{\boldsymbol{v}}_x\Big[\sum_{k=0}^n |\Phi(k)|\Big] \ge \mathsf{E}^{\bar{\boldsymbol{v}}}_{x-e^{m_0+1}+e^{m_0}}\Big[\sum_{k=0}^n |\Phi(k)|\Big] \ge V_n(x - e^{m_0+1} + e^{m_0}),$$

which establishes (i). The bound (ii) is an immediate consequence of (i). □

To approximate a solution to the ACOE (2.2,2.3), define the function h_n by

$$h_n(x) = V_n(x) - V_n(\boldsymbol{\theta}), \qquad x \in \mathsf{X}, n \ge 1.$$

Using Theorem 3.2 and Lemma 4.1 we can easily prove the following uniform bounds on the functions $\{h_n\}$.

LEMMA 4.2. *There is a constant b such that for any x and any n,*

$$0 \le h_n(x) \le b|x|^2.$$

PROOF. The lower bound on h_n is immediate, from Lemma 4.1.

To establish an upper bound, construct a history dependent policy $\bar{\boldsymbol{w}}^n$ on the time window $[0, n + \boldsymbol{\sigma}_\theta \wedge n]$ as follows, based upon the optimal n-stage policy $\boldsymbol{v}^n$, and an arbitrary stationary Markov policy $\widetilde{\boldsymbol{w}}^n = (\widetilde{w}, \widetilde{w}, \dots)$:

$$\bar{w}_k = \begin{cases} \widetilde{w}(\Phi(k)) & k \le \boldsymbol{\sigma}_\theta; \\ v^{n-1-k+\boldsymbol{\sigma}_\theta}(\Phi(k)) & k > \boldsymbol{\sigma}_\theta. \end{cases}$$

the policy is unchanged on $[0,n]$ if $\boldsymbol{\sigma}_\theta \ge n$. We then have

$$\begin{aligned} V_n(x) &\le \mathsf{E}_x^{\bar{\boldsymbol{w}}}\Big[\sum_{k=0}^{n} |\Phi(k)|\Big] \\ &\le \mathsf{E}_x^{\bar{\boldsymbol{w}}}\Big[\sum_{k=0}^{n\wedge\boldsymbol{\sigma}_\theta} |\Phi(k)|\Big] + \mathsf{E}_x^{\bar{\boldsymbol{w}}}\Big[\mathbb{1}\{\boldsymbol{\sigma}_\theta < n\} \sum_{k=\boldsymbol{\sigma}_\theta}^{n} |\Phi(k)|\Big] \\ &\le \mathsf{E}_x^{\widetilde{\boldsymbol{w}}}\Big[\sum_{k=0}^{\boldsymbol{\sigma}_\theta} |\Phi(k)|\Big] + \mathsf{E}_x^{\bar{\boldsymbol{w}}}\Big[\sum_{k=\boldsymbol{\sigma}_\theta}^{\boldsymbol{\sigma}_\theta+n} |\Phi(k)|\Big] \end{aligned}$$

It follows from the strong Markov property that

$$V_n(x) \le \mathsf{E}_x^{\widetilde{\boldsymbol{w}}}\Big[\sum_{k=0}^{\boldsymbol{\sigma}_\theta} |\Phi(k)|\Big] + V_n(\boldsymbol{\theta})$$

To obtain a quadratic bound on h_n, choose for $\widetilde{\boldsymbol{w}}$ any stationary Markov policy for which the associated fluid limit model is L_2 stable. By Theorem 3.2 (ii) we then have for a quadratic function V,

$$h_n(x) = V_n(x) - V_n(\boldsymbol{\theta}) \le \mathsf{E}_x^{\widetilde{\boldsymbol{w}}}\Big[\sum_{k=0}^{\boldsymbol{\sigma}_\theta} |\Phi(k)|\Big] \le V(x),$$

which completes the proof of the lemma. □

Using ideas similar to [**Sen96**] we can then prove the following result. We suspect that in fact $\{h_n\}$ is convergent, but this remains open.

THEOREM 4.3. *Let (h, v) be any limit point of the sequence $\{h_n, v^n\}$ generated by the value iteration algorithm, and let η_* denote*

$$\eta_* = \limsup_{n\to\infty} V_n(\boldsymbol{\theta})/n.$$

Then the triplet (h, v, η^*) solves the optimality equations (2.2,2.3), and the stationary Markov policy $\boldsymbol{v} = (v, v, \dots)$ is average cost optimal. □

From this result we recover a special case of the existence result of [**WS87**], and we also find that property (iv) of the control actions can be assumed without loss of generality.

COROLLARY 4.1. The network model possesses a non-idling optimal stationary Markov policy.

PROOF. Using Lemma 4.1 and the formula

$$\begin{aligned} V_{n+1}(x) &= |x| + \lambda V_n(x + e^1) + \sum_{j=1}^{K} \mu_j V_n(x + (e^{j+1} - e^j) v_j^n(x)) \\ &= |x| + \lambda V_n(x + e^1) + \min_{a \in \mathcal{A}(x)} \sum_{j=1}^{K} \mu_j V_n(x + (e^{j+1} - e^j) a_j) \end{aligned}$$

one may show by induction that each of the optimal policies $\{\boldsymbol{v}^n\}$ can be assumed to be non-idling. The result then follows from Theorem 4.3. □

5. Policy Iteration for Networks and their Fluid Models

We continue to investigate the scheduling problem here under assumptions (i)–(iv) on the action space. In view of Corollary 4.1, the non-idling condition (iv) may be assumed without loss of generality. This condition is crucial in what follows since we require the lower bound (3.1).

The policy iteration algorithm is another procedure which in principle can be used to compute an optimal policy. Suppose that a stationary Markov policy $\boldsymbol{w}^{n-1} = (w^{n-1}, w^{n-1}, \dots)$ is given, and assume that h_{n-1} satisfies the Poisson equation,

$$P_{w^{n-1}(x)} h_{n-1}(x) = h_{n-1}(x) - |x| + \eta_{n-1},$$

where η_{n-1} is a constant (presumed to be the steady state cost with this policy). The relative value functions $\{h_n\}$ are not uniquely defined: If h_n satisfies Poisson's equation, then so does $h_n + b$ for any constant b. However, if the controlled chain is c-regular we define h_n uniquely as

$$h_n(x) = \mathsf{E}_x^{\boldsymbol{w}^n}\Big[\sum_{i=0}^{\boldsymbol{\tau}_\theta - 1} (c(\Phi(i)) - \pi(c))\Big].$$

Given h_{n-1}, one then attempts to find an improved stationary Markov policy $\boldsymbol{w}^n$ by choosing, for each x,

$$w^n(x) = \underset{a\in\mathcal{A}(x)}{\arg\min}\, P_a h_{n-1}(x). \tag{5.1}$$

Once $\boldsymbol{w}^n$ is found, stationary Markov polices $\boldsymbol{w}^{n+1}, \boldsymbol{w}^{n+2}, \dots$ may be computed by induction, so long as the appropriate Poisson equation may be solved, and the minimization above has a solution. Sample path versions of this algorithm have been developed recently in [**Cao96**].

Theorem 5.1 shows that policy improvement always converges for this network model if the initial policy is stabilizing. To initiate the algorithm, one can choose one of the known stabilizing policies such as last buffer-first served or first buffer-first served [**CZ96, DW96, KK94**], or one may choose a policy which is *optimal* for the fluid model.

THEOREM 5.1 (Meyn [**Mey95b**]). *If the initial stationary Markov policy $\boldsymbol{w}^0$ is chosen so that the fluid limit model is L_2 stable, then the PIA produces successive stationary Markov policies $\{\boldsymbol{w}^n\}$ and associated relative value functions $\{h_n\}$ which satisfy the following properties:*

i: *For some constant $0 < N < \infty$,*

$$\inf_{x\in\mathsf{X},\, n\geq 0} h_n(x) > -N;$$

ii: *(Almost decreasing property): There exists a sequence of functions $\{g_n : n \geq 0\}$ such that*

$$g_n(x) \leq g_{n-1}(x) \leq \cdots \leq g_0(x), \qquad x \in \mathsf{X},\ n \geq 0,$$

and for some sequence of positive numbers $\{\alpha_k, \beta_k\}$,

$$g_n(x) = \alpha_n h_n(x) + \beta_n, \qquad n \geq 0,\ x \in \mathsf{X},$$

with $\alpha_k \downarrow 1$, $\beta_k \downarrow 0$ as $k \to \infty$.

Hence, the relative value functions are pointwise convergent to the function $h(x) := \lim_n g_n(x)$.

iii: *The costs are decreasing* $\eta_0 \geq \eta_1 \geq \cdots \geq \eta_m \geq \cdots$.
iv: *Any pointwise limit point* $\boldsymbol{w}$ *of the* $\{\boldsymbol{w}^n\}$ *is an optimal stationary Markov policy.*
v: *Each of the policies* $\{\boldsymbol{w}^n\}$ *and any limiting policy* $\boldsymbol{w}$ *possesses a fluid limit model which is* L_2*-stable.*

□

From this result we can establish a strong connection between optimization of the network, and optimization of the fluid model under an L_1 optimal control criterion. A policy $\boldsymbol{w}^*$ is called *optimal for the fluid limit model* if for any other policy $\boldsymbol{w}$ which is L_2 stable,

$$\liminf_{T\to\infty} \liminf_{|x|\to\infty} \Bigl(\mathsf{E}_x^{\boldsymbol{w}}\Bigl[\int_0^T |\phi^x(\tau)|\,d\tau\Bigr] - \mathsf{E}_x^{\boldsymbol{w}^*}\Bigl[\int_0^T |\phi^x(\tau)|\,d\tau\Bigr]\Bigr) \geq 0.$$

Since by Theorem 3.2 L_2-stability is equivalent to uniform boundedness of the total cost, this stability condition can be assumed without any real loss of generality. If the paths of the fluid limit model are purely deterministic, this form of optimality amounts to minimality of the total cost

$$\int_0^\infty |\phi(\tau)\,d\tau.$$

Currently, there is much interest in directly addressing methods for the synthesis of optimal policies for fluid network models with this cost criterion [**FAR95, Wei94, DEM96**].

The result below establishes the desired connection between the optimal control of the network and its associated fluid limit model.

Theorem 5.2 (Meyn [**Mey95b**]). *If the initial policy* $\boldsymbol{w}^0$ *is chosen so that the fluid limit model is* L_2 *stable, then*

i: *For each* $n \geq 1$,

$$\liminf_{T\to\infty} \liminf_{|x|\to\infty} \Bigl(\mathsf{E}_x^{\boldsymbol{w}^{n-1}}\Bigl[\int_0^T |\phi^x(\tau)|\,d\tau\Bigr] - \mathsf{E}_x^{\boldsymbol{w}^n}\Bigl[\int_0^T |\phi^x(\tau)|\,d\tau\Bigr]\Bigr) \geq 0.$$

Hence if $\boldsymbol{w}^0$ *is optimal for the fluid limit model, so is* $\boldsymbol{w}^n$ *for any* n.
ii: *Any optimal policy* $\boldsymbol{w}^*$ *generated by the algorithm is also optimal for the fluid limit model.*
iii: *With* h *equal to the relative value function for any optimal policy* $\boldsymbol{w}^*$ *generated by the algorithm,*

$$\limsup_{T\to\infty} \limsup_{|x|\to\infty} \Bigl|\frac{h(x)}{|x|^2} - \mathsf{E}_x^{\boldsymbol{w}^*}\Bigl[\int_0^T |\phi^x(\tau)|\,d\tau\Bigr]\Bigr| = 0.$$

Hence, when properly normalized, the relative value function approximates the value function for the fluid model control problem.

□

6. Convergence to the Fluid Policy

Theorem 5.2 is a qualitative statement about the structure of an optimal policy. Since the optimal policy for the fluid model is defined by linear switching curves [**FAR95**], Theorem 5.2 implies that the optimal discrete policy is also defined by similar switching curves, whenever the state is large. To illustrate the way in which

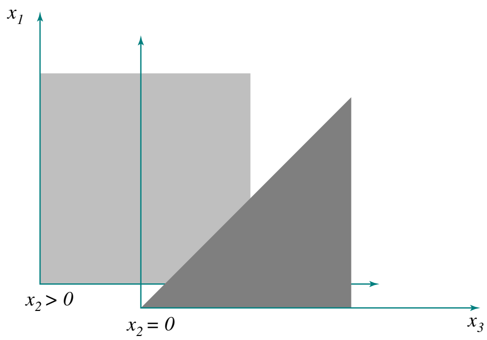

FIGURE 5. The three buffer example with $\rho_2 = 9/10$ and $\rho_1 = 9/11$. In this illustration, the grey regions indicate those states for which buffer three is given exclusive service.

this convergence takes place, we consider the simple network model introduced in Figure 4.

The fluid model for the network given in Figure 4 has the following form

$$\dot{\phi}_1(t) = \lambda - \mu_1 u_1(t), \qquad \dot{\phi}_2(t) = \mu_1 u_1(t) - \mu_2 u_2(t), \qquad \dot{\phi}_3(t) = \mu_2 u_2(t) - \mu_3 u_3(t),$$

where the controls satisfy for all t, $0 \leq u_2(t) \leq 1$ and $0 \leq u_1(t) + u_2(t) \leq 1$. One policy which minimizes the cost function

$$C(\phi) = \int_0^\infty |\phi(t)|\,dt = \sum_{i=1}^{3} \int_0^\infty \phi_i(t)\,dt,$$

is defined as follows, where γ is a positive constant defined by the parameters of the network.

i: Serve $\phi_3(t)$ exclusively ($u_3(t) = 1$) whenever $\phi_2(t) > 0$ and $\phi_3(t) > 0$;
ii: Serve $\phi_3(t)$ exclusively whenever $\phi_2(t) = 0$, and $\phi_3(t)/\phi_1(t) > \gamma$;
iii: Give $\phi_1(t)$ partial service with $u_1(t) = \mu_2/\mu_1$ whenever $\phi_2(t) = 0$, and

$$0 < \phi_3(t)/\phi_1(t) \leq \gamma$$

We consider a special case where machine two is the bottleneck with $\rho_2 = \lambda/\mu_2 = 9/10$. The load at the first machine is defined as $\rho_1 = \lambda/\mu_1 + \lambda/\mu_3 = 9/11$. Under these conditions, the constant γ is equal to one, and hence the optimal policy is of the form illustrated in Figure 5 (see [**Wei94**]).

We have computed the optimal policy for this model with the performance index

$$J_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathsf{E}_x[|\Phi(k)|].$$

To compute the policy numerically we used value iteration with $n = 7,000$. The buffer levels were truncated so that $x_i < 45$ for all i. This gives rise to a finite state space MDP with $45^3 = 91,125$ states. Boundary effects can be severe. For instance, for a loss model, if buffer one is nearly full then it is always desirable to serve buffer three since the resulting overflows at buffer one reduce the steady state cost. Since these dynamics have nothing to do with real network objectives, in the illustrations below we show the policy in the region $x_i < 25$ for all i. Since the cost

$J_n(\boldsymbol{\theta})$ was always found to be less than twelve, the threshold 25 represents a large state.

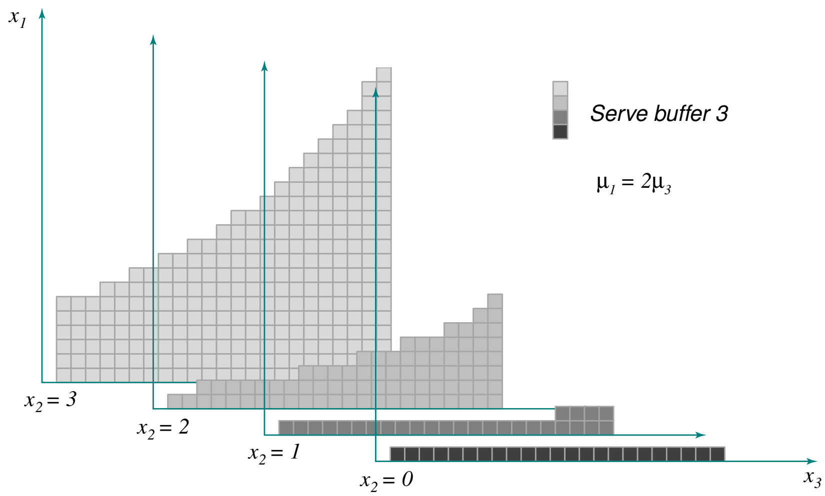

FIGURE 6. Optimal policy when the last buffer is slow: $\lambda/\mu_1 = \frac{2}{3}\rho_1$, $\lambda/\mu_3 = \frac{1}{3}\rho_1$. The optimal steady state cost: $\mathsf{E}[|x|] \approx 11.3$.

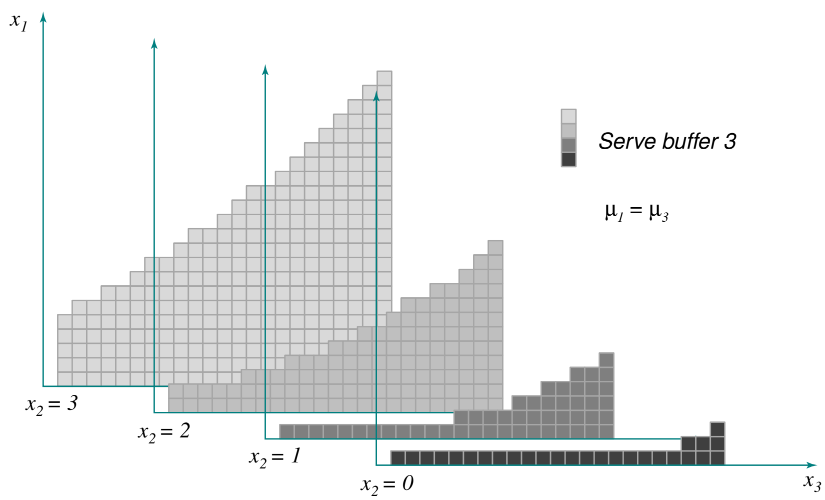

FIGURE 7. Optimal policy in the balanced case: $\lambda/\mu_1 = \lambda/\mu_3 = \frac{1}{2}\rho_1$. The optimal steady state cost: $\mathsf{E}[|x|] \approx 11.3$.

In Figures 6–8 the results of three experiments are given. The grey areas indicate those states for which buffer three is given priority. For example, in Figure 6 we see that buffer one has priority when the state is $x = (2, 1, 2)'$, while buffer three has priority when $x = (2, 1, 3)'$.

In Figure 6 where $\mu_1 = 2\mu_3$ it is seen that the optimal policy resembles FBFS (first buffer-first served) when the state falls significantly below its mean. When

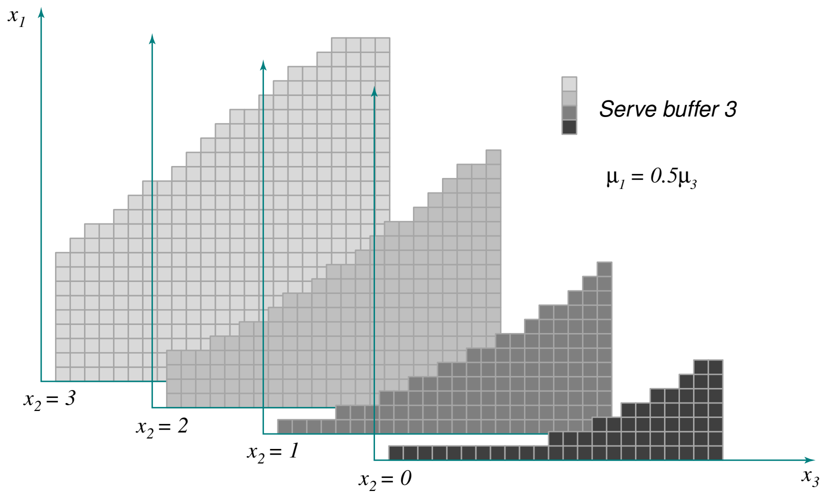

FIGURE 8. Optimal policy when the last buffer is fast: $\lambda/\mu_1 = \frac{1}{3}\rho_1$, $\lambda/\mu_3 = \frac{2}{3}\rho_1$. The optimal steady state cost: $\mathsf{E}[|x|] \approx 11.6$.

the state is greater than its mean value, the optimal policy does resemble the fluid policy. In the next two figures we consider networks where the relative speed of the third server is increasingly fast. Qualitatively, the same behavior is observed in each case.

7. A Policy for Discrete Queueing Networks

We conclude this paper with a proposed approach to the scheduling problem for the real system model of interest, which is typically a network with discrete customers. To obtain a suitable policy we search for some method of translating the continuous control to the discrete network. In the past dozen years much research has concentrated on the optimal control of networks in heavy traffic, where a Brownian system model may be used to approximate network behavior. One again faces the problem of translating the continuous control for the Brownian model to the original discrete network. In Harrison's recent treatment of the scheduling problem presented in [**Har96**], the author states that

> "...optimal control policies for Brownian system models, even if they can be determined, do not translate in any obvious way into implementable control policies for the more finely structured processing network models of original interest."

We face the same problem when given an optimal control for a fluid model. Although the synthesis of an optimal control for a fluid model appears to be easier than the synthesis problem for a Brownian model, the problem of *translation* is at least as difficult.

The examples given in [**JGDZ96**] indicate that the unmodified fluid policy may perform poorly on the discrete network model. In fact, for the example treated in Section 6 the unmodified optimal fluid policy cannot be optimal for the discrete network since with this policy the controlled network possesses a *sub-optimal* fluid limit model! To see this, suppose that the state is initially large with $\Phi_1(0) \gg \Phi_3(0)$, and that $\Phi_2(0) = 0$, so that buffer one is scheduled for service at time 0. Once the service at buffer one is complete, the second buffer will then receive a customer, and hence Φ_2 will no longer be zero. Assuming that the policy is implemented in a pre-emptive/resume manner, the next customer at buffer one cannot receive service until buffer two has completed its service at which time it is once again empty. This cycle will repeat until the first time that $\Phi_1 \leq \Phi_3$. It follows that the fluid limit of this policy satisfies

$$u_2(t) = \frac{1/\mu_2}{1/\mu_1 + 1/\mu_2} < 1, \qquad \text{whenever } \phi_2 = 0,\ \phi_1/\phi_3 > 1.$$

This differs from the optimal fluid policy, which never allows the bottleneck to idle when $\phi_1 + \phi_2 > 0$.

There are many polcies which do give rise to the optimal fluid policy. One such policy is defined as follows: serve buffer one at time k if and only if either buffer three is equal to zero, or

$$\Phi_1(k) > \Phi_3(k) - C + D\exp(\Phi_2(k)),$$

where C and D are positive constants. When the state $\Phi(k)$ is large, buffer one will receive service beginning at time k if $\Phi_1(k) > \Phi_3(k)$, and $\Phi_2(k) = O(\log(|\Phi(k)|)) = o(|\Phi(k)|)$. This ensures that the second buffer is never starved of work, so that the fluid limit model does satisfy $u_2(t) \equiv 1$ whenever $\phi_1(t) + \phi_2(t) > 0$. In the experiments that we have conducted, we have found that the optimal policy is similar in form to this policy.

In this section we describe a simple approach based upon an affine shift of the fluid policy which gives excellent results in the examples we have considered. This approach is similar to the modification of LBFS to form the "least slack policy" [**LRK94**], but is more generally applicable since the re-entrant line structure is not required. The approach described here is applicable to virtually any routing or scheduling problem for discrete networks with probabilistic routing.

In the paper [**FZ96**] the authors similarly argue that a tractable solution to a network control problem is found by first considering a deterministic fluid control problem. The setting there is different in that instead of considering large states, the authors examine the case of relatively frequent machine breakdowns.

For the general fluid network model, it may be shown that there are linear switching curves which define the optimal policy. Assume for simplicity that the policy is of the form

$$\textit{serve buffer } j \textit{ at machine } \sigma \textit{ if } j = \arg\max_{i:s(i)=\sigma}\{{\gamma_i}' x\}$$

where $\{\gamma_i\}$ are vectors in $\mathbb{R}^k$, and x is the state of the system. One approach then is to use the policy

$$\textit{serve buffer } j \textit{ at machine } \sigma \textit{ if } j = \arg\max_{i:s(i)=\sigma}\{{\gamma_i}'(x - \bar{x})\}$$

where $\bar{x}$ is a constant in $\mathbb{R}^K_+$ which defines an affine shift of the linear switching curves found in the optimal fluid policy. By shifting the origin to the value $\bar{x}$, the resulting policy draws the state towards this value whenever the state becomes large. One again faces the problem of choosing the shift $\bar{x}$. By examining the optimal policies in several examples the mean value $\bar{x} = \mathsf{E}[x]$ is suggested. The resulting policy thus has the appealing interpretation that it attempts to regulate the state around its optimal mean value. We have no proof that this approach will be effective in general, but the examples we have considered give motivation for further study.

We conclude with three experiments based upon the two machine example illustrated in Figure 4. The network parameters are identical to those used in the three examples given in Section 6 with $\rho_1 = 9/11$ and $\rho_2 = 9/10$. We take the crude estimate $\bar{x} = (3, 7, 3)'$ for the mean so that we can apply the same policy in each experiment. This mean corresponds to a steady state cost of $\mathsf{E}[|x|] = 13$, which is at least 10% greater than that obtained for any of the optimal policies. A more highly tuned value for $\bar{x}$ may give better results, but we have not tried other values. With this value of $\bar{x}$ we obtain the policy

(7.1) *serve buffer* 1 *at machine* 1 *if* $x_2 - 7 \leq 0$ and $x_1 - 3 \geq x_3 - 3$.

A drawback to this policy is that it requires numerous switchovers between buffers one and three. This is easily resolved by making switching decisions on a time scale greater than the time scale of the service times as in the BIGSTEP method of [**Har96**].

In the case where the first buffer (last buffer) is fast, we find that FBFS (LBFS) is nearly optimal, although no priority policy performs as well as the policy (7.1). The most interesting case tested is where $\mu_1 = \mu_3$. The data presented in Figure 9 shows the loss for each policy p in this balanced case. The values are in percentage,

$$\text{Loss}\,(p) = 100 \times [(J_n^p(\boldsymbol{\theta})/J_n^*(\boldsymbol{\theta})) - 1]$$

where $J_n^p(\boldsymbol{\theta})$ is the cost using the specific policy p, and $J_n^*(\boldsymbol{\theta})$ is the optimal cost. We again take $n = 7,000$. It is seen that the policy (7.1) is within 5% of the optimal cost, while LBFS and FBFS perform far worse.

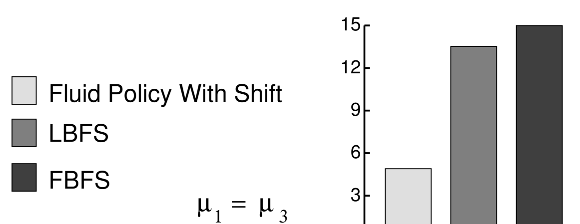

FIGURE 9. The loss in percentage of three policies compared with the optimal cost. The policy (7.1) is within 5% of the optimal cost.

Figure 10 gives a more complete comparison of the statistics of the controlled network using the policy (7.1), the optimal policy, and the two priority policies. It is seen here that the state is regulated in the mean near the target value $\bar{x} = (3, 7, 3)'$. Using this policy the standard deviation at each buffer is also *less* than that of the optimal policy.

For completeness we present in Figures 11 and 12 comparative data in the unbalanced case where either the first or last server is fast. We see that similar results continue to hold: the policy (7.1) is always near-optimal; the state is regulated in the mean to the value $\bar{x}$; and the standard deviation of buffer levels is also well behaved.

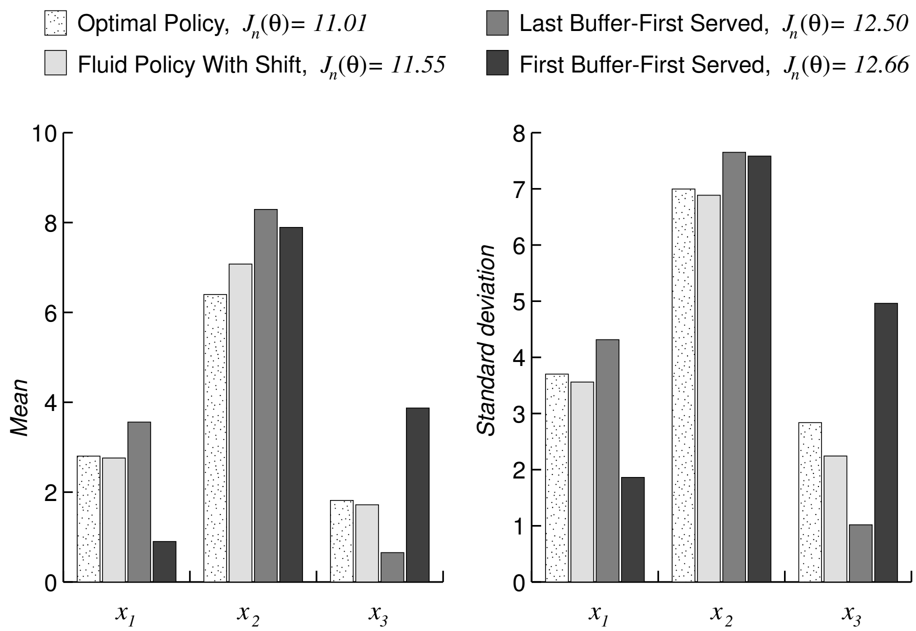

FIGURE 10. A comparison of four policies. The first machine is balanced with $\lambda/\mu_1 = \lambda/\mu_3$.

8. Conclusions

This paper reveals some appealing connections between optimization of discrete networks and their fluid models. To make this a practical approach to scheduling there are many unresolved issues such as,

1. *Can we bound the complexity of optimal fluid policies?*
2. *Can the performance of a suboptimal policy based upon the fluid optimal policy be bounded, say, through a linear program?*
3. *Is the suboptimal policy stable?*
4. *Is there a connection between this approach and that based upon a Brownian approximation?*

This paper invites more questions than it answers, but hopefully the reader will agree that they are interesting questions.

Acknowledgement

The author wishes to thank David K. Eng for his creative input during the course of this research and for his assistance with computer simulation.

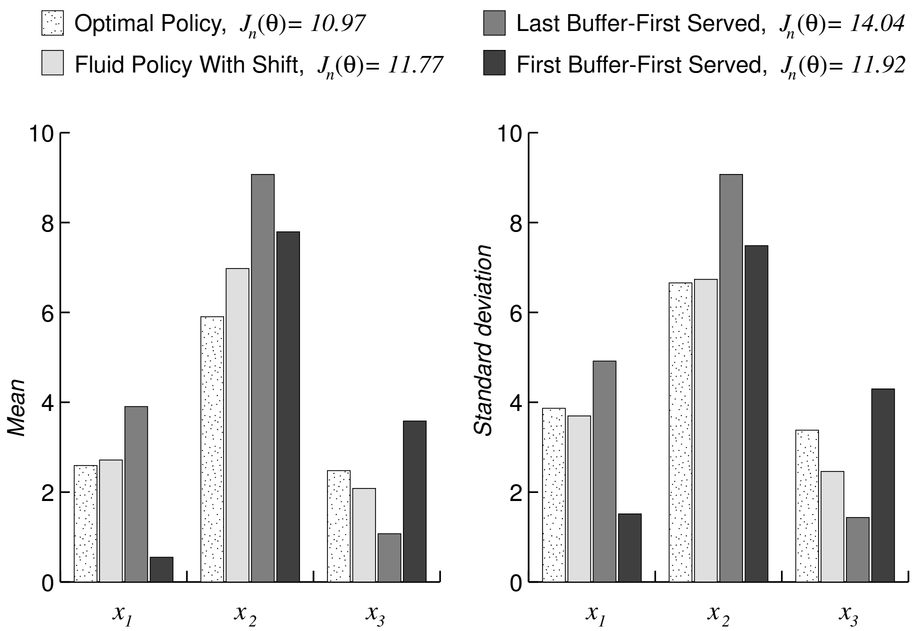

FIGURE 11. A comparison of four policies when last buffer is slow: $\lambda/\mu_1 = \frac{2}{3}\rho_1$, $\lambda/\mu_3 = \frac{1}{3}\rho_1$.

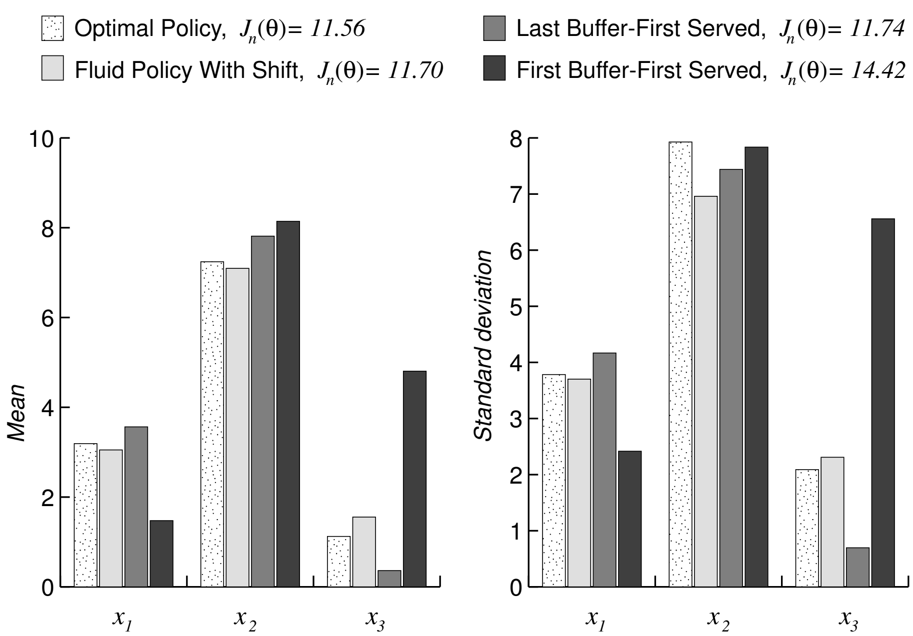

FIGURE 12. A comparison of four policies when the last buffer is fast: $\lambda/\mu_1 = \frac{1}{3}\rho_1$, $\lambda/\mu_3 = \frac{2}{3}\rho_1$.

References

[ABF+93] A. Arapostathis, V. S. Borkar, E. Fernandez-Gaucherand, M. K. Ghosh, and S. I. Marcus, *Discrete-time controlled Markov processes with average cost criterion: a survey*, SIAM J. Control Optim. **31** (1993), 282–344.

[Bor91] V. S. Borkar, *Topics in controlled Markov chains*, Pitman Research Notes in Mathematics Series # 240, Longman Scientific & Technical, UK, 1991.

[Cao96] Xi-Ren Cao, *Potentials, average cost Markov decision problems, and single sample path based optimization*, submitted for publication, 1996.

[CF95] R. Cavazos-Cadena and E. Fernandez-Gaucherand, *Value iteration in a class of average controlled Markov chains with unbounded costs: Necessary and sufficient conditions for pointwise convergence*, Proceedings of the 34th IEEE Conference on Decision and Control (New Orleans, LA), 1995, To appear in Journal of Applied Probability, pp. 2283–2288.

[Che96] H. Chen, *Optimal routing control of a fluid model*, Technical Note, 1996.

[CM91] H. Chen and A. Mandelbaum, *Discrete flow networks: Bottlenecks analysis and fluid approximations*, Mathematics of Operations Research **16** (1991), 408–446.

[CZ96] H. Chen and H. Zhang, *Stability of multiclass queueing networks under priority service disciplines*, Technical Note, 1996.

[Dai95] J. G. Dai, *On the positive Harris recurrence for multiclass queueing networks: A unified approach via fluid limit models*, Ann. Appl. Probab. **5** (1995), 49–77.

[DBT95] D. Gamarnik D. Bertsimas and J. N. Tsitsiklis, *Stability conditions for multiclass fluid queueing networks*, Tech. report, Massachusets Institute of Technology, 1995.

[DEM96] J. Humphrey D. Eng and S.P. Meyn, *Fluid network models: Linear programs for control and performance bounds*, Proceedings of the 13th IFAC World Congress (San Francisco, California) (J. Cruz J. Gertler and M. Peshkin, eds.), vol. B, 1996, pp. 19–24.

[DM95] J. G. Dai and S.P. Meyn, *Stability and convergence of moments for multiclass queueing networks via fluid limit models*, IEEE Trans. Automat. Control **40** (1995), 1889–1904.

[DV96] J. Dai and J. H. Vande Vate, *Virtual stations and the capacity of two–station queueing networks*, Preprint, March 1996, School of Industrial and Systems Engineering, Georgia Institute of Technology.

[DW96] J. Dai and G. Weiss, *Stability and instability of fluid models for certain re-entrant lines*, Mathematics of Operations Research **21** (1996), no. 1, 115–134.

[FAR95] D. Bertsimas F. Avram and M. Ricard, *Fluid models of sequencing problems in open queueing networks: an optimal control approach*, Technical Report, Massachussetts Institute of Technology, 1995.

[FZ96] Ngo-Tai Fong and Xun Yu Zhou, *Hierarchical feedback controls in two-machine flow shops under uncertainty*, Proceedings of the 36th Conference on Decision and Control (Kobe, Japan), 1996, pp. 1743–1748.

[GM96] P. W. Glynn and S. P. Meyn, *A Lyapunov bound for solutions of Poisson's equation*, Ann. Probab. **24** (1996).

[Haj84] B. Hajek, *Optimal control of two interacting service stations*, IEEE Trans. Automat. Control **AC-29** (1984), 491–499.

[Har96] J.M. Harrison, The BIGSTEP approach to flow management in stochastic processing networks, Stochastic Networks Theory and Applications, 57–89, Stochastic Networks Theory and Applications, Clarendon Press, Oxford, UK, 1996, pp. 57–89, F.P. Kelly, S. Zachary, and I. Ziedins (ed.).

[HLL95] O. Hernández-Lerma and J. B. Lassere, *Discrete time Markov control processes I*, IPN, Departamento de Matematicas, Mexico, and LAAS-CNRS, France, 1995, to appear.

[JGDZ96] D. H. Yeh J. G. Dai and C. Zhou, *The QNET method for re-entrant queueing networks with priority disciplines*, *Operations Research*, to appear, 1996.

[KK94] S. Kumar and P. R. Kumar, *Fluctuation smoothing policies are stable for stochastic re-entrant lines*, *Proceedings of the 33rd IEEE Conference on Decision and Control*, December 1994.

[KM96] P.R. Kumar and S.P. Meyn, *Duality and linear programs for stability and performance analysis queueing networks and scheduling policies*, IEEE Transactions on Automatic Control **41** (1996), no. 1, 4–17.

[Lip75] S. Lippman, *Applying a new device in the optimization of exponential queueing systems*, Operations Research **23** (1975), 687–710.

[LRK94] Steve C. H. Lu, Deepa Ramaswamy, and P. R. Kumar, *Efficient scheduling policies to reduce mean and variance of cycle–time in semiconductor manufacturing plants*, IEEE Transactions on Semiconductor Manufacturing **7** (1994), no. 3, 374–385.

[Mey95a] S. P. Meyn, *Transience of multiclass queueing networks via fluid limit models*, Ann. Appl. Probab. **5** (1995), 946–957.

[Mey95b] S.P. Meyn, *The policy improvement algorithm for Markov decision processes with general state space*, To appear, IEEE Trans. Auto. Control, 1995.

[MT93] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, London, 1993.

[Per93] J. Perkins, *Control of push and pull manufacturing systems*, Ph.D. thesis, University of Illinois, Urbana, IL, September 1993, Technical report no. UILU-ENG-93-2237 (DC-155).

[Put94] M. L. Puterman, *Markov Decision Processes*, Wiley, New York, 1994.

[RS93] R. K. Ritt and L. I. Sennott, *Optimal stationary policies in general state space Markov decision chains with finite action set*, Mathematics of Operations Research **17** (1993), no. 4, 901–909.

[Sen89] L.I. Sennott, *Average cost optimal stationary policies in infinite state Markov decision processes with unbounded cost*, Operations Res. **37** (1989), 626–633.

[Sen96] L.I. Sennott, *The convergence of value iteration in average cost Markov decision chains*, Operations Research Letters **19** (1996), 11–16.

[Wei94] G. Weiss, *On the optimal draining of re-entrant fluid lines*, Tech. report, Georgia Georgia Institute of Technology and Technion, 1994.

[WS87] R. Weber and S. Stidham, *Optimal control of service rates in networks of queues*, Adv. Appl. Probab. **19** (1987), 202–218.

[XC93] 6.S.H. Xu and H. Chen, *A note on the optimal control of two interacting service stations*, IEEE Transactions on Automatic Control **38** (1993), 187–189.

Coordinated Science Laboratory and the University of Illinois, 1308 W. Main Street, Urbana, IL 61801, URL `http://black.csl.uiuc.edu:80/~meyn`

E-mail address: `s-meyn@uiuc.edu`