Stability and Asymptotic Optimality of Generalized MaxWeight Policies*

Sean Meyn†

May 11, 2007

Abstract

It is shown that stability of the celebrated MaxWeight or back pressure policies is a consequence of the following interpretation: either policy is myopic with respect to a surrogate value function of a very special form, in which the "marginal disutility" at a buffer vanishes for vanishingly small buffer population. This observation motivates the *h-MaxWeight policy*, defined for a wide class of functions h. These policies that share many of the attractive properties of the MaxWeight policy:

(i) The policy does not require arrival rate data.

(ii) The h-myopic policy is stabilizing when h is a perturbation of a monotone linear function, or a monotone Lyapunov function for the fluid model.

(iii) A perturbation of the relative value function for a workload relaxation gives rise to a myopic policy that is approximately average-cost optimal in heavy traffic, with logarithmic regret.

The first results are obtained for a completely general stochastic network model. Asymptotic optimality is established for the general scheduling model with a single bottleneck.

Keywords: Queueing Networks, Routing, Scheduling, Optimal Control.

AMS subject classifications: Primary: 90B35, 68M20, 90B15;
Secondary: 93E20, 60J20.

---

*Financial support from the National Science Foundation ECS-0523620 and ITMANET DARPA RK 2006-07284 is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

†Department of Electrical and Computer Engineering and the Coordinated Sciences Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, U.S.A. (meyn@uiuc.edu).

1 Introduction

While it is popular to cite the curse of dimensionality when discussing optimization of stochastic networks, there are many classes of effective policies that are easily implemented, require limited information, and have other attractive properties.

Tassiulas and Ephremides showed in [48] that a myopic policy is universally stabilizing for a class of stochastic networks, provided the surrogate value function used in the definition of the policy is quadratic,

$$h(x) = \tfrac{1}{2} x^{\mathrm{T}} D x, \qquad x \in \mathbb{R}^\ell, \tag{1}$$

with $D > 0$ a diagonal matrix. The resulting policy is known as MaxWeight or back-pressure, depending on the context. This work has been extended in multiple directions over the past fifteen years [17, 47, 43], and in particular these policies are known to be approximately optimal in heavy traffic under certain conditions on the network [50, 45, 31].

These results are fragile. For example, for a general stochastic scheduling model, diagonal quadratics are one of a very few function classes for which the myopic policy is known to be stabilizing.[1] Moreover, even the related Proportional Fair scheduler is known to be destabilizing for certain stochastic models [1]. In contrast, stability of the fluid model under a myopic policy is virtually universal [10, 7, 33].

To explain this gap, consider the following two models. The controlled random walk (CRW) model evolves on $\mathbb{Z}_+^\ell$ according to the recursion,

$$Q(t+1) = Q(t) + B(t+1)U(t) + A(t+1), \qquad t \geq 0,\ Q(0) = x\,. \tag{2}$$

The allocation sequence $\boldsymbol{U}$ evolves on $\mathbb{Z}_+^{\ell_u}$ for some integer ℓ_u; the arrival sequence $\boldsymbol{A}$ is ℓ-dimensional, and $\boldsymbol{B}$ is an $\ell \times \ell_u$ matrix sequence, each with integer-valued entries. The fluid model $\boldsymbol{q} = \{q(t) : t \geq 0\}$ satisfies the linear equations,

$$q(t) = x + Bz(t) + \alpha t, \qquad t \geq 0,\ x \in \mathbb{R}_+^\ell, \tag{3}$$

where B and α are the mean values of $B(t)$ and $A(t)$, respectively, and $\boldsymbol{z}$ is the cumulative allocation process evolving on $\mathbb{R}_+^{\ell_u}$. The fluid model (3) is also expressed as the ODE model,

$$\tfrac{d^+}{dt} q(t) = B\zeta(t) + \alpha\,, \qquad t \geq 0, \tag{4}$$

where $\zeta(t) \in \mathbb{R}_+^{\ell_u}$ denotes the allocation rate vector at time t.

The allocation sequence $\boldsymbol{U}$ and the rate process $\boldsymbol{\zeta}$ satisfy common linear constraints, $U(t) \in \mathsf{U}$, $\zeta(t) \in \mathsf{U}$, where $\mathsf{U} \subset \mathbb{R}_+^{\ell_u}$ is a polyhedral set of the specific form,

$$\mathsf{U} := \{u \in \mathbb{R}^{\ell_u} : u \geq \mathsf{0}, \quad Cu \leq \mathsf{1}\}. \tag{5}$$

The matrix C is called the constituency matrix: Its entries are assumed to be binary, and further assumptions are imposed in Theorem 1.1. Each row of C corresponds to a station

[1]While not precisely a myopic policy, the stability analysis contained in [13] implies that the h-myopic policy is stabilizing with h of the form $h(x) = \sum h_i(x_i)$, provided that $\{h_i\}$ satisfy certain Lipschitz bounds. The exponential rule of [44] is also similar to a myopic policy.

(or 'resource') in the network: For a given $j \in \{1,\dots,\ell\}$ we let $s(j) \in \{1,\dots,\ell_m\}$ denote the corresponding station. That is, the unique value of s satisfying $C_{s,j} = 1$.

Suppose that $h\colon \mathbb{R}_+^\ell \to \mathbb{R}_+$ is any C^1 function that vanishes only at the origin. The *h-myopic policy* is defined for the fluid and stochastic models by the respective feedback laws,

$$\begin{aligned}\phi^{\mathrm{F}}(x) &= \mathop{\arg\min}_{\zeta\in\mathsf{U}(x)} \langle \nabla h\,(x), B\zeta+\alpha\rangle, &(6)\\ \phi^{\mathrm{D}}(x) &= \mathop{\arg\min}_{u\in\mathsf{U}_\diamond(x)} \mathsf{E}[h(Q(t+1)) - h(Q(t)) \mid Q(t)=x, U(t)=u], &(7)\end{aligned}$$

where $\mathsf{U}(x)$ denotes the set of admissible values of ζ when $q = x$,

$$\mathsf{U}(x) = \{u \in \mathsf{U} : v_i := (Bu+\alpha)_i \ge 0 \text{ when } x_i = 0\}\,, \tag{8}$$

and $\mathsf{U}_\diamond(x) := \mathsf{U}(x) \cap \{0,1\}^{\ell_u}$. It is shown in Proposition 2.2 that the MaxWeight policy coincides with the myopic policy for the fluid model when h is equal to the quadratic (1). This motivates an alternative policy for the stochastic model, the *h-MaxWeight policy*,

$$\phi^{\mathrm{MW}}(x) = \mathop{\arg\min}_{u\in\mathsf{U}_\diamond(x)} \langle \nabla h\,(x), Bu+\alpha\rangle, \qquad x \in \mathbb{Z}_+^\ell. \tag{9}$$

Note that $\phi^{\mathrm{MW}} = \phi^{\mathrm{D}}$ when h is a linear function of x.

The policy (6) is stabilizing for the fluid model under mild assumptions on the function h (see [10] and [7, Thm. 12.5] for linear functions, and [33, Proposition 11] for a smooth norm on $\mathbb{R}^\ell$.) The proof is based on establishing that the 'drift' defined by,

$$\tfrac{d^+}{dt} h(q(t)) = \langle \nabla h\,(q(t)), \tfrac{d^+}{dt} q(t)\rangle, \qquad t \ge 0, \tag{10}$$

is strictly negative when $q(t) \neq \mathsf{0}$.

The myopic policy (7) for the stochastic model may or may not be stabilizing, depending upon the particular network and the structure of the function h. One difficulty is that the corresponding drift for the stochastic model,

$$\mathsf{E}[h(Q(t+1)) - h(Q(t)) \mid Q(t)=x, U(t)=u] \tag{11}$$

can be positive for certain values of x on the boundary of the state space. This important distinction between the two models is illustrated in the following two examples.

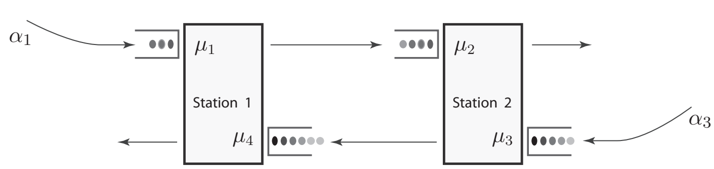

Figure 1: The Kumar-Seidman-Rybko-Stolyar model

Instability in the model of Rybko and Stolyar Consider the model of Kumar and Seidman, and Rybko and Stolyar shown in Figure 1 [25, 42]. With c the ℓ_1 norm, $c(x) = |x| = \sum x_i$, the myopic policy for the fluid model gives priority to the exit buffers if no machine is starved of work. Suppose that the parameters satisfy,

$$\mu_1 > \mu_2 \text{ and } \mu_3 > \mu_4. \tag{12}$$

If for example $x_1 > 0$ and $x_4 > 0$ yet $x_2 = x_3 = 0$ we then have

$$\phi_1^{\mathrm{F}}(x) = \mu_2\mu_1^{-1}, \qquad \phi_4^{\mathrm{F}}(x) = 1 - \phi_1^{\mathrm{F}}(x).$$

The myopic policy for the stochastic model is very different: The optimization (7) defines $\phi_4^{\mathrm{F}}(x) = 1$ if $x_4 \geq 1$, and $\phi_2^{\mathrm{F}}(x) = 1$ if $x_2 \geq 1$. This is precisely the policy found to be destabilizing in [42].

Work-stoppage under a myopic policy The myopic policy for the tandem queues with h linear may be entirely irrational. Consider the pair of queues in tandem illustrated in Figure 2. Suppose that a linear cost function is given $c(x) = c_1x_1 + c_2x_2$, with $c_2 > c_1$. The c-myopic policy for the fluid model is non-idling at Station 2, while at Station 1,

$$\phi_1^{\mathrm{F}}(x) = \begin{cases} 0 & \text{if } x_2 > 0 \\ \min(1, \mu_2\mu_1^{-1}) & \text{if } x_2 = 0,\ x_1 > 0. \end{cases} \tag{13}$$

The c-myopic policy is path-wise optimal when $\mu_1 \geq \mu_2$.

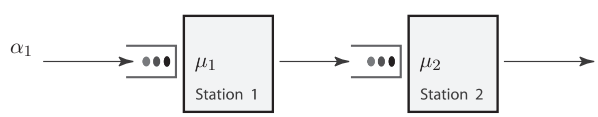

Figure 2: Tandem queues.

For a CRW model defined consistently with the fluid model we have for $x \in \mathbb{Z}_+^2$,

$$\begin{aligned} \phi^{\mathrm{MW}}(x) = \phi^{\mathrm{D}}(x) &= \arg\min_{u \in \mathsf{U}_\diamond(x)} \mathsf{E}[c(Q(t+1)) \mid Q(t) = x,\ U(t) = u] \\ &= \arg\min_{u \in \mathsf{U}_\diamond(x)} \big(c_1(\alpha_1 - \mu_1u_1) + c_2(\mu_1u_1 - \mu_2u_2)\big). \end{aligned}$$

At Station 2 this policy is non-idling, while at Station 1,

$$\phi_1^{\mathrm{MW}}(x) = \arg\min_{u \in \mathsf{U}_\diamond(x)} \big((c_2 - c_1)\mu_1u_1\big).$$

That is, Station 1 is *always idle* under our assumption that $c_2 > c_1$!

Instability is a consequence of *additional constraints in the stochastic model*. The choices are limited in the CRW model, so from certain states on the boundary it is not possible to find an allocation u such that the drift in (11) is negative. Why then is this possible when h is a diagonal quadratic?

We show in this paper that the key property that is required is the derivative condition,

$$\frac{\partial}{\partial x_j} h(x) = 0 \qquad \text{when } x_j = 0. \tag{14}$$

For a quadratic we have $\nabla h(x) = Dx$, and hence (14) does hold when D is diagonal.

With h interpreted as an approximate value function, the derivative $\frac{\partial}{\partial x_j} h(x)$ represents the "marginal disutility" of an additional increment of inventory at buffer j. If this marginal disutility is zero, then it is reasonable to shift inventory to this buffer when possible. Thus starvation of resources is avoided, which is the cause of instability in these two examples.

In this paper we make these informal observations precise. Moreover, to obtain a wide class of policies we describe a perturbation technique used to modify a given function so that (14) holds. Suppose that c is a norm on $\mathbb{R}^\ell$, such as $c(x) = \sum |x_i|$, and that $h_0 \colon \mathbb{R}^\ell \to \mathbb{R}_+$ is any C^1 function that satisfies the dynamic programing *inequality* for the fluid model,

$$\min_{u \in \mathsf{U}(x)} \langle \nabla h_0(x), Bu + \alpha \rangle \le -c(x), \qquad x \in \mathbb{R}^\ell_+. \tag{15}$$

With ϕ^{F} defined in (6) using h_0, and $v := B\phi^{\mathrm{F}}(x) + \alpha$, the bound (15) is equivalent to the functional inequality $\langle \nabla h_0, v \rangle \le -c$.

A perturbation of h_0 is obtained through a change of variables: For fixed $\theta \ge 1$ we denote

$$\tilde{x}_i := x_i + \theta(e^{-x_i/\theta} - 1), \quad \text{for any } i \text{ and } x, \tag{16}$$

and let $\tilde{x}$ denote the corresponding vector $\tilde{x} := (\tilde{x}_1, \dots, \tilde{x}_\ell)^{\mathrm{T}} \in \mathbb{R}^\ell_+$. The function h is then defined by,

$$h(x) = h_0(\tilde{x}), \qquad x \in \mathbb{R}^\ell_+. \tag{17}$$

An application of the chain rule shows that (14) holds. The first main result of this paper is based on this observation:

Theorem 1.1. *Consider the model (2) satisfying the following conditions:*

(i) *The entries of the constituency matrix are binary-valued, its rows are orthogonal, and the sum of any column is equal to unity.*

(ii) *The i.i.d. process $(\boldsymbol{A}, \boldsymbol{B})$ has integer entries, and a finite second moment.*

(iii) *$B_{ij}(t) \ge -1$ for each i, j and t, and for each $j \in \{1, \dots, \ell_u\}$ there exists a unique value $i_j \in \{1, \dots, \ell\}$ satisfying,*

$$B_{ij}(t) \ge 0 \qquad a.s. \;\; i \ne i_j. \tag{18}$$

(iv) *The function $h_0 \colon \mathbb{R}^\ell \to \mathbb{R}_+$ satisfies the following:*

(a) *Smoothness: The gradient ∇h_0 is Lipschitz continuous,*

(b) *Monotonicity: $\nabla h_0(x) \in \mathbb{R}^\ell_+$ for $x \in \mathbb{R}^\ell_+$,*

(c) *The dynamic programing inequality (15) holds, with c a norm on $\mathbb{R}^\ell$.*

Then, there exists $\theta_0 < \infty$ and $b_h < \infty$ such that for any $\theta \ge \theta_0$, the following bound holds under the h-MaxWeight policy:

$$\mathcal{D}h\,(x) = \mathsf{E}[h(Q(t+1)) - h(Q(t)) \mid Q(t) = x] \le -\tfrac{1}{2}c(x) + b_h. \tag{19}$$

Consequently,

$$n^{-1}\mathsf{E}\Big[\sum_{t=0}^{n-1} c(Q(t))\Big] \le 2n^{-1}h(x) + 2b_h, \qquad n \ge 1, x \in \mathbb{Z}_+^\ell. \tag{20}$$

Proof. The bound (19) is established in Section 2.2.2, and (20) then follows from the Comparison Theorem 2.1. □

The inequality (19) is a *Lyapunov drift condition* of the form developed in [37, 15], and also similar to the bounds used in [11, 5, 26, 24, 39] to obtain performance bounds for networks. Under natural assumptions on the model the bound (19) implies that the controlled network is geometrically ergodic, so that the mean $\mathsf{E}[c(Q(t))]$ converges to its steady-state value geometrically fast from each initial condition [37, 38, 24, 33].

To obtain finer performance bounds we impose further structure on the model. For simplicity we restrict to a scheduling model with deterministic routing: for each $i \in \{1,\dots,\ell\}$, after processing at buffer i a customer either enters some buffer $i_+ \in \{1,\dots,\ell\}$, or exits the system. The *routing matrix* R is the $\ell \times \ell$ matrix defined for $i, j \in \{1,\dots,\ell\}$ as $R_{ij} = \mathbf{1}_{j=i_+}$. The routing matrix satisfies $R^\ell = 0_{\ell\times\ell}$ - this ensures that each customer receives at most ℓ services during its lifetime in the network.

The CRW scheduling model is described by the recursion,

$$Q(t+1) = Q(t) + \sum_{i=1}^{\ell} M_i(t+1)[-\mathbf{1}^i + \mathbf{1}^{i_+}]U_i(t) + A(t+1), \qquad Q(0) = x\,, \tag{21}$$

where $\boldsymbol{M}_i$ is a Bernoulli sequence for each i. This is of the form (2) with,

$$B(t) = -[I - R^{\mathrm{T}}]M(t), \qquad t \ge 1.$$

The condition (18) follows from the assumptions on $\boldsymbol{M}$ and R, where $i_j = j$ for each $j \in \{1,\dots,\ell\}$.

The vector load $\rho \in \mathbb{R}_+^{\ell_m}$ is given by,

$$\rho = CM^{-1}[I - R^{\mathrm{T}}]^{-1}\alpha, \tag{22}$$

where M is a diagonal matrix, equal to the common mean of $M(t)$. The network load is the maximum, $\rho_\bullet = \max_{1\le i\le \ell} \rho_i$.

We restrict to models of a generalized 'Kelly type' in which service statistics are determined by the station, not the buffer. Station s is said to be *homogeneous* if the random variables $\{M_j(t) : s(j) = s\}$ are all identical, and the network is called homogeneous if each station satisfies this condition. Homogeneity is used to construct a one-dimensional workload relaxation of the form introduced in the thesis of N. Laws [30] (see also [29, 40].) A one-dimensional workload process $\boldsymbol{W}$ corresponding to the most heavily loaded station is compared to its relaxation $\widehat{\boldsymbol{W}}$, which is a version of a single queue. For each $t \geq 0$ the lower bound holds with probability one,

$$W(t) \geq \widehat{W}^*(t), \qquad t \geq 0, \tag{23}$$

under *any* policy for $\boldsymbol{Q}$, where the relaxation $\widehat{\boldsymbol{W}}^*$ is controlled using the non-idling policy. Relaxations of this form and multi-dimensional extensions are the subject of [9, 35].

Based on the solution to the average-cost optimality equation for the single queue, we construct a function h_0 such that the resulting h-MaxWeight policy with h defined in (17) is approximately average-cost optimal for $\boldsymbol{Q}$. Letting $\widehat{\eta}^*$ denote the optimal cost for the relaxation, and η the average-cost obtained under the h-MaxWeight policy, we have for some fixed constant $k_0 < \infty$, independent of load,

$$\widehat{\eta}^* \leq \eta \leq \widehat{\eta}^* + k_0 \log(\widehat{\eta}^*). \tag{24}$$

In this case we say that the policy is *heavy-traffic asymptotically optimal* (HTAO) *with logarithmic regret*.

This form of approximate optimality is the focus of several recent papers. A closely related form of asymptotic optimality was formulated in [34, Section 4.2], and verified for a class of multiclass networks with multiple bottlenecks and renewal inputs. Also closely related are results in the heavy traffic theory of stochastic networks, e.g. [21, 18, 27, 3, 2], as well as the aforementioned contributions [50, 45, 31, 43]. All of these asymptotic results require the *complete resource pooling* condition, meaning that there is a single bottleneck in heavy traffic. The diffusion approximations obtained in these papers *suggest* a bound of the form,

$$\eta \leq \widehat{\eta}^* + o(\sqrt{\widehat{\eta}^*}).$$

This heuristic has been established rigorously only in special cases, such as the single queue in the pioneering work of Kingman [22, 23], and the recent result [16] for generalized Jackson networks.

Heavy-traffic asymptotically optimality of the form (24) was obtained for the first time in [35] in several examples, based on a general Lyapunov condition similar to (19). The main idea is to take the optimal value function for the fluid model, and introduce a penalty function to account for possible starvation when the state reaches the boundary of $\mathbb{R}^\ell_+$. In each example considered, a single policy is proposed that is independent of network load, based on a switching curve with logarithmic growth. For example, for the tandem queues, the policy has the form

$$\textit{Serve buffer } 1 \textit{ if } x_2 \leq \beta \log(1 + x_1/\beta), \tag{25}$$

where $\beta > 0$ is a sufficiently large constant.

The approach used here is similar: The function h_0 is chosen as an approximation to a *fluid value function*. Rather than a penalty function, the change of variables (17) is used to construct a stabilizing h-MaxWeight policy and, under stronger conditions, HTAO.

Perturbed test function techniques are widely used for establishing weak limits of a family of stochastic processes, and this technique has been applied in the heavy-traffic theory of stochastic networks [28, 14]. The connection with the techniques in this paper appears to be only superficial.

Two significant contributions in the present paper are worth highlighting,

(i) In each example considered in [35] the policy was explicitly constructed based on switching curves of the form (25). This requires considerable intuition for more complex models. In the two main results Theorem 1.1 and Theorem 3.1 the policy is derived from the given value function via the minimization (9).

(ii) This is the first paper to give a completely general approach to HTAO for average cost, and in particular bounds of the form (24) for a general class of models.

The remainder of the paper is organized as follows. Section 2 concerns stability of h-MaxWeight policies. Asymptotic optimality is treated in Section 3: Theorem 3.1 establishes a bound of the form (24) for a family of models with increasing load. Section 4 contains conclusions and possible extensions.

2 MaxWeight policies

In this section we consider the general CRW model (2) under the assumptions of Theorem 1.1: The sequence $\{A(t), B(t)\}$ is i.i.d. with a finite second moment, and (18) holds for $B(t)$. This is a very general model:

(i) Controlled routing from buffer i is modeled by allowing $i_j = i$ for more than one $j \in \{1, \dots, \ell_u\}$. Then, $U_j(t) = 1$ indicates that a customer at buffer i is routed to buffer i_j^+.

(ii) It is straightforward to model a flexible server, as found in the processor sharing models of [19, 20].

(iii) Assembly-disassembly systems can be modeled. For example, following service completion at buffer i_j, a customer can spawn several new jobs. This is modeled by defining positive values of $B_{ij}(t)$ for several values of $i \neq i_j$.

(iv) On interpreting an arrival as an increment of demand, the CRW model (2) can be used to model inventory systems. In this setting, an entry of $U(t)$ can model an order for new raw material [8].

All of the policies considered in this paper are stationary, and deterministic: For the CRW model it is assumed that there is a function $\phi\colon \mathbb{Z}_+^\ell \to \mathsf{U}_\diamond$ such that

$$U(t) = \phi(Q(t)), \qquad t \geq 0. \tag{26}$$

Hence $\boldsymbol{Q}$ is a time-homogeneous Markov chain on $\mathbb{Z}_+^\ell$, with transition matrix denoted P. That is, for each $x, y \in \mathbb{Z}_+^\ell$ and $t \geq 0$,

$$P(x,y) = \mathsf{P}\{Q(t+1) = y \mid Q(t) = x\} = \mathsf{P}\{x + B(1)\phi(x) + A(1) = y\}.$$

For any function $g\colon \mathbb{R}^\ell \to \mathbb{R}$, the *generator* $\mathcal{D}$ is defined as the difference operator,

$$\mathcal{D}g\,(x) := \mathsf{E}[g(Q(t+1)) - g(Q(t)) \mid Q(t) = x] = \sum_{y \in \mathbb{Z}_+^\ell} P(x,y)[g(y) - g(x)].$$

The analysis of the h-MaxWeight policy is based on bounds on the drift $\mathcal{D}h$. The first step is an application of the Mean Value Theorem, which implies the following representation for any $Q(t) \in \mathbb{Z}_+^\ell$, and any $t \geq 0$,

$$\begin{aligned} h(Q(t+1)) - h(Q(t)) &= \langle \nabla h\,(\bar{Q}), \Delta(t+1)\rangle \\ &= \langle \nabla h\,(Q(t)), \Delta(t+1)\rangle \\ &\quad + \langle \nabla h\,(\bar{Q}) - \nabla h\,(Q(t)), \Delta(t+1)\rangle \end{aligned} \tag{27}$$

where $\Delta(t+1) := Q(t+1) - Q(t)$, and $\bar{Q} \in \mathbb{R}_+^\ell$ lies on the line connecting $Q(t+1)$ and $Q(t)$. Consequently,

$$\mathcal{D}h\,(x) = \langle \nabla h\,(x), v\rangle + b_h, \tag{28}$$

where

$$v(x) = \mathsf{E}[\Delta(t+1) \mid Q(t) = x], \qquad b_h(x) = \mathsf{E}[\langle \nabla h\,(\bar{Q}) - \nabla h\,(Q(t)), \Delta(t+1)\rangle \mid Q(t) = x]. \tag{29}$$

To deduce stability based on (28) we obtain a bound on $\langle \nabla h\,(x), v\rangle$ under the given policy. We then show that the second term $b_h(x)$ is relatively small in magnitude.

We begin with a review of some Lyapunov theory for Markov and fluid models.

2.1 Stochastic stability

When does a stabilizing policy exist for a network model? How do we test for stability? To answer these questions we consider first the fluid model.

Denote the velocity set for the fluid model by

$$\mathsf{V} := \{v = B\zeta + \alpha : \zeta \in \mathsf{U}\}. \tag{30}$$

In the general setting of this section, the "load condition" $\rho_\bullet < 1$ translates to the following,

$$\textit{The origin is an interior point of } \mathsf{V}. \tag{31}$$

If (31) holds then there exists $\varepsilon > 0$ such that the vector $v^x = -\varepsilon x/|x|$ lies in V for each $x \in \mathbb{R}_+^\ell$. Setting $\zeta = v^x$ from the initial condition $q(0) = x$ we have $q(t) = q(0) - (\varepsilon/|x|)t$ for $0 \leq t \leq |x|/\varepsilon$. For a given policy for the fluid model, such as (7), we consider the value function,

$$J(x) = \int_0^\infty c(q(t;x))\,dt. \tag{32}$$

We let J^* denote the minimum of (32) over all policies. Setting $h = J$ in (6) we obtain the drift inequality,

$$\min_{u \in \mathsf{U}(x)} \langle \nabla J\,(x), Bu + \alpha \rangle \le -c(x), \qquad x \in \mathbb{R}_+^\ell,$$

and this inequality is equality when $h = J^*$. We thereby obtain solutions to the dynamic programing inequality (15).

We now turn to the CRW model. The general form of the Lyapunov condition considered here is Condition (V3), or the special case known as Foster's criterion [37]. All involve bounds on the generator applied to a function $V\colon \mathbb{Z}_+^\ell \to \mathbb{R}_+$.

(i) The controlled network satisfies *Foster's criterion* if there is a constant $b < \infty$ and a finite set $S \subset \mathbb{Z}_+^\ell$ such that

$$\mathcal{D}V\,(x) \le -1 + b\mathbf{1}_S(x), \qquad x \in \mathbb{Z}_+^\ell. \tag{\textbf{V2}}$$

(ii) The network satisfies Condition (V3) if for a function $f\colon \mathbb{Z}_+^\ell \to [1,\infty)$,

$$\mathcal{D}V\,(x) \le -f(x) + b\mathbf{1}_S(x), \qquad x \in \mathbb{Z}_+^\ell\,, \tag{\textbf{V3}}$$

where again $b < \infty$ and $S \subset \mathbb{Z}_+^\ell$ is a finite set.

We have the following simple but useful result relating the policies ϕ^{MW} and ϕ^{D}:

Proposition 2.1. *Suppose that (V3) holds under the h-MaxWeight policy with V a constant multiple of h. Then the same bound holds for the h-myopic policy.*

Proof. If $V = kh$ for some $k < \infty$ we then have under the h-myopic policy,

$$\begin{aligned} P_{\phi^{\mathrm{D}}} V\,(x) &= k \mathop{\arg\min}_{u \in \mathsf{U}_\diamond(x)} \mathsf{E}[h(Q(t+1)) \mid Q(t) = x, U(t) = u] \\ &\le k\mathsf{E}[h(Q(t+1)) \mid Q(t) = x, U(t) = \phi^{\mathrm{MW}}(x)] \\ &= P_{\phi^{\mathrm{MW}}} V\,(x) \le V(x) - f(x) + b\mathbf{1}_S(x). \end{aligned}$$

□

The most common approach to establishing (V3) is to construct a function $h\colon \mathbb{Z}_+^\ell \to (0,\infty)$ and a constant $\overline{\eta} < \infty$ that solve the *Poisson inequality*,

$$\mathcal{D}h \le -c + \overline{\eta}. \tag{33}$$

If c has bounded sublevel sets (e.g. c defines a norm on $\mathbb{R}^\ell$), then this implies (V3) with $V = h$, $f = 1 + c/2$, $b = \overline{\eta} + 1$, and S is the sublevel set $S = \{x : c(x) \le 2(\overline{\eta} + 1)\}$. The Comparison Theorem implies that the steady state mean of c is bounded by $\overline{\eta}$ when (33) holds. Theorem 2.1 is also the most common approach to obtaining bounds on expectations involving stopping times. For a proof see [37].

Theorem 2.1. (Comparison Theorem) *Suppose that the non-negative functions V, f, g satisfy the bound,*

$$\mathcal{D}V \le -f + g. \tag{34}$$

Then for each $x \in \mathbb{Z}_+^\ell$ and any stopping time τ we have,

$$\mathsf{E}_x\Big[\sum_{t=0}^{\tau-1} f(Q(t))\Big] \le V(x) + \mathsf{E}_x\Big[\sum_{t=0}^{\tau-1} g(Q(t))\Big].$$

□

The average cost is finite under (V3), provided f dominates the cost function. The following result follows from Theorems 14.0.1 and 17.0.1 of [37].

Theorem 2.2. *Consider the CRW model (2) controlled using a stationary policy. Suppose that there exists a solution to (V3) satisfying $k_0^{-1}\|x\| \le c(x) \le k_0 f(x)$ for some $k_0 < \infty$ and all $x \in \mathbb{Z}_+^\ell$. Suppose moreoever that the controlled network is $\mathsf{0}$-irreducible: For each $x \in \mathbb{Z}_+^\ell$,*

$$\sum_{t=0}^{\infty} \mathsf{P}\{Q(t) = \mathsf{0} \mid Q(0) = x\} > 0, \tag{35}$$

and that $P(\mathsf{0},\mathsf{0}) = \mathsf{P}\{A(t) = \mathsf{0}\} > 0$. Then the following hold:

(i) $\boldsymbol{Q}$ *is an aperiodic Markov chain satisfying $\eta := \pi(c) < \infty$ for all $x \in \mathbb{Z}_+^\ell$.*

(ii) *The Law of Large Numbers holds: For each initial condition,*

$$\eta(n) := n^{-1} \sum_{t=0}^{n-1} c(Q(t)) \to \eta, \qquad n \to \infty, \text{ a.s.}$$

(iii) *The mean ergodic theorem holds: For each initial condition,*

$$\mathsf{E}[c(Q(t))] \to \eta, \qquad t \to \infty.$$

□

The simplest example is the CRW model for the single server queue defined by the recursion,

$$Q(t+1) = Q(t) - S(t+1)U(t) + A(t+1), \qquad t \in \mathbb{Z}_+, \tag{36}$$

with given initial condition $Q(0) = x \in \mathbb{Z}_+$. A solution to (V3) is obtained with $f(x) = 1 + x$ and $V = J^*$, where the fluid value function J^* is the quadratic,

$$J^*(x) = \tfrac{1}{2}\frac{x^2}{\mu - \alpha}. \tag{37}$$

Theorem 2.3 establishes formulae for the steady-state mean as well as the associated solution to Poisson's equation with c the identity function ($c(x) = x$ for $x \in \mathbb{R}_+$),

$$\mathcal{D}h\,(x) - x + \eta, \qquad x \in \mathbb{Z}_+. \tag{38}$$

The formula (40) for the steady-state mean may be viewed as an analog of the celebrated *Pollaczek-Khintchine formula* for the M/G/1 queue. The proof is based on refinements of the Comparison Theorem applied to the function $V = J^*$ [35, 36].

Theorem 2.3. *Consider the CRW queueing model (36) satisfying $\rho = \alpha/\mu < 1$, and define*

$$m^2 = \mathsf{E}[(S(1) - A(1))^2], \quad m_A^2 = \mathsf{E}[A(1)^2], \quad \sigma^2 = \rho m^2 + (1-\rho)m_A^2. \tag{39}$$

Then,

(i) *There is a unique invariant probability measure π on $\mathbb{Z}_+$, with steady-state mean*

$$\eta := \mathsf{E}_\pi[Q(0)] = \tfrac{1}{2}\frac{\sigma^2}{\mu - \alpha} \tag{40}$$

(ii) *A solution to Poisson's equation (38) is the quadratic,*

$$h(x) = J^*(x) + \tfrac{1}{2}\mu^{-1}\Big(\frac{m^2 - m_A^2}{\mu - \alpha}\Big)x\,, \qquad x \in \mathbb{Z}_+. \tag{41}$$

□

We now illustrate the application of Theorem 2.2 using the scheduling model (21). The MaxWeight policy is defined by (9) with h a quadratic,

$$\phi^{\mathrm{MW}}(x) = \mathop{\arg\min}_{u \in \mathsf{U}_\diamond(x)} \langle Bu + \alpha, Dx\rangle, \qquad x \in \mathbb{Z}_+^\ell, \tag{42}$$

where $D > 0$ is a diagonal matrix. Proposition 2.2 implies that ϕ^{MW} coincides with the h-myopic policy for the fluid model. This result is a special case of Proposition 2.4 that follows.

Proposition 2.2. *For each $x \in \mathbb{Z}_+^\ell$, the allocation $\phi^{\mathrm{MW}}(x)$ defined by the MaxWeight policy for the CRW scheduling model can be expressed as a solution to the linear program,*

$$\begin{aligned} \phi^{\mathrm{MW}}(x) = \quad & \arg\max \quad x^{\mathrm{T}} D(I - R^{\mathrm{T}})Mu \\ & \text{s.t.} \qquad Cu \le \mathbf{1}, \quad u \ge \mathbf{0}\,. \end{aligned} \tag{43}$$

□

Proposition 2.2 easily leads to a proof of stability.

Theorem 2.4. *Suppose that $\rho_\bullet < 1$ in the CRW scheduling model. Then, for any diagonal matrix $D > 0$, the network controlled under the MaxWeight policy satisfies (V3) with $V(x) = \frac{1}{2}x^{\mathrm{T}}Dx$ and $f(x) = 1 + \varepsilon_0|x|$ for some $\varepsilon_0 > 0$ and all $x \in \mathbb{R}_+^\ell$.*

Proof. Since (31) holds when $\rho_\bullet < 1$, there exists $\varepsilon > 0$ such that the vector v with coefficients $v_i = -\varepsilon$, $i \geq 1$, lies in V for each x. By definition there exists $u \in \mathsf{U}$ such that $(-I + R^{\mathrm{T}})Mu + \alpha = v$, so that by Proposition 2.2,

$$\langle B\phi^{\mathrm{MW}}(x) + \alpha, \nabla h\,(x)\rangle = \langle B\phi^{\mathrm{MW}}(x) + \alpha, Dx\rangle \leq v^{\mathrm{T}} Dx = -\varepsilon \sum D_{ii} x_i \leq -\varepsilon_0 |x|,$$

with $\varepsilon_0 = \varepsilon(\min_i D_{ii})$.

Note that u may not be feasible if $x_i = 0$ for some i. Thanks to Proposition 2.2 this is irrelevant since we are only seeking bounds on the value of (43).

We thus arrive at a version of the Poisson inequality,

$$\mathcal{D}_{\mathrm{MW}} h\,(x) := \mathsf{E}_{\mathrm{MW}}[h(Q(t+1)) - h(Q(t)) \mid Q(t) = x] \leq -\varepsilon_0 |x| + b_D,$$

with

$$b_D := \tfrac{1}{2} \max_{x' \in \mathbb{Z}_+^\ell, u \in \mathsf{U}_\diamond(x)} \mathsf{E}[(Q(t+1) - Q(t))^{\mathrm{T}} D (Q(t+1) - Q(t)) \mid Q(t) = x',\ U(t) = u].$$

□

2.2 Perturbed functions

We now analyze the drift $\mathcal{D}h$ represented in (28) to establish stability of the h-MaxWeight policy. We return to the general CRW model (2).

An application of the chain rule of differentiation shows that,

Proposition 2.3. *For any C^1 function h_0, the function h defined in (17) satisfies the derivative conditions (14). We have the explicit representations,*

(i) *The first derivative is given by,*

$$\nabla h\,(x) = [I - M_\theta] \nabla h_0\,(\tilde{x}), \tag{44}$$

where,

$$M_\theta = \mathrm{diag}\,(e^{-x_i/\theta}), \qquad x \in \mathbb{R}^\ell. \tag{45}$$

(ii) *If h_0 is C^2, then the Hessian of h is,*

$$\nabla^2 h\,(x) = [I - M_\theta] \nabla^2 h_0\,(x) [I - M_\theta] + \theta^{-1} \mathrm{diag}\,(M_\theta \nabla h_0\,(\tilde{x})). \tag{46}$$

Hence h is convex provided h_0 is both convex and monotone.

□

A key step in the proof of Theorem 1.1 is to generalize Proposition 2.2. At the same time, in Proposition 2.4 we express the h-MaxWeight policy in terms of *generalized Klimov indices*, determined by the weights,

$$\Theta_j(x) := -\sum_i B_{ij} \frac{\partial}{\partial x_i} h\,(x), \qquad x \in \mathbb{Z}_+^\ell,\ j \in \{1, \ldots, \ell_u\}. \tag{47}$$

Proposition 2.4. *Suppose that Assumptions (i)-(iii) of Theorem 1.1 hold, and that h is any C^1 monotone function satisfying the derivative conditions (14). Then, for each $x \in \mathbb{Z}_+^\ell$, the allocation $\phi^{\mathrm{MW}}(x)$ defined by the h-MaxWeight policy can be expressed as a solution to the linear program,*

$$\begin{aligned}\phi^{\mathrm{MW}}(x) = \quad & \arg\min \quad \langle Bu, \nabla h\,(x)\rangle \\ & \quad \text{s.t.} \qquad Cu \le \mathsf{1}, \quad u \ge \mathsf{0}\,.\end{aligned} \tag{48}$$

Moreover, the policy can be described as follows: Given $Q(t) = x$, for each station $s(j) \in \{1,\dots,\ell_m\}$ let $j_s^* \in \arg\max\{\Theta_j(x) : s(j) = s\}$. If $\Theta_{j_s^*}(x) < 0$ then this station is idle, otherwise server j_s^* receives priority. That is,

$$U_{j_s^*}(t) = \mathbf{1}\{\Theta_{j_s^*}(x) \ge 0\}. \tag{49}$$

Proof. The proof requires that we demonstrate that the minimum in (9) can be relaxed to a minimum over all of U.

The coefficient of u_j in the objective function of (48) is given by $-\Theta_j$. Monotonicity implies that each partial derivative of h is non-negative. This assumption combined with (18) implies that

$$\Theta_j(x) \le 0 \text{ whenever } x_{i_j} = 0. \tag{50}$$

It then follows that the optimizer u^* of the linear program (48) satisfies without loss of generality $u_j^* = 0$ whenever $x_{i_j} = 0$. This shows that $u^* \in \mathsf{U}(x)$ for $x \in \mathbb{Z}_+^\ell$, which proves (48).

To show that u^* can be chosen in $\mathsf{U}_\diamond$ we argue that optimizers of linear programs can be chosen among the extreme points in the constraint region. The extreme points for this linear program are all contained in $\{0,1\}^\ell$. The fact that the extreme-point optimizers are of the form (49) is then obvious. $\square$

Consider the special case in which h_0 is linear.

2.2.1 Perturbed linear function

Suppose that $h_0(x) = c^{\mathrm{T}}x$, where the vector $c \in \mathbb{R}_+^\ell$ has non-zero coefficients, so that the function h can be expressed,

$$h(x) = \sum_{i=1}^{\ell} c_i \tilde{x}_i, \qquad x \in \mathbb{R}_+^\ell. \tag{51}$$

An application of Proposition 2.3 shows that the derivative condition (14) holds, and that the first and second derivatives are given by,

$$\nabla h\,(x) = [I - M_\theta]c, \qquad \nabla^2 h\,(x) = \theta^{-1}\mathrm{diag}\,(M_\theta c). \tag{52}$$

Hence the function h is monotone and strictly convex.

We show in Proposition 2.5 that the h-MaxWeight policy is stabilizing provided $\theta \ge 1$ is suitably large.

Proposition 2.5. *Suppose that Assumptions (i)-(iii) of Theorem 1.1 hold, along with the stabilizability condition (31). Then, there exists $\theta_0 > 0$ such that the following hold for all $\theta \geq \theta_0$:*

(i) *The controlled network satisfies Foster's Criterion. The function V in (V2) can be taken as a constant multiple of h.*

(ii) *Condition (V3) holds: There exists $\varepsilon_2 > 0$, $b_2 < \infty$, and a finite set S satisfying,*

$$\mathcal{D}V \leq -f + b_2 \mathbf{1}_S$$

with $V = 1 + \frac{1}{2}h^2$, $f = 1 + \varepsilon_2 h$.

(iii) *Suppose that for some $\varepsilon > 0$ the arrival process satisfies $\mathsf{E}[\exp(\varepsilon \|A(t)\|)] < \infty$. Then Condition (V4) holds: For some $\varepsilon_e > 0$, $\delta_e > 0$, $b_e < \infty$, and a finite set S,*

$$\mathcal{D}V \leq -\delta_e V + b_e \mathbf{1}_S$$

with $V = \exp(\varepsilon_e h)$. Hence $\boldsymbol{Q}$ is geometrically ergodic provided (35) holds [37].

Proof. We apply the second-order Mean Value Theorem to obtain,

$$\begin{aligned} h(Q(t+1)) - h(Q(t)) = \langle \nabla h\,(Q(t)), \Delta(t+1) \rangle \\ + \tfrac{1}{2}\Delta(t+1)^{\mathrm{T}} \big[\nabla^2 h\,(\bar{Q})\big] \big(\Delta(t+1)\big), \end{aligned} \tag{53}$$

where again $\bar{Q} \in \mathbb{R}_+^\ell$ lies on the line connecting $Q(t+1)$ and $Q(t)$. This implies the identity (28) with b_h redefined as,

$$b_h(x) = \tfrac{1}{2}\mathsf{E}\big[\Delta(t+1)^{\mathrm{T}} \nabla^2 h\,(\bar{Q})\, \Delta(t+1) \mid Q(t) = x\big].$$

The expression for the second derivative in (52) then gives,

$$\mathcal{D}h\,(x) = \langle \nabla h\,(x), v \rangle + \theta^{-1} b_\Delta,$$

where

$$b_\Delta = \tfrac{1}{2}\|c\| \sup_{x' \in \mathbb{Z}_+^\ell, u \in \mathsf{U}_\diamond(x')} \mathsf{E}[\|\Delta(t+1)\|^2 \mid Q(t) = x', U(t) = u] < \infty.$$

We now obtain an upper bound on $\langle \nabla h\,(x), v \rangle$ under the h-MaxWeight policy. The expression for the first derivative in (52) implies the bound,

$$\frac{\partial}{\partial x_i} h\,(x) = c_i(1 - e^{-x_i/\theta}) \geq \underline{c}(1 - e^{-x_i/\theta}), \quad 1 \leq i \leq \ell,$$

with $\underline{c} := \min_j c_j$. Exactly as in the proof of Theorem 2.4 we can consider arbitrary $v \in \mathsf{V}$ to obtain bounds on the value of (48). This is justified by Proposition 2.4. The stabilizability condition (31) implies that there exists $\varepsilon > 0$ such that the vector with components $v_i = -\varepsilon$, $1 \leq i \leq \ell$, lies in V for each $x \in \mathbb{R}_+^\ell$. By definition, there exists $u \in \mathsf{U}$ satisfying $Bu + \alpha = v$. Consequently, under the h-MaxWeight policy,

$$\mathcal{D}h\,(x) \le -\varepsilon\underline{c}\max_i(1 - e^{-x_i/\theta}) + \theta^{-1}b_\Delta.$$

Suppose that $|x| \ge \ell\theta$. Then $x_i \ge \theta$ for at least one i, and we obtain the bound,

$$\mathcal{D}h\,(x) \le -\tfrac{1}{2}\varepsilon\underline{c} + \theta^{-1}b_\Delta, \qquad \text{if } |x| \ge \ell\theta. \tag{54}$$

The right hand side is negative provided $\theta > 2b_\Delta/(\varepsilon\underline{c})$. Fixing θ satisfying this bound we obtain the desired solution to (V2) with $V = 2(\varepsilon\underline{c})^{-1}h$, and $S = \{x : |x| < \ell\theta\}$. This establishes (i).

To establish (ii) we begin with the identity,

$$\tfrac{1}{2}[h(Q(t+1))]^2 - \tfrac{1}{2}[h(Q(t))]^2 = h(Q(t))(h(Q(t+1) - h(Q(t))) + \tfrac{1}{2}[h(Q(t+1)) - h(Q(t))]^2.$$

On taking conditional expectations of both sides we obtain $\mathcal{D}V\,(x) = h(x)[\mathcal{D}h\,(x)] + b_{\Delta 2}(x)$, where

$$b_{\Delta 2} = \tfrac{1}{2} \sup_{x' \in \mathbb{Z}_+^\ell, u \in \mathsf{U}_\diamond(x')} \mathsf{E}[[h(Q(t+1)) - h(Q(t))]^2 \mid Q(t) = x', U(t) = u] < \infty.$$

Applying (i) we obtain a version of the Poisson inequality (33) with this V, which implies that (V3) also holds.

Part (iii) follows from (i) combined with [37, Theorem 16.3.1] (see also [33, Theorem 4].) $\square$

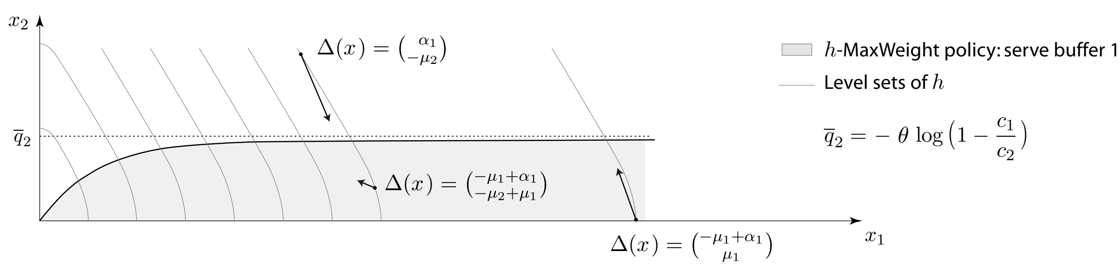

Figure 3: The perturbed cost function defined in (17) with h_0 linear on $\mathbb{R}^2$ satisfies $\min_u\langle\nabla h\,(x), Bu+\alpha\rangle < 0$ for each non-zero $x \in \mathbb{R}^2_+$. This geometry is illustrated in this figure using the tandem queues. The contour plots shown are the level sets $\{x : h(x) = r\}$ for $r = 1, 2, \dots$.

Tandem queues: Emergence of a threshold policy Suppose that we replace the linear cost function used in (13) with the convex cost function $h\colon \mathbb{R}^2_+ \to \mathbb{R}_+$ defined in (51). Figure 3 shows a plot of the sublevel sets of this function.

The h-MaxWeight policy defined in (9) minimizes the inner-product,

$$\langle Bu + \alpha, \nabla h\,(x)\rangle = -\mu_1u_1c_1(1 - e^{-x_1/\theta}) + (\mu_1u_1c_1 - \mu_2u_2c_2)(1 - e^{-x_2/\theta}) + \langle\alpha, \nabla h\,(x)\rangle.$$

The h-MaxWeight policy is thus non-idling at Station 2, and at Station 1 the policy can be expressed as a switching curve,

$$\phi_1^{\text{MW}}(x) = \mathbf{1}\{-c_1(1 - e^{-x_1/\theta}) + c_2(1 - e^{-x_2/\theta}) \le 0\}, \qquad x_1 \ge 1.$$

For small values of x_1 a first order Taylor series gives the approximation $\phi_1^{\text{MW}}(x) \approx \mathbf{1}\{x_2 \le (c_1/c_2)x_1\}$. If $x_1 \gg \theta$ then the h-MaxWeight policy can be approximated by a threshold policy, $\phi_1^{\text{MW}}(x) \approx \mathbf{1}\{x_2 \le \overline{q}_2\}$, where the threshold $\overline{q}_2$ is the solution to the equation $c_2(1 - e^{-\overline{q}_2/\theta}) = c_1$. That is,

$$\overline{q}_2 = \theta\Big|\log\Big(1 - \frac{c_1}{c_2}\Big)\Big|. \tag{55}$$

Figure 3 illustrates ϕ^{MW} when $c_1 = 1$, $c_2 = 3$, and $\theta = 10$. The asymptote (55) is $\overline{q}_2 = 10\log(3/2) \approx 4$ in this special case.

For comparison consider the discounted-cost optimal policy, minimizing

$$\sum_{t=0}^{\infty}(1+\gamma)^{-t-1}\mathsf{E}\big[c(Q(t)) \mid Q(t) = x\big],$$

for a given $\gamma > 0$. Letting $h_\gamma^*(x)$ denote the minimizing value, the optimal policy is expressed as the h_γ^*-myopic policy, and the discounted-cost dynamic programming equation holds,

$$\gamma h_\gamma^*(x) = c(x) + \min_{u\in\mathsf{U}_\diamond(x)} \mathsf{E}[h_\gamma^*(Q(t+1)) - h_\gamma^*(Q(t)) \mid Q(t) = x, U(t) = u]$$

Consider the stochastic model described by the recursion,

$$Q(t+1) = Q(t) + (-\mathbf{1}^1 + \mathbf{1}^2)U_1(t)M_1(t) - \mathbf{1}^2U_2(t)M_2(t) + \mathbf{1}^1A_1(t+1), \qquad t \ge 0,$$

in which the statistics of $\Phi(t) := (M_1(t), M_2(t), A_1(t))^{\text{T}}$, $t \ge 1$, are consistent with a model obtained through uniformization: $\boldsymbol{\Phi}$ is i.i.d., with marginal distribution defined by,

$$\mathsf{P}\{\Phi(t) = \mathbf{1}^1\} = \mu_1, \quad \mathsf{P}\{\Phi(t) = \mathbf{1}^2\} = \mu_2, \quad \mathsf{P}\{\Phi(t) = \mathbf{1}^3\} = \alpha_1, \tag{56}$$

and that $\mu_1 + \mu_2 + \alpha_1 = 1$. We take $c_1 = 1$, $c_2 = 3$, $\rho_1 = 9/11$, $\rho_2 = 9/10$.

For any finite γ the following approximation holds,

$$\lim_{r\to\infty} r^{-1}h_\gamma^*([rx]) = \gamma^{-1}c(x), \qquad x \in \mathbb{R}_+^2,$$

where $[\,\cdot\,]$ denotes the component-wise integer part of a vector. Hence for large x far from the boundary, the optimal policy coincides with the c-myopic policy. In fact, it can be shown that the optimal policy is approximated by a static threshold, similar to the policy shown in Figure 3 (see [36].) Two examples are shown in Figure 4.

2.2.2 Perturbed value function

We now consider a function h_0 that serves as a Lyapunov function for the fluid model. For example, we might take $h_0 = J^*$ the optimal fluid value function defined below (32). Our goal is to complete the proof of Theorem 1.1, which amounts to establishing (V3) in the form of the Poisson inequality (19) with $V = 2h$.

Before proving the theorem we present an example to illustrate the structure of the h-myopic policy. In simple examples, it is approximated by a switching curve with logarithmic growth.

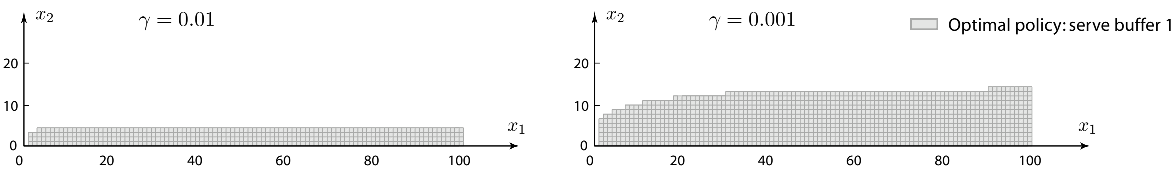

Figure 4: Discounted cost optimal policy for the tandem queues with cost parameters $(c_1, c_2) = (1, 3)$. The load parameters are $\rho_1 = 9/10$, $\rho_2 = 9/11$, and the linear cost defined by $c_1 = 1$, $c_2 = 3$. On the left $\gamma = 0.01$ and on the right $\gamma = 0.001$.

Tandem queues: translation of the optimal policy If $\rho_1 < \rho_2$ and $c_1 < c_2$ then the fluid value function is purely quadratic,

$$J^*(x) = \tfrac{1}{2}\frac{c_1}{\mu_2 - \alpha_1}(x_1 + x_2)^2 + \tfrac{1}{2}\frac{c_2 - c_1}{\mu_2}x_2^2, \qquad x \in \mathbb{R}^2_+. \tag{57}$$

Letting $h_0 = J^*$, the assumption of Theorem 1.1 (c) is satisfied with equality,

$$\min_{u \in \mathsf{U}(x)} \langle \nabla h_0\,(x), Bu + \alpha\rangle = -c(x) \qquad x \in \mathbb{R}^\ell_+.$$

The derivative conditions (14) fail, so we do not know if the h_0-MaxWeight policy is stabilizing for the CRW model.

To compute the h-MaxWeight policy we write (57) as

$$h_0(x) = J^*(x) = \tfrac{1}{2}d_1(x_1 + x_2)^2 + \tfrac{1}{2}d_2x_2^2, \qquad x \in \mathbb{R}^2_+,$$

so that the gradient of $h(x) = h_0(\tilde{x})$ can be expressed,

$$\nabla h(x) = [I - M_\theta]\nabla h_0\,(\tilde{x}) = \begin{pmatrix} d_1(\tilde{x}_1 + \tilde{x}_2)(1 - e^{-x_1/\theta}) \\ (d_1(\tilde{x}_1 + \tilde{x}_2) + d_2\tilde{x}_2)(1 - e^{-x_2/\theta}) \end{pmatrix}$$

Writing $Bu + \alpha = (-\mu_1u_1 + \alpha_1, \mu_1u_1 + \mu_2u_2)^{\mathrm{T}}$, we obtain for any $x \in \mathbb{Z}^\ell_+$, $u \in \mathsf{U}(x)$,

$$\begin{aligned}\langle \nabla h\,(x), Bu + \alpha\rangle &= \mu_1u_1\big[d_1(e^{-x_1/\theta} - e^{-x_2/\theta})(\tilde{x}_1 + \tilde{x}_2) + d_2(1 - e^{-x_2/\theta})\tilde{x}_2\big] \\ &\quad - \mu_2u_2\big[d_1(1 - e^{-x_2/\theta})(\tilde{x}_1 + \tilde{x}_2) + d_2(1 - e^{-x_2/\theta})\tilde{x}_2\big] \\ &\quad + \alpha_1d_1(1 - e^{-x_1/\theta})(\tilde{x}_1 + \tilde{x}_2)\end{aligned}$$

Minimizing over u we see that the policy is non-idling at Station 2. At Station 1 we have $u_1 = 1$ if and only if $x_1 \geq 1$ and the coefficient of u_1 is non-positive. That is, the policy at Station 1 is defined by the switching curve described by the equation,

$$d_1(e^{-x_1/\theta} - e^{-x_2/\theta})(\tilde{x}_1 + \tilde{x}_2) + d_2(1 - e^{-x_2/\theta})\tilde{x}_2 = 0. \tag{58}$$

When x_1 is large we obtain the approximation,

$$x_2 \approx s(x_1) := \theta\log\Big(1 + \frac{d_1}{d_2}x_1\Big), \tag{59}$$

where by (57),

$$\frac{d_1}{d_2} = \Big(\frac{c_2}{c_1} - 1\Big)^{-1} \frac{1}{1-\rho_2}.$$

This is an approximation to (58) in the sense that for all sufficiently large x_1 there is a unique x_2 such that (x_1, x_2) solve the equation (58), and the ratio $x_2/s(x_1)$ tends to unity as $x_1 \to \infty$.

A policy defined by a switching curve $s(x_1)$ of the form given in (59) is similar to the policy introduced in [35] to obtain HTAO (see eq. (25) and the surrounding discussion.)

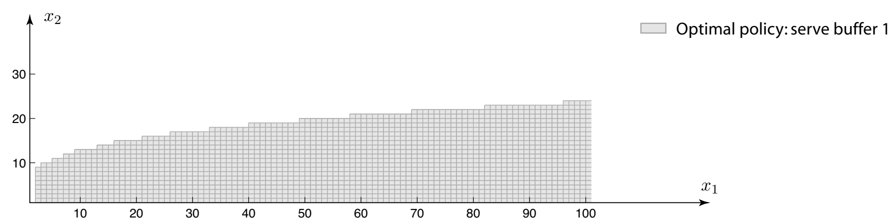

Figure 5: Average cost optimal policy for the tandem queues with $c_1 = 1$, $c_2 = 3$, $\rho_1 = 9/11$, $\rho_2 = 9/10$.

Consider now the average-cost optimal policy for the CRW model with statistics defined in (56), and linear cost with$(c_1, c_2) = (1, 3)$. It is known that the average-cost optimal policy exists, and that it is h^*-myopic with respect to the relative value function (see [6] and [36, Chapter 9].) Moreover, Theorem 7.2 of [32] implies that the following approximation holds,

$$\lim_{r\to\infty} r^{-2} h^*([rx]) = J^*(x), \qquad x \in \mathbb{R}^2_+.$$

The average-cost optimal policy for the CRW model is shown in Figure 5 with $\rho_1 = 9/11 < \rho_2$. This policy can be represented by a switching curve s that is concave and unbounded in x_1, similar to (59).

The proof of Theorem 1.1 is organized in the following two lemmas.

Lemma 2.1. *Under the assumptions of Theorem 1.1 we have under the h-MaxWeight policy, for some constant* k_1*,*

$$\langle \nabla h\,(x), v^{\mathrm{MW}} \rangle \le -c(x) + k_1 \log(1 + \|x\|), \qquad x \in \mathbb{Z}^\ell_+,$$

where $v^{\mathrm{MW}} = B\phi^{\mathrm{MW}}(x) + \alpha$*.*

Proof. Fix a constant $\beta_- \ge \theta$, and define

$$s_-(r) = \beta_- \log(1 + r/\beta_-), \qquad r \ge 0.$$

We impose the following constraint on the velocity vector v.

$$v_i \ge 0 \quad \textit{whenever} \quad x_i < s_-(\|x\|), \qquad i = 1, \dots, \ell. \tag{60}$$

The minimum of $\langle \nabla h\,(x), v\rangle$ over v satisfying these constraints provides a bound under the h-MaxWeight policy. Proposition 2.4 is critical here so that we can ignore lattice constraints and boundary constraints as we search for bounds on this inner product.

The purpose of (60) is to obtain the following bound:

$$- e^{-x_i/\theta} v_i \le |v_i| \big(1 + \|x\|/\beta_-\big)^{-\beta_-/\theta}, \qquad i = 1, \dots, \ell. \tag{61}$$

Since h_0 is assumed monotone we have $\nabla h_0 \colon \mathbb{R}^\ell_+ \to \mathbb{R}^\ell_+$, and applying (44) we obtain,

$$\langle \nabla h\,(x), v\rangle \le \langle \nabla h_0\,(\tilde{x}), v\rangle + \|v\| \|\nabla h_0\,(\tilde{x})\| \big(1 + \|x\|/\beta_-\big)^{-\beta_-/\theta}.$$

Since ∇h_0 is also Lipschitz and $\beta_- \ge \theta$, this gives for some constant k_0,

$$\langle \nabla h\,(x), v\rangle \le \langle \nabla h_0\,(\tilde{x}), v\rangle + k_0 \tag{62}$$

To bound (62) we shift $\tilde{x}$ as follows: Let $\tilde{x}^- \in \mathbb{Z}^\ell_+$ denote the vector with components

$$\tilde{x}^-_i = \lfloor (\tilde{x}_i - s_-(\|x\|))_+ \rfloor, \qquad i = 1, \dots, \ell,$$

where $\lfloor \cdot \rfloor$ denotes the integer part. In view of (15), there exists $u \in \mathsf{U}(x)$ such that with $v = Bu + \alpha$,

$$\langle \nabla h_0\,(\tilde{x}^-), v\rangle \le -c(\tilde{x}^-).$$

Moreover, we have $\tilde{x}^-_i = 0$ whenever the constraint on x_i in (60) is active. Since $u \in \mathsf{U}(x)$, this implies that the vector $v = Bu + \alpha$ satisfies $v_i \ge 0$. That is, v satisfies the constraint (60).

Using this v in (62) gives

$$\begin{aligned}
\langle \nabla h\,(x), v\rangle &\le \langle \nabla h_0\,(\tilde{x}^-), v\rangle + \langle \nabla h_0\,(\tilde{x}) - \nabla h_0\,(\tilde{x}^-), v\rangle + k_0 \\
&\le -c(\tilde{x}^-) + \|v\| \|\nabla h_0\,(\tilde{x}) - \nabla h_0\,(\tilde{x}^-)\| + k_0 \\
&\le -c(x) + |c(x) - c(\tilde{x}^-)| + \|v\| \|\nabla h_0\,(\tilde{x}) - \nabla h_0\,(\tilde{x}^-)\| + k_0.
\end{aligned}$$

This completes the proof since c and ∇h_0 are each Lipschitz. ☐

Lemma 2.2. *Under the assumptions of Theorem 1.1 we have under the h-MaxWeight policy, for some constant k_2,*

$$\mathcal{D}h\,(x) \le \langle \nabla h\,(x), v^{\mathrm{MW}}\rangle + k_2(1 + \theta^{-1}\|x\|), \qquad x \in \mathbb{Z}^\ell_+,$$

where $v^{\mathrm{MW}} = B\phi^{\mathrm{MW}}(x) + \alpha$.

Proof. The first order Mean Value Theorem (27) results in the representation (28) with b_h defined in (29). The Cauchy-Schwartz inequality gives,

$$\mathcal{D}h\,(x) \le \langle \nabla h\,(x), v^{\mathrm{MW}}\rangle + \mathsf{E}\big[\|\nabla h\,(\bar{Q}) - \nabla h\,(Q(t))\|^2 \mid Q(t) = x\big]^{\frac{1}{2}} \mathsf{E}\big[\|\Delta(t+1)\|^2 \mid Q(t) = x\big]^{\frac{1}{2}} \tag{63}$$

It remains to bound the right hand side.

Given $Q(t)=x$, an application of Proposition 2.3 gives,

$$\nabla h\,(\bar{Q}) - \nabla h\,(x) = [I - M_\theta(\bar{Q})](\nabla h_0\,(\bar{Q}) - \nabla h_0\,(\tilde{x})) + [M_\theta(x) - M_\theta(\bar{Q})]\nabla h_0\,(\tilde{x}),$$

and applying the triangle inequality, the first expectation at right in (63) is bounded as follows,

$$\begin{aligned}\mathsf{E}\big[\|\nabla h\,(\bar{Q}) - \nabla h\,(Q(t))\|^2 \mid Q(t)=x\big]^{\frac{1}{2}} &\le \mathsf{E}\big[\|[I - M_\theta(\bar{Q})](\nabla h_0\,(\bar{Q}) - \nabla h_0\,(\tilde{x}))\|^2 \mid Q(t)=x\big]^{\frac{1}{2}} \\ &\quad + \mathsf{E}\big[\|[M_\theta(x) - M_\theta(\bar{Q})]\nabla h_0\,(\tilde{x})\|^2 \mid Q(t)=x\big]^{\frac{1}{2}}\end{aligned} \tag{64}$$

To bound the first term on the right hand side of (64) we apply the Lipschitz condition on h_0: For some constant k_1,

$$\|\nabla h_0\,(\bar{Q}) - \nabla h_0\,(x)\| \le k_1 \|\Delta(t+1)\|, \qquad x \in \mathbb{Z}_+^\ell.$$

Hence the first term is bounded over x.

The second term is bounded using the Mean Value Theorem. The ith diagonal element of $[M_\theta(x) - M_\theta(\bar{Q})]$ admits the bound,

$$\begin{aligned}|e^{-x_i/\theta} - e^{-\bar{Q}_i/\theta}| &= e^{-x_i/\theta}|1 - e^{-(\bar{Q}_i - x_i)/\theta}| \\ &\le e^{-x_i/\theta}(1 - e^{-\overline{\Delta}_i/\theta})\mathbf{1}\{\bar{Q}_i > x_i\} \\ &\quad + e^{-x_i/\theta}(e^{\ell_u/\theta} - 1)\mathbf{1}\{\bar{Q}_i < x_i\}\end{aligned}$$

where $\overline{\Delta}_i = A_i(1) + \sum_j |B_{ij}(1)|$, and we have used the fact that $\sum_j B_{ij}(1) \ge -\ell_u$ under (18). The right hand side can be bounded through a second application of the Mean Value Theorem, giving

$$|e^{-x_i/\theta} - e^{-\bar{Q}_i/\theta}| \le e^{-x_i/\theta}(e^{\ell_u/\theta} - e^{-\overline{\Delta}_i/\theta}) \le \theta^{-1} e^{\ell_u/\theta}(\ell_u + \overline{\Delta}_i).$$

The Lipschitz condition on ∇h_0 and second moment conditions on $(\boldsymbol{A}, \boldsymbol{B})$ then imply that for some $k_3 < \infty$,

$$\begin{aligned}\mathsf{E}\big[\|[M_\theta(x) - M_\theta(\bar{Q})]\nabla h_0\,(\tilde{x})\|^2 \mid Q(t)=x\big]^{\frac{1}{2}} &\le \theta^{-1} e^{\ell_u/\theta}(\sqrt{\ell}\ell_u + \mathsf{E}[\|\overline{\Delta}\|^2]^{\frac{1}{2}})\|\nabla h_0\,(\tilde{x})\| \\ &\le k_3 \theta^{-1}(1 + \|x\|).\end{aligned}$$

This combined with (63) and (64) completes the proof. ☐

Proof of Theorem 1.1. Combining the bounds given in Lemma 2.1 and Lemma 2.2 gives under the h-MaxWeight policy,

$$\mathcal{D}h\,(x) \le -c(x) + k_1 \log(1 + \|x\|) + k_2(1 + \theta^{-1}\|x\|), \qquad x \in \mathbb{Z}_+^\ell.$$

Choosing $\theta > 0$ sufficiently large, we obtain a version of the Poisson inequality (33). ☐

3 Relaxations and heavy traffic

We now establish HTAO under the h-MaxWeight policy for a specifically chosen function h, and under further restrictions on the network model.

Throughout this section we consider the homogeneous scheduling model (21) subject to the following conventions. The load parameters defined in (22) are expressed,

$$\rho_i = \lambda_i/\mu_i, \qquad 1 \le i \le \ell_m,$$

where μ_i is the common mean of $\{M_j(t) : s(j) = i\}$, and λ_i is the ith component of $C[I - R^{\mathrm{T}}]^{-1}\alpha$. It is assumed throughout this section that $\rho_1 = \max_{1\le i\le \ell} \rho_i$, so that $\rho_\bullet = \rho_1$. This can be achieved by choice of indices. We let $\xi \in \mathbb{Z}_+^\ell$ denote the first column of the $\ell \times \ell_m$ matrix $[I - R]^{-1}C^{\mathrm{T}}$, so that $\xi^{\mathrm{T}}\alpha = \lambda_1$.

Homogeneity implies that the random variables $\{M_j(t) : s(j) = 1\}$ are all identical; We let $S_1(t)$ denote their common value, and we let $L_1(t) = \xi^{\mathrm{T}} A(t)$. The one-dimensional workload process $W(t) = \langle \xi, Q(t)\rangle$ evolves as,

$$W(t+1) = W(t) - S_1(t+1) + S_1(t+1)\boldsymbol{\iota}(t) + L_1(t+1), \qquad t \ge 0, \tag{65}$$

where $\boldsymbol{\iota}(t) := 1 - [CU(t)]_1$.

The one-dimensional relaxation is defined on the same probability space with $\boldsymbol{Q}$, and evolves as a controlled random walk analogous to (65):

$$\widehat{W}(t+1) = \widehat{W}(t) - S_1(t+1) + S_1(t+1)\widehat{\boldsymbol{\iota}}(t) + L_1(t+1), \qquad t \ge 0. \tag{66}$$

The idleness process $\{\widehat{\boldsymbol{\iota}}(t)\}$ is assumed to be non-negative, and adapted to $\{\widehat{W}(t), S_1(t), L_1(t)\}$. The relaxation is denoted $\widehat{\boldsymbol{W}}^*$ when controlled using the non-idling policy, $\widehat{\boldsymbol{\iota}}^*(t) = \mathbf{1}\{\widehat{W}^*(t) \ge 1\}$. In this case we have (23), provided each process has the common initialization $\widehat{W}^*(0) = W(0) = \langle \xi, Q(0)\rangle$, which we assume henceforth.

For a convex cost function $c\colon \mathbb{R}_+^\ell \to \mathbb{R}_+$ we define the *effective cost* $\overline{c}\colon \mathbb{R}_+ \to \mathbb{R}_+$ as the value of the convex program,

$$\begin{aligned} \overline{c}(w) = \min \quad & c(x) \\ \text{s.t.} \quad & \xi^{\mathrm{T}} x = w,\ x \in \mathbb{R}_+^\ell. \end{aligned} \tag{67}$$

For each $w \in \mathbb{R}_+$ an *effective state* $\mathcal{X}^*(w)$ is defined to be any vector $x^* \in \mathbb{R}_+^\ell$ that achieves the minimum in (67):

$$\mathcal{X}^*(w) = \underset{x\in\mathbb{R}_+^\ell}{\arg\min}\Big(c(x) : \xi^{\mathrm{T}} x = w\Big). \tag{68}$$

It follows from the definitions that the following bound holds for all t:

$$c(Q(t)) \ge \overline{c}(W(t)) \ge \overline{c}(\widehat{W}^*(t)). \tag{69}$$

3.1 Starvation and relaxations

It is helpful to consider a workload relaxation to see how starvation arises under a myopic policy.

3.1.1 Linear cost function

If $c(x) = c^{\mathrm{T}} x$, $x \in \mathbb{R}^{\ell}_{+}$, then the effective state in a one-dimensional relaxation is given by

$$\mathcal{X}^*(w) = \Bigl(\frac{1}{\xi_{i^*}} 1^{i^*}\Bigr) w, \qquad w \in \mathbb{R}_+,$$

where the index i^* is any solution to $c_{i^*}/\xi_{i^*} = \min_{1\le i\le \ell}\bigl(c_i/\xi_i\bigr)$. The effective cost is given by the linear function, $\overline{c}(w) = c(\mathcal{X}^*(w)) = (c_{i^*}/\xi_{i^*})w$, $w \in \mathbb{R}_+$.

This underlines the conflict that arises frequently in network optimization: Optimization of an idealized model dictates zero inventory at various stations, while in a more realistic model, adopting a "zero-inventory policy" results in starvation of resources.

3.1.2 Quadratic cost function

If $c\colon \mathbb{R}^{\ell}_{+} \to \mathbb{R}_+$ is quadratic, of the form $c(x) = \frac{1}{2} x^{\mathrm{T}} D x$, $x \in \mathbb{R}^{\ell}_{+}$ for a symmetric matrix D, then the effective state is again linear in the workload value in any one-dimensional relaxation.

For the scheduling model considered in this section the workload vector ξ has non-negative entries. Suppose that D^{-1} also has non-negative entries. In this case we have the explicit expression,

$$\mathcal{X}^*(w) = \Bigl(\bigl(\xi^{\mathrm{T}} D^{-1} \xi\bigr)^{-1} D^{-1}\xi\Bigr) w, \qquad w \in \mathbb{R}_+,$$

and the effective cost is the one-dimensional quadratic,

$$\overline{c}(w) = \tfrac{1}{2}\bigl(\xi^{\mathrm{T}} D^{-1}\xi\bigr)^{-1} w^2, \qquad w \in \mathbb{R}_+.$$

For example, if $D > 0$ is diagonal, and ξ has strictly positive entries, then the effective state $\mathcal{X}^*(w)$ has strictly positive entries for any $w > 0$. In conclusion, the conflict observed for linear cost functions does not arise when using a quadratic function satisfying these conditions.

3.2 Logarithmic regret

We now construct a policy satisfying (78) with c a linear cost function. The policy is defined as the h-MaxWeight policy for a specific function h.

We saw in Section 3.1.1 that the effective state $x^* = \mathcal{X}^*(w)$ can be constructed so that $x^*_i = 0$ for all but one $i \in \{1, \ldots, \ell\}$. By choice of indices we assume that $x^*_i = 0$ for $i \ge 2$ and any $w \in \mathbb{R}_+$. Consequently, the effective cost is given by,

$$\overline{c}(w) = c(\mathcal{X}^*(w)) = \frac{c_1}{\xi_1} w, \qquad w \in \mathbb{R}_+. \tag{70}$$

We assume moreover that the solution to (67) is unique, which amounts to the following strict bound,

$$\frac{c_i}{\xi_i} > \frac{c_1}{\xi_1}, \qquad i = 2, \ldots, \ell. \tag{71}$$

Theorem 2.3 implies the following formula for the steady-state cost for the relaxation,

$$\widehat{\eta}^* = \tfrac{1}{2} \frac{\sigma^2_{\bullet}}{\mu_1 - \lambda_1} \frac{c_1}{\xi_1}, \tag{72}$$

where $\sigma_\bullet^2 := \rho_\bullet \mathsf{E}[(S_1(1) - L_1(1))^2] + (1-\rho_\bullet)\mathsf{E}[(L_1(1))^2]$. From (69) we evidentally have $\widehat{\eta}^* \le \eta^*$. To establish (24) we require a bound in the reverse direction.

We now introduce a family of network models, parameterized by a scalar $\kappa \in [1, \infty)$ that represents load. It is assumed that, for some fixed $\kappa_0 \ge 1$, $\bar{\rho} < 1$ we have

$$\begin{aligned} \rho_\bullet := \rho_1 = 1 - 1/\kappa \qquad & \text{for each } \kappa \in [\kappa_0, \infty), \\ \textit{and} \quad \rho_i \le \bar{\rho} \qquad & \text{for each } \kappa \text{ and } i \ge 2. \end{aligned} \tag{73}$$

The one-dimensional workload process $\boldsymbol{W}$ is given in (65), where we suppress the dependency of $\{\boldsymbol{Q}, \boldsymbol{W}, \boldsymbol{S}, \boldsymbol{L}\}$ on the parameter κ to simplify notation.

The fluid model for the one-dimensional workload process is expressed $\frac{d^+}{dt} w(t) = -(\mu_1 - \lambda_1) + \iota(t)$ with $\iota(t) \ge 0$. Given the cost function $\overline{c}$ given in (70), the fluid value function is given by,

$$\widehat{J}^*(w) = \tfrac{1}{2} \frac{c_1}{\xi_1} \frac{w^2}{\mu_1} \kappa, \qquad w \ge 0. \tag{74}$$

The solution to Poisson's equation (41) is the sum of $\widehat{J}^*$ and a linear function of w. We take the function h_0 used to define the h-MaxWeight policy as a different perturbation of $\widehat{J}^*$: Fix a positive constant $b > 0$, and define

$$h_0(x) = \widehat{J}^*(\xi^{\mathrm{T}} x) + \tfrac{1}{2} b \big(c(x) - \overline{c}(\xi^{\mathrm{T}} x)\big)^2. \tag{75}$$

We then take $h(x) = h_0(\tilde{x})$, or

$$h(x) := \widehat{J}^*(\widetilde{w}) + \tfrac{1}{2} b \big(c(\tilde{x}) - \overline{c}(\widetilde{w})\big)^2, \qquad x \in \mathbb{R}_+^\ell, \tag{76}$$

where $\tilde{x}$ is the ℓ-dimensional vector with components given in (16), and $\widetilde{w} := \sum \xi_i \tilde{x}_i$.

For each κ we denote by $\eta = \eta(\kappa)$ the steady state cost under the policy for the CRW model, and $\widehat{\eta}^*$ the optimal average cost for the one-dimensional relaxation. Applying (73), the representation (72) becomes,

$$\widehat{\eta}^* = \tfrac{1}{2} \frac{\sigma_\bullet^2}{\mu_1} \frac{c_1}{\xi_1} \kappa. \tag{77}$$

Note that η and $\widehat{\eta}^*$ are each unbounded as the network load $\rho_\bullet$ approaches unity, and $\widehat{\eta}^*$ is of order κ. Hence (24) implies the bounds,

$$\eta^* \le \eta \le \eta^* + k_* \log(\kappa), \qquad\qquad \kappa \ge \kappa_0. \tag{78}$$

Theorem 3.1. *Suppose that the following hold for the parameterized family of CRW scheduling models.*

(i) *The effective state is unique: Equation (71) holds.*

(ii) *The load parameters satisfy (73). Hence $\rho^\kappa \to \rho^\infty$ as $\kappa \to \infty$, where $\rho_1^\infty = 1$ and $\rho_i^\infty < 1$ for $i \ge 2$.*

(iii) *The random variables $\{A^\kappa(t), S_s^\kappa(t) : t \ge 1, \kappa \in [1, \infty], s \ge 1\}$ are defined on a common probability space, and are monotone in κ:*

$$S_s^\kappa(t) \downarrow S_s^\infty(t), \quad A^\kappa(t) \uparrow A^\infty(t), \qquad \kappa \to \infty, \ \text{a.s.}, \ t \ge 1, \ s \in \{1, \dots, \ell_m\}.$$

(iv) *For each $\kappa \geq 1$ and each station $s \in \{1, \ldots, \ell_m\}$ the distribution of $S_s^\kappa(t)$ is Bernoulli. The distribution of $A^\kappa(t)$ is supported on $\mathbb{Z}_+^\ell$ and satisfies $\sup_\kappa \mathsf{E}[\|A^\kappa(t)\|^2] < \infty$. The joint process $(\boldsymbol{A}^\kappa, \boldsymbol{S}^\kappa)$ satisfies, for some $\varepsilon > 0$ independent of κ and s,*

$$\mathsf{P}\{S_s^\kappa(t) = 1 \text{ and } A^\kappa(t) = \mathsf{0}\} \geq \varepsilon. \tag{79}$$

Then, there exists $\theta_0 > 0$ and $b_0 > 0$ such that the following conclusions hold for the h-MaxWeight policy with h defined in (76), with $\theta \geq \theta_0$ and $b \geq b_0$: There exists $\kappa_0 > 0$ such that the controlled network is ergodic, in the sense that it is $\mathsf{0}$-irreducible and (V3) holds, for each $\kappa \geq \kappa_0$. Moreoever, the family of controlled networks satisfies the bound (78) for some fixed $k_* < \infty$. That is, the policy is heavy-traffic asymptotically optimal with logarithmic regret.

The proof is based on the construction of a Lyapunov function $V\colon \mathbb{Z}_+^\ell \to \mathbb{R}_+$ satisfying a refinement of (V3),

$$\mathsf{E}_x[V(Q(1))] \leq V(x) - c(x) + \widehat{\eta}^* + \mathcal{E}(x), \qquad x \in \mathbb{Z}_+^\ell, \tag{80}$$

where the error $\mathcal{E}$ has at most logarithmic growth,

$$\mathcal{E}(x) = k_e\Big(\log(\kappa + c(x)) + \kappa/(1 + (\xi^{\mathrm{T}} x)^2)\Big), \qquad x \in \mathbb{R}_+^\ell,\ \kappa \geq 1. \tag{81}$$

This is achieved using the same steps used in Section 2.2 to establish stability of the h-MaxWeight policy: First we obtain a bound on the term $\langle \nabla h\,(x), v\rangle$ appearing in (28). We then decompose the term $b_h(x)$ into a bounded term, and a term whose mean is equal to $\widehat{\eta}^*$.

3.3 Proof of Theorem 3.1

Throughout this section we assume that the assumptions of Theorem 3.1 hold. We let $\mathcal{E}$ denote the function of x and κ given in (81). The constant k_e may differ in each appearance.

We first establish irreducibility:

Proposition 3.1. *Under the assumptions of Theorem 3.1 the policy ϕ^{MW} is $\mathsf{0}$-irreducible for each $\kappa < \infty$.*

Proof. This follows from Proposition 2.2, together with convexity and monotonicity of the function h.

Monotonicity implies that the h-MaxWeight policy is "weakly non-idling': For any t,

$$\sum_{i=1}^{\ell} U(t) \geq 1 \ \text{ whenever } Q(t) \neq \mathsf{0}.$$

Combining this with (79) we can conclude that for some $\delta > 0$, and any non-zero $x \in \mathsf{X}_\diamond$,

$$\mathsf{P}\big\{\text{A service is completed at time } t \text{ and } A(t) = \mathsf{0} \mid Q(t-1) = x\big\} \geq \delta, \qquad t \geq 1.$$

Since each customer in the network requires service at most ℓ times, it follows that $P^T(x, \mathsf{0}) \geq \delta^T$ for $T = \ell|x|$. This establishes $\mathsf{0}$-irreducibility.

Aperiodicity also follows from (79) since $P(\mathsf{0}, \mathsf{0}) = \mathsf{P}\{A(t) = \mathsf{0}\} > 0$. ◻

Recall that the proof of Theorem 1.1 was based on Lemmas 2.1 and 2.2. The following two propositions are refinements of these results.

Proposition 3.2. *Under the h-MaxWeight policy we have for some $\varepsilon > 0$, independent of b and x,*

$$\langle \nabla h\,(x), v^{\mathrm{MW}} \rangle \leq -\overline{c}(w) - \varepsilon\Big(b|c(x) - \overline{c}(w)| + \|M_\theta(x)\xi\|\kappa w\Big) + \mathcal{E}(x), \tag{82}$$

where $w = \xi^{\mathrm{T}}x$ and $v^{\mathrm{MW}} = B\phi^{\mathrm{MW}}(x) + \alpha$. □

Proposition 3.3. *Under the h-MaxWeight policy we have for some $k_2 < \infty$, independent of b, κ, and x,*

$$\mathcal{D}h\,(x) \leq \langle \nabla h\,(x), v^{\mathrm{MW}} \rangle + \widehat{\eta}^*(x) + k_2\theta^{-1}\Big(b|c(x) - \overline{c}(w)| + \|M_\theta(x)\xi\|\kappa w\Big) + \mathcal{E}(x), \tag{83}$$

where $v^{\mathrm{MW}} = B\phi^{\mathrm{MW}}(x) + \alpha$ and

$$\widehat{\eta}^*(x) := \tfrac{1}{2}\frac{\kappa}{\mu_1}\frac{c_1}{\xi_1}\sigma_\bullet^2(x), \quad \textit{with} \quad \sigma_\bullet^2(x) := \mathsf{E}[(\xi^{\mathrm{T}}\Delta(t+1))^2 \mid Q(t) = x]$$

□

We begin with the proof of Proposition 3.2. Note that,

$$c(x) - \overline{c}(w) = \sum_{i=2}^{\ell}\big(c_i - (c_1/\xi_1)\xi_i\big)x_i, \tag{84}$$

which is non-negative by (71). That is, we have $|c(x) - \overline{c}(w)| = c(x) - \overline{c}(w)$. To prove the proposition we first apply Proposition 2.3 to obtain a representation for the gradient of h,

$$\nabla h\,(x) = \nabla \widehat{J}^*(\widetilde{w})[I - M_\theta]\xi + b(c(\tilde{x}) - \overline{c}(\widetilde{w}))[I - M_\theta]\big(c - \frac{c_1}{\xi_1}\xi\big), \tag{85}$$

with

$$\nabla \widehat{J}^*(w) = \frac{c_1}{\xi_1}\frac{w}{\mu_1 - \lambda_1} = \frac{c_1}{\xi_1}\frac{\kappa}{\mu_1}w, \qquad w \geq 0.$$

A representation for the drift easily follows:

Lemma 3.1. *For any $v \in \mathsf{V}$, $x \in \mathbb{R}^\ell_+$ we have,*

$$\begin{aligned}\langle \nabla h\,(x), v \rangle = \nabla \widehat{J}^*(\widetilde{w})\langle \xi, v\rangle - \nabla \widehat{J}^*(\widetilde{w}) \sum_{i=1}^{\ell} e^{-x_i/\theta}\xi_i v_i& \\ + [c(\tilde{x}) - \overline{c}(\widetilde{w})]\sum_{i=2}^{\ell}(1 - e^{-x_i/\theta})b_i v_i&.\end{aligned} \tag{86}$$

with $b_i = b(c_i - (c_1/\xi_1)\xi_i)$, $i \geq 2$. □

Fix constants $\beta_+ > \beta_- > 0$, and define for $r \ge 0$,

$$s_-(r) = \beta_- \log(1 + r/\beta_-), \quad s_+(r) = \beta_+ \log(1 + r/\beta_+). \tag{87}$$

It is assumed throughout that $\beta_- \ge 3\theta^{-1}$.

Similar to the proof of Lemma 2.1, to bound (86) we impose the following constraint on the velocity vector v.

$$v_i \ge 0 \quad \textit{if } x_i \le s_-(\xi^{\mathrm{T}} x) - \beta_- \log(|\xi|) \textit{ and } \xi_i > 0, \tag{88}$$

where $|\xi| := \sum \xi_i$. The minimum of $\langle \nabla h(x), v \rangle$ over v satisfying (88) provides a bound under the h-MaxWeight policy. The following two results imply that (88) is feasible for a policy that is non-idling at Station 1.

Lemma 3.2. *For each $x \in \mathbb{R}^\ell_+$ we have $\max\{x_i : \xi_i > 0\} \ge s_-(\xi^{\mathrm{T}} x) - \beta_- \log(|\xi|)$.*

Proof. Letting $x^\infty = \max\{x_i : \xi_i > 0\}$, we have $\xi^{\mathrm{T}} x \le x^\infty |\xi|$, and hence by concavity of the logarithm, with $w = \xi^{\mathrm{T}} x$,

$$\begin{aligned} s_-(w) = \beta_- \log(1 + w/\beta_-) &\le \beta_- \log(1 + x^\infty |\xi|/\beta_-) \\ &\le \beta_- \Big(\log(|\xi|) + (1 + x^\infty |\xi|/\beta_- - |\xi|)/|\xi| \Big). \end{aligned}$$

The right hand side is bounded above by $\beta_- \log(|\xi|) + x^\infty$ since $|\xi| \ge 1$. This gives the desired bound. ☐

We now establish a set of feasible values for v.

Lemma 3.3. *There exists $\kappa_v \ge 1$ and $\varepsilon_v > 0$ such that for each $\kappa \ge \kappa_v$ we have*

$$\big\{ v : \|v\| \le \varepsilon_v \quad \textit{and} \quad \langle \xi, v \rangle \ge -(\mu_1 - \lambda_1) \big\} \subset \mathsf{V}.$$

Proof. The velocity space V^κ is a polyhedron for each κ, and as $\kappa \to \infty$ these sets converge to a polyhedron whose interior is non-empty, with a single face meeting the origin given by $\{\xi^{\mathrm{T}} v = 0\}$. ☐

Proof of Proposition 3.2. The drift obtained with any $v \in \mathsf{V}$ provides a bound on the drift obtained using the h-MaxWeight policy. As in the proof of Lemma 2.1, we apply Proposition 2.4 to relax lattice constraints and boundary constraints in our construction of v.

We take $v \in \mathsf{V}$ of the specific form $v = v^* + v^0$ with $v^* = -(\mu_1 - \lambda_1)\xi_1^{-1} \mathbf{1}^1$ and v^0 orthogonal to ξ so that $\xi^{\mathrm{T}} v = \xi^{\mathrm{T}} v^* = -(\mu_1 - \lambda_1)$.

We fix $\varepsilon_0 > 0$ and choose $v^0_i = -\varepsilon_0$ for each i satisfying $\xi_i = 0$. We take $v^0_i = 0$ for all but (at most) three indices satisfying $\xi_i > 0$. In Cases (ii) and (iii) below there are just two non-null indices, denoted $i_\ominus$ and $i_\oplus$, with $v^0_{i_\ominus} < 0$ and $v^0_{i_\oplus} > 0$.

It is assumed that $|v^0_i| \le \varepsilon_0$ for each i. Lemma 3.3 implies that we can choose $\varepsilon_0 > 0$ independent of κ so that any such v lies in V^κ for each κ.

To complete the specification of v we introduce the index sets,

$$
\begin{aligned}
I_\oplus &= \{i \ge 1 : \xi_i > 0,\ x_i \le s_-(\xi^{\mathrm T} x) - \beta_- \log(|\xi|)\} \\
I_\ominus &= \{i \ge 2 : \xi_i > 0,\ x_i > s_+(\xi^{\mathrm T} x)\}
\end{aligned}
$$

The choice of v^0 depends upon these sets as follows:

(i) If $I_\oplus = \emptyset$ and $I_\ominus = \emptyset$ then $v^0_i = 0$ for all i satisfying $\xi_i > 0$.

(ii) If $I_\oplus = \emptyset$ and $I_\ominus \neq \emptyset$ then we take $i_\ominus \in I_\ominus$ arbitrary with $v^0_{i_\ominus}\xi_{i_\ominus} = -\varepsilon_0$, and $i_\oplus = 1$ with $v^0_{i_\oplus}\xi_{i_\oplus} = \varepsilon_0$.

(iii) If $I_\oplus \neq \emptyset$ and $I_\ominus = \emptyset$ then we take

$$
i_\oplus \in \arg\min\{x_i : i \in I_\oplus\}, \quad i_\ominus = 1,
$$

with $v^0_{i_\oplus}\xi_{i_\oplus} = \varepsilon_0^2$ and $v^0_{i_\ominus}\xi_{i_\ominus} = -\varepsilon_0^2$.

(iv) If $I_\oplus \neq \emptyset$ and $I_\ominus \neq \emptyset$ then

$$
i_\oplus \in \arg\min\{x_i : i \in I_\oplus\}, \quad i_\ominus \in \arg\max\{x_i : i \ge 2,\ \xi_i > 0\},
$$

and we also take $v^0_1 > 0$ with,

$$
v^0_{i_\oplus}\xi_{i_\oplus} = \varepsilon_0^2, \quad v^0_1\xi_1 = \varepsilon_0 - \varepsilon_0^2, \quad v^0_{i_\ominus}\xi_{i_\ominus} = -\varepsilon_0.
$$

The added complexity in the second two cases is due to the positive drift induced by $v_{i_\oplus}$. By imposing the constraint that this is of order ε_0^2 rather than ε_0 we can maintain a negative overall drift.

This choice of v satisfies (88). Moreover, under the assumption that $\beta_- \ge 3\theta^{-1}$ we have for some constant k_e and all x,

$$
\big|\nabla \widehat{J}^*(\widetilde{w}) e^{-x_i/\theta}\xi_i v_i\big| \le k_e \kappa (1 + w^2)^{-1} \le \mathcal{E}(x), \qquad i \notin I_\oplus. \tag{89}
$$

If $I_\oplus$ is non-empty then by our choice of v we must have for some $\varepsilon_1 > 0$,

$$
-\nabla \widehat{J}^*(\widetilde{w}) e^{-x_\oplus/\theta}\xi_\oplus v_\oplus \le -\varepsilon_1 \|M_\theta(x)\xi\| \kappa w.
$$

From Lemma 3.1 and (89) we obtain the bound,

$$
\begin{aligned}
\langle \nabla h\,(x), v\rangle \le -(\mu_1 - \lambda_1)\nabla \widehat{J}^*(\widetilde{w}) - \varepsilon_1 \|M_\theta(x)\xi\|\kappa w \\
+ [c(\tilde{x}) - \overline{c}(\widetilde{w})] \sum_{i=2}^{\ell} (1 - e^{-x_i/\theta}) b_i v_i + \mathcal{E}(x).
\end{aligned}
$$

To complete the proof we argue that the following bound holds: For some $\varepsilon_2 > 0$,

$$
[c(\tilde{x}) - \overline{c}(\widetilde{w})] \sum_{i=2}^{\ell} (1 - e^{-x_i/\theta}) b_i v_i \le -\varepsilon_2 b [c(\tilde{x}) - \overline{c}(\widetilde{w})] + \mathcal{E}(x). \tag{90}
$$

It is here that we require the bound $v_i \leq \varepsilon_0^2$ for each $i \geq 2$, and the fact that at most one value of v_i is positive.

If $x_i > s_+(\xi^{\mathrm{T}} x)$ for some $i \geq 2$ (not necessarily satisfying $\xi_i > 0$) then $v_i = -\varepsilon_0$ for some $i \geq 2$. In fact, with $i_- \in \arg\max\{x_i : i \geq 2\}$ we have $v_{i_-} = -\varepsilon_0$, and from (84) we obtain $c(x) - \overline{c}(w) \leq |c| x_{i_-}$. Consequently,

$$\sum_{i=2}^{\ell} (1 - e^{-x_i/\theta}) b_i v_i \leq -(b_{i_-}\varepsilon_0 - b_{i_\oplus}\varepsilon_0^2) + \varepsilon_0 b_{i_-} e^{-[c(x) - \overline{c}(w)]/(|c|\theta)}$$

The bound (90) follows for $\varepsilon_0 > 0$ sufficiently small: fix $\varepsilon_0 < \min_{i,j \geq 2} b_i/b_j$ and set $\varepsilon_2 = \min_{i,j\geq 2}(b_i\varepsilon_0 - b_j\varepsilon_0^2)/b$. Note that the positive term $b_{i_\oplus}\varepsilon_0^2$ is absent in cases (i) or (ii), so that we are considering the worst case in which $I_\oplus \neq \emptyset$.

If $x_i \leq s_+(\xi^{\mathrm{T}} x)$ for each $i \geq 2$ then it may be impossible to guarantee the negative drift $v_i = -\varepsilon_0$ for any $i \geq 2$. But this is irrelevant since in this case,

$$[c(\tilde{x}) - \overline{c}(\widetilde{w})] \leq \mathcal{E}(x),$$

so that (90) follows trivially. □

Proof of Proposition 3.3. We begin with a representation of the form (28) based on a second-order Mean Value Theorem of the form (53). We write $h_0(x) = \frac{1}{2} x^{\mathrm{T}} H_0 x$, with

$$H_0 = \frac{\kappa}{\mu_1}\frac{c_1}{\xi_1}\xi\xi^{\mathrm{T}} + b(c - (c_1/\xi_1)\xi)(c - (c_1/\xi_1)\xi)^{\mathrm{T}}.$$

Based on this expression combined with the Mean Value Theorem we obtain,

$$\begin{aligned}\mathcal{D}h\,(x) = \langle \nabla h\,(x), v\rangle &+ \tfrac{1}{2}\mathsf{E}\big[\Delta(t+1)^{\mathrm{T}} H_0 \Delta(t+1) \mid Q(t) = x\big] \\ &+ \tfrac{1}{2}\mathsf{E}\big[\Delta(t+1)^{\mathrm{T}}\big(\nabla^2 h\,(\bar{Q}) - H_0\big)\Delta(t+1) \mid Q(t) = x\big].\end{aligned} \tag{91}$$

We also have by definition of $\widehat{\eta}^*(x)$,

$$\mathsf{E}\big[\Delta(t+1)^{\mathrm{T}} H_0 \Delta(t+1) \mid Q(t) = x\big] = \widehat{\eta}^*(x) + b\mathsf{E}\big[\big((c - (c_1/\xi_1)\xi)^{\mathrm{T}}\Delta(t+1)\big)^2 \mid Q(t) = x\big]. \tag{92}$$

We apply Proposition 2.3 to bound the final term in (91):

$$\nabla^2 h\,(x) - H_0 = -[M_\theta H_0 + M_\theta H_0] + M_\theta H_0 M_\theta + \theta^{-1}\mathrm{diag}\,(M_\theta \nabla h_0(\tilde{x})). \tag{93}$$

We have for any $\Delta \in \mathbb{R}^\ell$,

$$\begin{aligned}\Delta^{\mathrm{T}}\big[&-[M_\theta H_0 + M_\theta H_0] + M_\theta H_0 M_\theta\big]\Delta \\ &= \kappa c_1/(\mu_1\xi_1)\big[-2(\Delta^{\mathrm{T}}\xi)(\Delta^{\mathrm{T}} M_\theta \xi) + (\Delta^{\mathrm{T}} M_\theta\xi)^2\big] + O(1) \\ &= \kappa c_1/(\mu_1\xi_1)\big[-(\Delta^{\mathrm{T}} M_\theta\xi) + 2(\Delta^{\mathrm{T}} M_\theta\xi)\big((\Delta^{\mathrm{T}} M_\theta\xi) - (\Delta^{\mathrm{T}}\xi)\big)\big] + O(1) \\ &\leq \kappa c_1/(\mu_1\xi_1)\big[-(\Delta^{\mathrm{T}} M_\theta\xi) + 2\|\Delta\|^2\|M_\theta\xi\|\|(I - M_\theta)\xi\|\big] + O(1)\end{aligned}$$

Applying the Mean Value Theorem as in the proof of Lemma 2.2 we obtain the crude bound, $\|(I - M_\theta)\xi\| \leq \theta^{-1}\kappa w$, and hence for some $k_0 < \infty$,

$$-[M_\theta H_0 + M_\theta H_0] + M_\theta H_0 M_\theta \le k_0(\theta^{-1}\|M_\theta \xi\|\kappa w + 1)I$$

Also, for a possibly larger constant k_0,

$$\begin{aligned}\|M_\theta \nabla h_0(\tilde{x})\| &= \|M_\theta\bigl(\kappa\mu_1^{-1}\overline{c}(\widetilde{w})\xi + b(c(\tilde{x}) - \overline{c}(\widetilde{w}))(c - (c_1/\xi_1)\xi)\bigr)\| \\ &\le k_0\bigl(\|M_\theta\xi\|\kappa w + b|c(\tilde{x}) - \overline{c}(\widetilde{w})|\bigr).\end{aligned}$$

Consequently, for some $k_0 < \infty$,

$$\nabla^2 h\,(x) - H_0 \le k_0\theta^{-1}\bigl(\kappa\|M_\theta\xi\|w + b|c(\tilde{x}) - \overline{c}(\widetilde{w})|\bigr)I + k_0 I.$$

This combined with (91) and (92) completes the proof. □

Proof of Theorem 3.1. Following Proposition 3.2 and Proposition 3.3, the proof of the theorem amounts to establishing the drift (80) for a function V derived from h. We define,

$$V(x) = h(x) + \tfrac{1}{2}\mu^{-1}\Bigl(\frac{m^2 - m_L^2}{\mu - \alpha}\Bigr)w\,, \qquad x \in \mathbb{Z}_+^\ell,$$

where $m^2 := \mathsf{E}[(S_1(1) - L_1(1))^2]$ and $m_L^2 \mathsf{E}[(L_1(1))^2]$. That is, we are re-introducing the linear term appearing in the solution to Poisson's equation for the relaxation. It is easy to show that for any policy,

$$\mathsf{E}\Bigl[\tfrac{1}{2}\mu^{-1}\Bigl(\frac{m^2 - m_L^2}{\mu - \alpha}\Bigr)\Bigl(W(t+1) - W(t)\Bigr) \mid Q(t) = x\Bigr] = \widehat{\eta}^* - \widehat{\eta}^*(x).$$

Hence the function V does satisfy (80).

This bound implies that (V3) holds, so that $\pi(c)$ is finite for any finite κ. An application of the Comparison Theorem gives,

$$\pi(c) \le \widehat{\eta}^* + \pi(\mathcal{E}).$$

From the form of $\mathcal{E}$ it follows that $\pi(c)$ is bounded by a constant times κ. Also, by Jensen's inequality,

$$\pi(c) \le \widehat{\eta}^* + k_e \log(\kappa + \pi(c)) + k_e\kappa\mathsf{E}_\pi[(1 + (\xi^{\mathrm{T}}Q(t)))^2)^{-1}]$$

so that for a possibly larger constant,

$$\pi(c) \le \widehat{\eta}^* + k_e \log(\kappa) + k_e\kappa\mathsf{E}_\pi[(1 + (\xi^{\mathrm{T}}Q(t)))^2)^{-1}].$$

Moreover, applying (23) we obtain,

$$\pi(c) \le \widehat{\eta}^* + k_e \log(\kappa) + k_e\kappa\mathsf{E}[(1 + (\widehat{W}^*(t))^2)^{-1}],$$

where the expectation is taken for the steady-state relaxation. Lemma A.2 of [35] implies that $\kappa\mathsf{E}[(1 + (\widehat{W}^*(t))^2)^{-1}]$ is uniformly bounded in κ, so this final bound completes the proof. □

4 Conclusions

The generalized MaxWeight policies proposed in this paper can be designed to capture all of the desirable features observed in Tassiulus' original policy. Depending upon the structure of h, the policy can be designed to depend only on local information as in the standard algorithm, or it can utilize more information if available.

A shortcoming of the perturbation technique used here is that the resulting policies are guaranteed to be stable only for $\theta > 0$ sufficiently large. It is likely that a universally stabilizing policy is obtained by letting θ grow slowly with $\|x\|$, say,

$$\theta(x) = \theta_0 \log(1 + \|x\|), \qquad x \in \mathbb{R}^\ell_+,$$

with $\theta_0 > 0$ arbitrary. Some of the elegance is then lost since h is no longer convex, so we are considering alternate approaches.

The generalization of Theorem 3.1 to multiple bottlenecks is a significant open problem. This is difficult because we do not have an explicit representation for the relative value function for the relaxation, and we do not know the optimal policy when the effective cost is not monotone.

Suppose that $\hat{h}_*\colon \mathbb{R}^n_+ \to \mathbb{R}$ solves the average-cost optimality equations for the n-dimensional relaxation, and consider the following generalization of (76),

$$h(x) := \hat{h}_*(\widetilde{w}) + \tfrac{1}{2} b[c(\tilde{x}) - \overline{c}(\widetilde{w})]^2, \qquad x \in \mathbb{R}^\ell_+. \tag{94}$$

This will depend upon κ in a parameterized model. The following conjecture is suggested:

> *Conjecture*: Theorem 3.1 can be extended to the case where there are precisely n bottlenecks as $\kappa \to \infty$ based on the h-MaxWeight policy with h given in (94).

If true, then this provides a valuable tool for constructing an effective policy in a complex but centralized network setting.

Analysis of the usual MaxWeight policy based on a workload relaxation of dimension $n \geq 2$ is contained in [36, Chapter 6]. Under general conditions the convex program that defines the effective cost can be solved explicitly. Based on the geometry of the level sets of the effective cost, a hedging technique is introduced to avoid excessive idleness of resources, and hopefully mirror an average-cost optimal policy.

A final topic of current interest is to bridge this work with recent approaches to machine learning. Given a parameterized family of functions $\{h_\alpha : \alpha \in \mathbb{R}^d\}$, we seek the value of α such that the h_α-MaxWeight policy has the best performance in this class. There are a variety of methods to find an optimizer based on simulation [4, 46, 49, 12, 41]. It is hoped that specialized algorithms can be constructed for networks based on the techniques introduced here.

References

[1] M. Andrews. Instability of the proportional fair scheduling algorithm for HDR. *Trans. Wireless Comm.*, 3(5):1422–1426, 2004.

[2] B. Ata and S. Kumar. Heavy traffic analysis of open processing networks with complete resource pooling: Asymptotic optimality of discrete review policies complete resource pooling: asymptotic optimality of discrete review policies. *Ann. Appl. Probab.*, 15(1A):331–391, 2005.

[3] S. L. Bell and R. J. Williams. Dynamic scheduling of a system with two parallel servers in heavy traffic with complete resource pooling: Asymptotic optimality of a continuous review threshold policy. *Ann. Appl. Probab.*, 11:608–649, 2001.

[4] D.P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming.* Atena Scientific, Cambridge, Mass, 1996.

[5] D. Bertsimas, D. Gamarnik, and J. N. Tsitsiklis. Stability conditions for multiclass fluid queueing networks. *IEEE Trans. Automat. Control*, 41(11):1618–1631, 1996.

[6] V. S. Borkar. Convex analytic methods in Markov decision processes. In *Handbook of Markov decision processes*, volume 40 of *Internat. Ser. Oper. Res. Management Sci.*, pages 347–375. Kluwer Acad. Publ., Boston, MA, 2002.

[7] H. Chen and D. D. Yao. *Fundamentals of queueing networks: Performance, asymptotics, and optimization*. Springer-Verlag, New York, 2001. Stochastic Modelling and Applied Probability.

[8] M. Chen, R. Dubrawski, and S. P. Meyn. Management of demand-driven production systems. *IEEE Trans. Automat. Control*, 49(2):686–698, May 2004.

[9] M. Chen, C. Pandit, and S. P. Meyn. In search of sensitivity in network optimization. *Queueing Syst. Theory Appl.*, 44(4):313–363, 2003.

[10] D. P. Connors, G. Feigin, and D. Yao. Scheduling semiconductor lines using a fluid network model. *IEEE Trans. on Robotics and Automation*, 10:88–98, 1994.

[11] I. Ch. Paschalidis D. Bertsimas and J. N. Tsitsiklis. Scheduling of multiclass queueing networks: Bounds on achievable performance. In *Workshop on Hierarchical Control for Real–Time Scheduling of Manufacturing Systems*, Lincoln, New Hampshire, October 16–18, 1992.

[12] D. P. de Farias and B. Van Roy. The linear programming approach to approximate dynamic programming. *Operations Res.*, 51(6):850–865, 2003.

[13] A. Eryilmaz, R. Srikant, and J. R. Perkins. Stable scheduling policies for fading wireless channels. *IEEE/ACM Trans. Netw.*, 13(2):411–424, 2005.

[14] S. N. Ethier and T. G. Kurtz. *Markov Processes : Characterization and Convergence.* John Wiley & Sons, New York, 1986.

[15] G. Fayolle, V. A. Malyshev, and M. V. Men′shikov. *Topics in the constructive theory of countable Markov chains*. Cambridge University Press, Cambridge, 1995.

[16] D. Gamarnik and A. Zeevi. Validity of heavy traffic steady-state approximations in generalized jackson networks. *Adv. Appl. Probab.*, 16(1):56–90, 2006.

[17] L. Georgiadis, W. Szpankowski, and L. Tassiulas. A scheduling policy with maximal stability region for ring networks with spatial reuse. *Queueing Syst. Theory Appl.*, 19(1-2):131–148, 1995.

[18] J. M. Harrison. *Brownian motion and stochastic flow systems*. Robert E. Krieger Publishing Co. Inc., Malabar, FL, 1990. Reprint of the 1985 original.

[19] J. M. Harrison. Heavy traffic analysis of a system with parallel servers: asymptotic optimality of discrete-review policies. *Ann. Appl. Probab.*, 8(3):822–848, 1998.

[20] J. M. Harrison and M. J. López. Heavy traffic resource pooling in parallel-server systems. *Queueing Syst. Theory Appl.*, 33:339–368, 1999.

[21] F.P. Kelly and C.N. Laws. Dynamic routing in open queueing networks: Brownian models, cut constraints and resource pooling. *Queueing Syst. Theory Appl.*, 13:47–86, 1993.

[22] J. F. C. Kingman. The single server queue in heavy traffic. *Proc. Cambridge Philos. Soc.*, 57:902–904, 1961.

[23] J. F. C. Kingman. On queues in heavy traffic. *J. Roy. Statist. Soc. Ser. B*, 24:383–392, 1962.

[24] P. R. Kumar and S. P. Meyn. Duality and linear programs for stability and performance analysis queueing networks and scheduling policies. *IEEE Trans. Automat. Control*, 41(1):4–17, 1996.

[25] P. R. Kumar and T. I. Seidman. Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems. *IEEE Trans. Automat. Control*, AC-35(3):289–298, March 1990.

[26] S. Kumar and P. R. Kumar. Performance bounds for queueing networks and scheduling policies. *IEEE Trans. Automat. Control*, AC-39:1600–1611, August 1994.

[27] H. J. Kushner. *Heavy traffic analysis of controlled queueing and communication networks*. Springer-Verlag, New York, 2001. Stochastic Modelling and Applied Probability.

[28] H. J. Kushner and G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.

[29] C. N. Laws and G. M. Louth. Dynamic scheduling of a four-station queueing network. *Prob. Eng. Inf. Sci.*, 4:131–156, 1990.

[30] N. Laws. *Dynamic routing in queueing networks*. PhD thesis, Cambridge University, Cambridge, UK, 1990.

[31] A. Mandelbaum and A. L. Stolyar. Scheduling flexible servers with convex delay costs: heavy-traffic optimality of the generalized $c\mu$-rule. *Operations Res.*, 52(6):836–855, 2004.

[32] S. P. Meyn. The policy iteration algorithm for average reward Markov decision processes with general state space. *IEEE Trans. Automat. Control*, 42(12):1663–1680, 1997.

[33] S. P. Meyn. Sequencing and routing in multiclass queueing networks. Part I: Feedback regulation. *SIAM J. Control Optim.*, 40(3):741–776, 2001.

[34] S. P. Meyn. Sequencing and routing in multiclass queueing networks. Part II: Workload relaxations. *SIAM J. Control Optim.*, 42(1):178–217, 2003.

[35] S. P. Meyn. Dynamic safety-stocks for asymptotic optimality in stochastic networks. *Queueing Syst. Theory Appl.*, 50:255–297, 2005.

[36] S. P. Meyn. Control techniques for complex networks. To appear, Cambridge University Press, 2007.

[37] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, second edition, 1993. online: `http://black.csl.uiuc.edu/~meyn/pages/book.html`.

[38] S. P. Meyn and R. L. Tweedie. Computable bounds for convergence rates of Markov chains. *Ann. Appl. Probab.*, 4:981–1011, 1994.

[39] J. R. Morrison and P. R. Kumar. New linear program performance bounds for queueing networks. *J. Optim. Theory Appl.*, 100(3):575–597, 1999.

[40] J. Ou and L. M. Wein. Performance bounds for scheduling queueing networks. *Ann. Appl. Probab.*, 2:460–480, 1992.

[41] B. Van Roy. Neuro-dynamic programming: Overview and recent trends. In E. Feinberg and A. Shwartz, editors, *Markov Decision Processes: Models, Methods, Directions, and Open Problems*, pages 43–82. Kluwer, Holland, 2001.

[42] A. N. Rybko and A. L. Stolyar. On the ergodicity of random processes that describe the functioning of open queueing networks. *Problemy Peredachi Informatsii*, 28(3):3–26, 1992.

[43] S. Shakkottai, R. Srikant, and A. L. Stolyar. Pathwise optimality of the exponential scheduling rule for wireless channels. *Adv. Appl. Probab.*, 36(4):1021–1045, 2004.

[44] S. Shakkottai and A. L. Stolyar. Scheduling for multiple flows sharing a time-varying channel: the exponential rule. In *Analytic methods in applied probability*, volume 207 of *Amer. Math. Soc. Transl. Ser. 2*, pages 185–201. Amer. Math. Soc., Providence, RI, 2002.

[45] A. L. Stolyar. Maxweight scheduling in a generalized switch: state space collapse and workload minimization in heavy traffic. *Adv. Appl. Probab.*, 14(1):1–53, 2004.

[46] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press (and on-line, http://www.cs.ualberta.ca/˜sutton/book/the-book.html), 1998.

[47] L. Tassiulas. Adaptive back-pressure congestion control based on local information. *IEEE Trans. Automat. Control*, 40(2):236–250, 1995.

[48] L. Tassiulas and A. Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Trans. Automat. Control*, 37(12):1936–1948, 1992.

[49] J. N. Tsitsiklis and B. Van Roy. An analysis of temporal-difference learning with function approximation. *IEEE Trans. Automat. Control*, 42(5):674–690, 1997.

[50] J. A. Van Mieghem. Dynamic scheduling with convex delay costs: the generalized c-μ rule. *Ann. Appl. Probab.*, 5(3):809–833, 1995.