Stability and Asymptotic Optimality of Generalized MaxWeight Policies*

Sean Meyn†

May 22, 2008

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 share many of the attractive properties of the MaxWeight policy:

(i) Arrival rate data is not required in the policy.

(ii) Under a variety of general conditions, the policy is stabilizing when h is a perturbation of a monotone linear function, a monotone quadratic, 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 general Markovian network model. Asymptotic optimality is established for a general Markovian scheduling model with a single bottleneck, and homogeneous servers.

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

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

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. A well-known example is the MaxWeight policy of Tassiulas and Ephremides [52]. This policy can be interpreted as a myopic policy for the associated fluid model with respect to a quadratic function,

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

with $D > 0$ a diagonal matrix. Stability theory for this and similar classes of policies has been extended in multiple directions over the past fifteen years [20, 51, 47, 17, 13], and in particular these policies are known to be approximately optimal in heavy traffic under certain conditions on the network — see [54, 48, 33], and the recent comprehensive results by Dai and Lin [14].

These results are fragile: Diagonal quadratics are one of a very few function classes for which the myopic policy is known to be stabilizing for general classes of network models. In contrast, stability of the fluid model under a myopic policy is virtually universal [10, 7, 35].

It is important to find broader classes of stabilizing policies for complex networks. It is known that the MaxWeight policy can perform poorly since it makes use of so little information [49].

To explain the gap between the stochastic and deterministic models we consider some simple, well-known examples. The stochastic model favored in this paper is the controlled random walk (CRW) model in which the queue length process $\boldsymbol{Q}$ evolves on $\mathbb{Z}_+^{\ell}$ in discrete time 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 allocation sequence $\boldsymbol{U}$ is subject to both integral and linear constraints of the form $U(t) \in \mathsf{U}_\diamond$, $t \geq 0$, where

$$\mathsf{U}_\diamond := \{u \in \{0, 1\}^{\ell_u} : Cu \leq \mathbf{1}\}. \tag{3}$$

The $\ell_m \times \ell_u$ matrix C is called the constituency matrix: Each of the rows of C corresponds to a 'resource' in the network. Its entries are assumed to be binary.

The CRW model is a generalization of the network model of Lippman obtained via uniformization [31]. Versions of this model appear throughout the communications and operations research literature, and in particular appear in the paper [52] that introduced the MaxWeight policy.

The fluid model $\boldsymbol{q} = \{q(t) : t \geq 0\}$ satisfies the linear equations,

$$q(t) = x + Bz(t) + \alpha t, \qquad t \geq 0, \tag{4}$$

where $x \in \mathbb{R}_+^{\ell}$ is the initial condition, 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 constraints on

z are analogous to those on U. We let U denote the convex hull $\mathsf{U} := \operatorname{conv}(\mathsf{U}_\diamond)$, and assume that for each $0 \leq t_0 < t_1$,

$$\frac{z(t_1) - z(t_0)}{t_1 - t_0} \in \mathsf{U}.$$

The fluid model (4) is also expressed as the ODE model,

$$\tfrac{d^+}{dt} q(t) = B\zeta(t) + \alpha, \qquad t \geq 0, \tag{5}$$

where $\zeta(t) \in \mathsf{U}$ denotes the allocation rate vector at time t, and the "+" denotes right derivative.

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,

$$\phi^{\mathrm{F}}(x) = \arg\min_{\zeta \in \mathsf{U}(x)} \langle \nabla h(x), B\zeta + \alpha \rangle, \tag{6}$$

$$\phi^{\mathrm{D}}(x) = \arg\min_{u \in \mathsf{U}_\diamond(x)} \mathsf{E}[h(Q(t+1)) - h(Q(t)) \mid Q(t) = x, U(t) = u], \tag{7}$$

where the "(x)" is used to capture boundary constraints,

$$\begin{aligned}
\mathsf{U}_\diamond(x) &= \{u \in \mathsf{U} : (B(t)u)_i \geq 0 \; a.s. \text{ when } x_i = 0\}, \\
\mathsf{U}(x) &= \{u \in \mathsf{U} : v_i := (Bu + \alpha)_i \geq 0 \text{ when } x_i = 0\},
\end{aligned} \tag{8}$$

The MaxWeight policy coincides with the *h*-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*,

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

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 [35, Proposition 11] for a smooth norm on $\mathbb{R}^\ell$.) The proof is based on establishing that the 'drift' defined by,

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

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

The *h*-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.
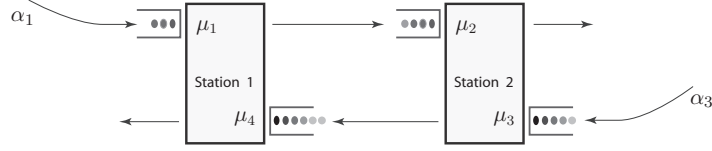
3

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 [27, 46]. A typical choice for h is a cost function c, and a typical cost function in network applications is the ℓ_1 norm, $c(x) = |x| = \sum x_i$. With $h = c$, 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 h-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 [46].

Work-stoppage under a myopic policy The h-myopic policy 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 h-myopic policy for the fluid model with $h = c$ 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 h-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 $x \in \mathbb{Z}_+^2$,

$$\begin{aligned} \phi^{\mathrm{MW}}(x) = \phi^{\mathrm{D}}(x) &= \arg\min_{u\in\mathsf{U}_\diamond(x)} \mathsf{E}[c(Q(t+1)) \mid Q(t) = x, \ U(t) = u] \\ &= \arg\min_{u\in\mathsf{U}_\diamond(x)} \big(c_1(\alpha_1 - \mu_1 u_1) + c_2(\mu_1 u_1 - \mu_2 u_2)\big). \end{aligned}$$

4

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

$$\phi_1^{\mathrm{MW}}(x) = \arg\min_{u \in \mathsf{U}_\diamond(x)}\big((c_2 - c_1)\mu_1 u_1\big).$$

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,

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

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,

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

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 $\langle \nabla h_0, v \rangle \leq -c$.

For example, if $\| \cdot \|_h$ is a any norm on $\mathbb{R}^\ell$ that is monotone and C^1 on $\mathbb{R}_+^\ell$, then (15) holds with $h_0(x) = \frac{1}{2}\|x\|_h^2$ and $c(\cdot) = \varepsilon_0\| \cdot \|_h$ for some $\varepsilon_0 > 0$. Another solution to the dynamic programing inequality is the quadratic (1) in which D is not necessarily diagonal, but satisfies $D_{ij} \geq 0$ and $D_{ii} > 0$ for each i, j. In some cases a fluid value function is piecewise quadratic, C^1, and satisfies (15) with equality. An example is contained in Section 2.2.2.

A perturbation of h_0 is obtained through a change of variables: For fixed $\theta \geq 1$ 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 $\tilde{x} := (\tilde{x}_1, \ldots, \tilde{x}_\ell)^{\mathrm{T}} \in \mathbb{R}_+^\ell$. 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 i.i.d. process $(\boldsymbol{A}, \boldsymbol{B})$ has integer entries, and a finite second moment.*

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

$$B_{ij}(t) \geq 0 \qquad a.s. \quad i \neq i_j. \tag{18}$$

(iii) *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 $\overline{\eta}_h < \infty$ such that for any $\theta \geq \theta_0$, the following bound holds under the h-MaxWeight policy:

$$\mathsf{E}[h(Q(t+1)) - h(Q(t)) \mid Q(t) = x] \leq -\tfrac{1}{2}c(x) + \tfrac{1}{2}\overline{\eta}_h. \tag{19}$$

Consequently,

$$n^{-1}\mathsf{E}\Big[\sum_{t=0}^{n-1} c(Q(t)) \mid Q(t) = x\Big] \leq 2n^{-1}h(x) + \overline{\eta}_h, \qquad n \geq 1, x \in \mathbb{Z}_+^\ell. \tag{20}$$

Proof. This is based on results obtained in Section 2.2.2: Combining the bounds obtained in Lemma 2.1 and Lemma 2.2 gives under the h-MaxWeight policy, for each $x \in \mathbb{Z}_+^\ell$,

$$\mathsf{E}[h(Q(t+1)) - h(Q(t)) \mid Q(t) = x] \leq -c(x) + k_{2.1}\log(1 + \|x\|) + k_{2.2}(1 + \theta^{-1}\|x\|),$$

where the constants are independent of θ. Choosing $\theta > 2k_{2.2}$, we obtain the bound (19).

Equation (20) then follows from the Comparison Theorem 2.1. $\qquad\square$

Assumption (ii) implies that the matrix $-B(t)$ is *Leontief* with probability one for each t, and that its expectation $-B = -\mathsf{E}[B(t)]$ is also Leontief. Bramson and Williams [6] call a network *unitary* if Assumption (ii) holds, and in addition the rows of C are orthogonal (interpreted as the absence of "simultaneous resource possession"). Relaxations of Assumption (ii) that imply stability of the MaxWeight policy are contained in [13, 14] (called *maximum pressure policies* in these papers).

For networks that are unitary the h-MaxWeight policy has a simple representation in terms of the *generalized Klimov indices,*

$$\Theta_j(x) := -\sum_i B_{ij}\frac{\partial}{\partial x_i}h(x), \qquad x \in \mathbb{Z}_+^\ell, \; j \in \{1, \ldots, \ell_u\}. \tag{21}$$

For a unitary model, for any j we denote by $s(j) \in \{1, \ldots, \ell_m\}$ the unique value of s satisfying $C_{s,j} = 1$.

Proposition 1.1. *Suppose that the CRW model is unitary. Suppose moreover that h is C^1, monotone, and satisfies the boundary conditions (14). Then, the h-MaxWeight policy can be described as follows: For each $s \in \{1, \ldots, \ell_m\}$ and $x \in \mathbb{Z}_+^\ell$, denote $\Theta_s^*(x) := \max\{\Theta_j(x) : s(j) = s\}$. if $\Theta_s^* < 0$ then $U_j(t) = 0$ whenever $s(j) = s$. Otherwise, priority is giving to buffers that achieve the maximum,*

$$\sum \{U_j(t) : s(j) = s, \ \Theta_j(x) = \Theta_s^*\} = 1.$$

Proof. For a unitary model the optimization (9) decouples into ℓ_m separate optimization problems, each with a single linear constraint obtained from the respective row of C. The proof is completed on noting that $-\Theta_j$ is the coefficient of u_j in the objective function of (9). □

A drawback to Theorem 1.1 is that stability holds only for $\theta > 0$ sufficiently large. Section 2.3 considers the alternative change of variables $\tilde{x}_i := x_i \log(1 + x_i/\theta)$, $i = 1, \ldots, \ell$. Theorem 2.5 shows that the resulting h-MaxWeight policy is stabilizing, for any fixed $\theta > 0$, in the sense that a version of (19) holds.

The inequality (19) is a *Lyapunov drift condition* of the form developed in [41, 18], and also similar to the bounds used in [11, 4, 28, 26, 43] 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 [41, 42, 40, 26, 35, 19]. Proposition 2.5 below contains sufficient conditions for geometric ergodicity for a particular version of the h-MaxWeight policy.

In Section 3 we move to an asymptotic, heavy-traffic setting to obtain finer performance bounds. Dai and Lin's recent paper [14] contains a comprehensive survey on the theory of networks in heavy traffic. While the results in Section 3 use language and some results from the heavy traffic literature, the goals and conclusions are very different from those of [14] or any other papers from this literature.

A heavy traffic analysis is based on the construction of a one-dimensional parameterized family of networks with increasing load. Let $\kappa > 0$ denote the parameter, and assume that the load increases to one as $\kappa \to \infty$. Letting $Q^\kappa(t; x)$ denote the queue-length process for the κth network at time t with initial condition x, a "Central Limit"-scaling is applied,

$$Q^{\kappa,\kappa}(t; x) := \frac{1}{\kappa} Q^\kappa(\kappa^2 t; \kappa x) \tag{22}$$

This is defined for all $t \in \mathbb{R}_+$ via linear interpolation.

In virtually all of the asymptotic results contained in the literature it is assumed that there is a single bottleneck in heavy traffic or, more generally, *complete resource pooling* [2, 48, 33, 1, 14]. Let $\xi \in \mathbb{R}^\ell$ denote the corresponding workload vector, and assume its entries are non-negative. Then the workload process $W^\kappa(t; x) = \xi^\mathsf{T} Q^\kappa(t; x)$ evolves on $\mathbb{R}_+$, and can be compared to a minimal workload process $\widehat{W}^\kappa(t; x)$. Section 3 restricts to a simplified setting in which a realization of the minimal process evolves as a simple queue,

$$\widehat{W}^\kappa(t+1) = \widehat{W}^\kappa(t) - S_1(t+1)\mathbb{1}\{\widehat{W}^\kappa(t) \geq 1\} + L_1(t+1), \qquad t \geq 0, \tag{23}$$

where $(\boldsymbol{S}_1, \boldsymbol{L}_1)$ is i.i.d. on $\mathbb{Z}_+^2$, and $\boldsymbol{S}_1$ is Bernoulli (see (83)). The load is given by $\rho_\bullet = \mathsf{E}[L_1(t)]/\mathsf{E}[S_1(t)]$.

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

$$\begin{aligned} \overline{c}(w) \;=\; \min \quad & c(x) \\ \text{s.t.} \quad & \xi^{\mathrm{T}} x = w, \; x \in \mathbb{R}_+^\ell. \end{aligned} \tag{24}$$

A sequence of policies is declared to be *heavy-traffic asymptotically optimal* (HTAO) if the following properties are verified:

State space collapse For each t, the scaled queue length process has asymptotically minimal cost, subject to its workload, in the sense that

$$\lim_{\kappa \to \infty} \big(c(Q^{\kappa,\kappa}(t;x)) - \overline{c}(W^{\kappa,\kappa}(t;x)) \big) = 0,$$

where the convergence is in probability. This is called state space collapse because typically it implies that the queue-length processes converge to a one-dimensional subspace,

$$\lim_{\kappa \to \infty} \big\| Q^{\kappa,\kappa}(t;x) - \mathcal{X}^*\big(Q^{\kappa,\kappa}(t;x))\big) \big\| = 0, \tag{25}$$

where $\mathcal{X}^*$ denotes the projection. The vector $\mathcal{X}^*(Q)$ is known as the *effective state* corresponding to Q (see (84)).

Asymptotic minimality The scaled workload process $W^{\kappa,\kappa}(t;x)$ and the scaled minimal workload process $\widehat{W}^{\kappa,\kappa}(t;x)$, defined as in (22), each converge in distribution to a reflected Brownian motion (RBM) with common drift and covariance.

A uniform version of HTAO is formulated in [36]. The paper considers multiclass networks with multiple bottlenecks and renewal inputs. Under the assumption that the effective cost is a monotone function of w, among other assumptions, the following uniform asymptotic bounds are obtained for a proposed policy: For any other policy, letting $\boldsymbol{Q}^{\kappa\prime}$ denote the resulting state process,

$$\frac{1}{T} \int_0^T c(Q^\kappa(t;x)) \, dt \leq \frac{1}{T} \int_0^T c(Q^{\kappa\prime}(t;x)) \, dt + O\big(\log\big((1-\rho_\bullet)^{-1}\big)\big), \quad 0 \leq T \leq \frac{1}{(1-\rho_\bullet)^3}. \tag{26}$$

The policies considered in [36] are based on those of [32], which are generalizations of the policy introduced in [2] for a particular example.

HTAO for MaxWeight and certain generalizations is established in the aforementioned papers [54, 48, 33, 14]. However, state space collapse for the MaxWeight policy is obtained with respect to an implicitly defined cost function [48]. In [14] an approximation is obtained: For a given linear cost function, the projection $\mathcal{X}^*$ is of the form,

$$\mathcal{X}^*(x) = (\xi^{\mathrm{T}} x) \frac{c_{i^*}}{\xi_{i^*}} \mathbf{1}^{i^*}$$

where $i^* \in \arg\min\{c_i/\xi_i\}$, and $\mathbf{1}^i$ denotes the ith basis element in $\mathbb{R}^\ell$. For each $\varepsilon > 0$ the authors construct a version of the MaxWeight policy satisfying,

$$\lim_{\kappa \to \infty} \mathsf{P}\big\{\big\|Q^{\kappa,\kappa}(t;x) - \mathcal{X}^*\big(Q^{\kappa,\kappa}(t;x))\big)\big\| > \varepsilon\big\|Q^{\kappa,\kappa}(t;x)\big\|\big\} = 0. \tag{27}$$

The present paper is concerned with steady-state performance. The average cost is denoted,

$$\eta = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathsf{E}[c(Q(t;x))] \tag{28}$$

where the limit is independent of x under the assumptions imposed in the main results. Note that HTAO as formulated above does not imply that the performance measured by average cost is approximately optimal, or even that η is finite. Heavy-traffic asymptotically optimality *suggests* a bound of the form,

$$\eta \le \widehat{\eta}^* + o(\widehat{\eta}^*) \tag{29}$$

where $\widehat{\eta}^* = \mathsf{E}[\overline{c}(\widehat{W}^\kappa(t))]$, $\overline{c}$ denotes the effective cost, and the expectation is in steady-state. This heuristic has been established rigorously only in special cases, based on the assumption of weak convergence of the scaled workload processes.

Suppose that the scaled workload processes converge in distribution to a reflected Brownian motion,

$$\widehat{W}^{\kappa,\kappa}(t;x) \xrightarrow{\mathrm{w}} \widehat{W}^\infty(t;x), \qquad \kappa \to \infty.$$

If moreover the scaled steady-state means are convergent, $\mathsf{E}[\overline{c}(\widehat{W}^{\kappa,\kappa}(t;x))] \to \widehat{\eta}^\infty = \mathsf{E}[\overline{c}(\widehat{W}^\infty(t))]$, $\kappa \to \infty$, then we can then reinterpret (29) as the limit

$$\lim_{\kappa \to \infty} \mathsf{E}[c(Q^{\kappa,\kappa}(t;x))] = \widehat{\eta}^\infty, \tag{30}$$

where the expectations are all in steady-state. The approximation (30) has been established for the single queue in the pioneering work of Kingman [24, 25], and for generalized Jackson networks [19, Corollary 2, p. 73]. Stolyar conjectures in [48] that the invariant distributions for $Q^{\kappa,\kappa}$ will convergence weakly under the MaxWeight policy, provided there is complete resource pooling. This together with uniform integrability would imply the limit (30).

The limit (27) also suggests an asymptotic bound, of the form

$$\eta \le \widehat{\eta}^* + O(\varepsilon\widehat{\eta}^*)$$

but no such result is established in [14], or elsewhere.

Section 3 treats heavy-traffic asymptotic optimality in an average-cost sense. The main result establishes *logarithmic regret*: for some fixed constant $k_0 < \infty$, independent of load,

$$\widehat{\eta}^* \le \eta \le \widehat{\eta}^* + k_0 \log(\widehat{\eta}^*) \tag{31}$$

This is a significant refinement of (29) since $\widehat{\eta}^* \to \infty$ as the network load approaches unity. Heavy-traffic asymptotically optimality of the form (31) was obtained for the first time in [37] in several examples, based on Lyapunov techniques similar to those used in this paper. 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

$$\text{Serve buffer } 1 \text{ if } x_2 \leq \beta \log(1 + x_1/\beta), \qquad (32)$$

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

The approach used in Section 3 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 with logarithmic regret.

The development is greatly simplified by imposing further structure on the model. The following assumptions are imposed in Theorem 3.1, the main result of Section 3 that establishes logarithmic regret.

(HTAO 1) The network is described by a scheduling model with deterministic routing: for each $i \in \{1, \ldots, \ell\}$, after processing at buffer i a customer either enters some buffer $i_+ \in \{1, \ldots, \ell\}$, or exits the system. The *routing matrix R* is the $\ell \times \ell$ matrix defined for $i, j \in \{1, \ldots, \ell\}$ as $R_{ij} = \mathbb{1}_{j=i_+}$. 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,

$$Q(t+1) = Q(t) + \sum_{i=1}^{\ell} M_i(t+1)[-1^i + 1^{i_+}]U_i(t) + A(t+1), \qquad Q(0) = x, \quad (33)$$

where $\boldsymbol{M}_i$ is a Bernoulli sequence for each i. The vector 1^{i_+} is taken to be zero if customers exit the system following service at buffer i.

(HTAO 2) The network is of a generalized 'Kelly type' in which service statistics are determined by the station, not the buffer. For each $s \in \{1, \ldots, \ell_m\}$, Station s is said to be *homogeneous* if the random variables $\{M_j(t) : s(j) = s\}$ are all identical. It is assumed that the network is homogeneous, meaning that each station satisfies this condition.

(HTAO 3) The cost function is linear.

(HTAO 4) There is a single bottleneck in heavy traffic, and the convex program (24) that defines the effective cost is assumed to possess a unique maximizer for each $w \in \mathbb{R}_+$.

We believe that many of these assumptions are non-essential. Potential extensions are surveyed in Section 4.

The CRW scheduling model (33) is of the form (2) with,

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

The Leontief condition (18) follows from the assumptions on $\boldsymbol{M}$ and R, where $i_j = j$ for each $j \in \{1, \ldots, \ell\}$. The vector load $\rho \in \mathbb{R}_+^{\ell_m}$ is given by,

$$\rho = CM^{-1}[I - R^{\mathrm{T}}]^{-1}\alpha, \qquad (34)$$

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

Homogeneity is used to construct a one-dimensional workload relaxation satisfying the recursion (23). A one-dimensional workload process $\boldsymbol{W}$ corresponding to the most heavily loaded station is compared to its (minimal) relaxation $\widehat{\boldsymbol{W}}^*$. For each $t \ge 0$ the lower bound holds with probability one,

$$W(t) \ge \widehat{W}^*(t), \qquad t \ge 0, \tag{35}$$

under *any* policy for $\boldsymbol{Q}$. This simplifies a proof of logarithmic regret since a lower bound on performance is explicit, $\widehat{\eta}^* = \mathsf{E}[\overline{c}(\widehat{W}^*(t))]$.

A relaxation of the form (23) was introduced in the thesis of N. Laws [30] to obtain performance bounds (see also [29, 44]). Relaxations of this form and multi-dimensional extensions for the purposes of control synthesis and performance approximation are the subject of [9, 37, 23, 39].

Two significant contributions in the present paper are worth highlighting,

(i) In each example considered in [37] the policy was explicitly constructed based on switching curves of the form (32). 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 (31) 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 (31) 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 $j \in \{1, \ldots, \ell_u\}$. 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 [21, 22]; Optimal policies for the fluid and CRW models are described in [35, p. 760] for a simple example.

(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 $i \ne i_j$.

11

(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

$$U(t) = \phi(Q(t)), \qquad t \geq 0. \tag{36}$$

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 $\mathcal{D}h$ for the stochastic model, based on the corresponding drift $\langle B\phi^{\mathrm{MW}}(x) + \alpha, \nabla h\,(x)\rangle$ for the fluid model. 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 $t \geq 0$,

$$
\begin{aligned}
h(Q(t+1)) - h(Q(t)) &= \langle \nabla h\,(\bar{Q}), \Delta(t+1)\rangle \\
&= \langle \nabla h\,(Q(t)), \Delta(t+1)\rangle \\
&\quad + \langle \nabla h\,(\bar{Q}) - \nabla h\,(Q(t)), \Delta(t+1)\rangle
\end{aligned}
\tag{37}
$$

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,

$$\mathcal{D}h\,(x) = \langle \nabla h\,(x), v\rangle + b_h, \tag{38}$$

where

$$v(x) = \mathsf{E}[\Delta(t+1) \mid Q(t) = x], \qquad b_h(x) = \mathsf{E}[\langle \nabla h\,(\bar{Q}) - \nabla h\,(Q(t)), \Delta(t+1)\rangle \mid Q(t) = x]. \tag{39}$$

To deduce stability based on (38) 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

$$\mathsf{V} := \{v = B\zeta + \alpha : \zeta \in \mathsf{U}\}. \tag{40}$$

12

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

$$\text{The origin is an interior point of } \mathsf{V}. \tag{41}$$

If (41) 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 $0 \le t \le |x|/\varepsilon$. For a given policy for the fluid model, such as (7), we consider the value function,

$$J(x) = \int_0^\infty c(q(t;x))\, dt. \tag{42}$$

We let J^* denote the minimum of (42) 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 $J = J^*$. We thereby obtain another class of functions that satisfy the dynamic programing inequality (15). The fluid value function is typically C^1 in workload models [38].

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 [41]. 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

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

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

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

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,

$$P_{\phi^{\mathrm{D}}} V(x) = k \arg\min_{u \in \mathsf{U}_\diamond(x)} \mathsf{E}[h(Q(t+1)) \mid Q(t) = x, U(t) = u]$$

$$\le k\mathsf{E}[h(Q(t+1)) \mid Q(t) = x, U(t) = \phi^{\mathrm{MW}}(x)]$$

$$= P_{\phi^{\mathrm{MW}}} V(x) \le V(x) - f(x) + b\mathbb{1}_S(x).$$

$\square$

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*,

$$\mathcal{D}h \le -c + \overline{\eta}. \tag{43}$$

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 (43) holds. Theorem 2.1 is also the most common approach to obtaining bounds on expectations involving stopping times. For a proof see [41].

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

$$\mathcal{D}V \le -f + g. \tag{44}$$

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

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

$\square$

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 [41]. A sufficient condition for 0-irreducibility is given in Proposition 3.1.

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 moreover that the controlled network is 0-irreducible: For each $x \in \mathbb{Z}_+^\ell$,*

$$\sum_{t=0}^\infty \mathsf{P}\{Q(t) = 0 \mid Q(0) = x\} > 0, \tag{45}$$

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

(i) *$\boldsymbol{Q}$ is an aperiodic Markov chain with unique invariant measure π. The average cost defined in (28) is finite, is independent of initial condition, and coincides with the mean with respect to π:*

$$\eta = \pi(c) := \sum \pi(x)c(x)$$

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

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

14

(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,

$$Q(t+1) = Q(t) - S(t+1)U(t) + A(t+1), \qquad t \in \mathbb{Z}_+, \tag{46}$$

with given initial condition $Q(0) = x \in \mathbb{Z}_+$. A solution to Poisson's inequality (43) is obtained with $c(x) \equiv x$ and $V = J^*$, where the fluid value function is quadratic,

$$J^*(x) = \tfrac{1}{2} \frac{x^2}{\mu - \alpha}. \tag{47}$$

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 $x \in \mathbb{R}_+$),

$$\mathcal{D}h(x) = -x + \eta, \qquad x \in \mathbb{Z}_+. \tag{48}$$

The formula (50) 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 $V = J^*$ [37, 39].

Theorem 2.3. *Consider the CRW queueing model (46) satisfying $\rho = \alpha/\mu < 1$, and define*

$$m^2 = \mathsf{E}[(S(1) - A(1))^2], \quad m_A^2 = \mathsf{E}[A(1)^2], \quad \sigma^2 = \rho m^2 + (1 - \rho)m_A^2. \tag{49}$$

Then,

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

$$\eta := \mathsf{E}_\pi[Q(0)] = \tfrac{1}{2} \frac{\sigma^2}{\mu - \alpha} \tag{50}$$

(ii) *A solution to Poisson's equation (48) is the quadratic,*

$$h(x) = J^*(x) + \tfrac{1}{2}\mu^{-1}\Big(\frac{m^2 - m_A^2}{\mu - \alpha}\Big)x, \qquad x \in \mathbb{Z}_+. \tag{51}$$

□

The MaxWeight policy is defined by (9) with h a quadratic,

$$\phi^{\mathrm{MW}}(x) = \arg\min_{u \in \mathsf{U}_\diamond(x)} \langle Dx, Bu + \alpha \rangle, \qquad x \in \mathbb{Z}_+^\ell, \tag{52}$$

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. *Suppose that Assumptions (i) and (ii) of Theorem 1.1 hold. Then, for each $x \in \mathbb{Z}_+^\ell$, the allocation $\phi^{\mathrm{MW}}(x)$ defined by the MaxWeight policy can be expressed as a solution to the linear program,*

$$\phi^{\mathrm{MW}}(x) = \quad \arg\max \quad x^{\mathrm{T}} D(I - R^{\mathrm{T}})Mu$$
$$\text{s.t.} \quad u \in \mathsf{U}.$$
(53)

$\square$

Proposition 2.2 easily leads to a proof of stability of the MaxWeight policy. Theorem 2.4 is essentially contained in earlier work [52, 13]. We present the short proof since the same ideas are used to prove Theorem 1.1 and generalizations that follow.

Theorem 2.4. *Suppose that $\rho_\bullet < 1$, and that Assumptions (i) and (ii) of Theorem 1.1 hold. Then, for any diagonal matrix $D > 0$, the network controlled under the MaxWeight policy has a solution to Poisson's inequality (43) with $V = h$ the quadratic defined in (1), and $c(x) \equiv \varepsilon_0 |x|$ for some $\varepsilon_0 > 0$.*

Proof. Since (41) 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. By definition there exists $u \in \mathsf{U}$ such that $Bu + \alpha = v$, 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 $\varepsilon_0 = \varepsilon(\min_i D_{ii})$. We thus arrive at a version of the Poisson inequality,

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

with

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

$\square$

2.2   Perturbed functions

We now analyze the drift $\mathcal{D}h$ represented in (38) 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,*

$$\nabla h(x) = [I - M_\theta] \nabla h_0(\tilde{x}),$$
(54)

where,

$$M_\theta = M_\theta(x) = \mathrm{diag}(e^{-x_i/\theta}), \qquad x \in \mathbb{R}_+^\ell.$$
(55)

16

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

$$\nabla^2 h\left(x\right) = [I - M_\theta]\nabla^2 h_0\left(\tilde{x}\right)[I - M_\theta] + \theta^{-1}\mathrm{diag}\left(M_\theta \nabla h_0\left(\tilde{x}\right)\right). \tag{56}$$

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.

Proposition 2.4. *Suppose that Assumptions (i) and (ii) 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,*

$$\phi^{\mathrm{MW}}(x) = \quad \arg\min \quad \langle Bu, \nabla h\left(x\right)\rangle$$
$$\mathrm{s.t.} \quad u \in \mathsf{U}. \tag{57}$$

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

Recall the definition of the generalized Klimov indices in (21), and the interpretation that $-\Theta_j$ is the coefficient of u_j in the objective function of (9). Monotonicity of h implies that each partial derivative of h is non-negative. This assumption combined with (18) implies that

$$\Theta_j(x) \leq 0 \text{ whenever } x_{i_j} = 0. \tag{58}$$

It then follows that the optimizer u^* of the linear program (57) 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 (57).

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 contained in $\mathsf{U}_\diamond$ because of the definition $\mathsf{U} := \mathrm{conv}\left(\mathsf{U}_\diamond\right)$. □

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,

$$h(x) = \sum_{i=1}^\ell c_i \tilde{x}_i, \qquad x \in \mathbb{R}_+^\ell. \tag{59}$$

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

$$\nabla h\left(x\right) = [I - M_\theta]c, \qquad \nabla^2 h\left(x\right) = \theta^{-1}\mathrm{diag}\left(M_\theta c\right). \tag{60}$$

Hence the function h is monotone and strictly convex.

17

We show in Proposition 2.5 that the h-MaxWeight policy is stabilizing provided $\theta \geq 1$ is suitably large. Although the dynamic programming inequality (15) fails for the linear function h_0, it does hold for the function $k_0 h_0^2$ where k_0 is a sufficiently large constant, and $c(x) = c^{\mathrm{T}} x$. Hence Proposition 2.5 (ii) could be deduced from Theorem 1.1.

Proposition 2.5. *Suppose that Assumptions (i) and (ii) of Theorem 1.1 hold, along with the stabilizability condition (41). Then, there exists $\theta_0 > 0$ such that the following hold under the h-MaxWeight policy with h_0 linear, provided $\theta \geq \theta_0$.*

(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,*

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

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,*

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

with $V = \exp(\varepsilon_e h)$. Hence $\boldsymbol{Q}$ is geometrically ergodic provided (45) holds [41].

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

$$\begin{aligned} h(Q(t+1)) - h(Q(t)) &= \langle \nabla h\left(Q(t)\right), \Delta(t+1) \rangle \\ &\quad + \tfrac{1}{2} \Delta(t+1)^{\mathrm{T}} \big[\nabla^2 h\left(\bar{Q}\right) \big] \big(\Delta(t+1)\big), \end{aligned} \tag{61}$$

where again $\bar{Q} \in \mathbb{R}_+^\ell$ lies on the line connecting $Q(t+1)$ and $Q(t)$. This implies the identity (38) with b_h redefined as,

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

The expression for the second derivative in (60) then gives,

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

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 $\langle \nabla h\left(x\right), v \rangle$ under the h-MaxWeight policy. The expression for the first derivative in (60) implies the bound,

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

18

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 (57). This is justified by Proposition 2.4. The stabilizability condition (41) implies that there exists $\varepsilon > 0$ such that the vector with components $v_i = -\varepsilon$, $1 \le i \le \ell$, 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,

$$\mathcal{D}h\,(x) \le -\tfrac{1}{2}\varepsilon\underline{c} + \theta^{-1}b_\Delta, \qquad \text{if } |x| \ge \ell\theta. \tag{62}$$

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 (43) with this V, which implies that (V3) also holds.

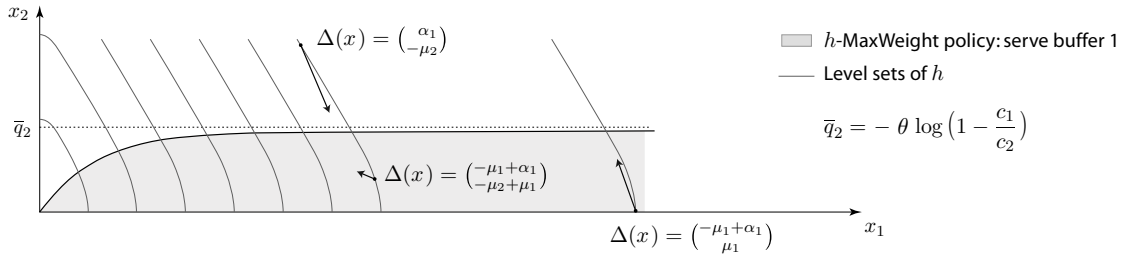Part (iii) follows from (i) combined with [41, Theorem 16.3.1] (see also [35, Theorem 4].) □

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 $r = 1, 2, \ldots$.

In the next example we find that the h-MaxWeight policy considered in Proposition 2.5 (i) mirrors the discounted-cost optimal policy.

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 (59). Figure 3 shows a plot of the sublevel sets of this function.

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

$$\langle Bu + \alpha, \nabla h\,(x) \rangle = -\mu_1 u_1 c_1 (1 - e^{-x_1/\theta}) + (\mu_1 u_1 c_1 - \mu_2 u_2 c_2)(1 - e^{-x_2/\theta}) + \langle \alpha, \nabla h\,(x) \rangle.$$

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

$$\phi_1^{\mathrm{MW}}(x) = \mathbb{1}\{-c_1(1 - e^{-x_1/\theta}) + c_2(1 - e^{-x_2/\theta}) \leq 0\}, \qquad x_1 \geq 1.$$

For small values of x_1 a first order Taylor series gives the approximation $\phi_1^{\mathrm{MW}}(x) \approx \mathbb{1}\{x_2 \leq (c_1/c_2)x_1\}$. For large x_1 there are three cases to consider depending on the relative sizes of c_1 and c_2. If $c_1 = c_2$ then the approximation is equality, $\phi_1^{\mathrm{MW}}(x) = \mathbb{1}\{x_2 \leq x_1\}$ for all x. If $c_1 > c_2$ then Station 1 does not idle for large x_1, exactly as in the c-myopic policy. The h-MaxWeight policy is most interesting when $c_2 > c_1$. In this case, for $x_1 \gg \theta$ the policy can be approximated by a threshold policy, $\phi_1^{\mathrm{MW}}(x) \approx \mathbb{1}\{x_2 \leq \overline{q}_2\}$, where the threshold $\overline{q}_2$ is the solution to the equation $c_2(1 - e^{-\overline{q}_2/\theta}) = c_1$. That is,

$$\overline{q}_2 = \theta \left| \log\left(1 - \frac{c_1}{c_2}\right) \right|. \tag{63}$$

Figure 3 illustrates ϕ^{MW} when $c_1 = 1$, $c_2 = 3$, and $\theta = 10$. The asymptote (63) 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,

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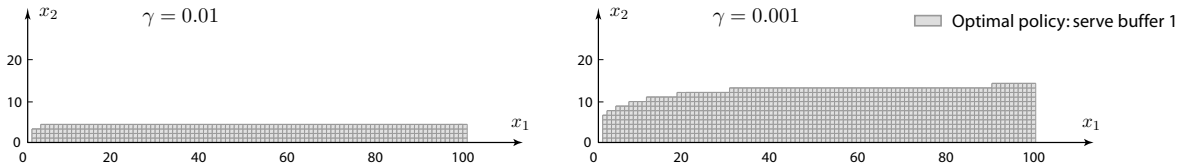

Figure 4: Discounted cost optimal policy for the tandem queues with cost parameters $(c_1, c_2) = (1, 3)$. The load parameters are $\rho_1 = 9/10$, $\rho_2 = 9/11$, and the linear cost defined by $c_1 = 1$, $c_2 = 3$. On the left $\gamma = 0.01$ and on the right $\gamma = 0.001$.

Consider the CRW model described by the recursion,

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

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

$$\mathsf{P}\{\Phi(t) = \mathbf{1}^1\} = \mu_1, \quad \mathsf{P}\{\Phi(t) = \mathbf{1}^2\} = \mu_2, \quad \mathsf{P}\{\Phi(t) = \mathbf{1}^3\} = \alpha_1, \tag{64}$$

with $\mu_1 + \mu_2 + \alpha_1 = 1$. 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, as seen in the two examples shown in Figure 4 (see [39, Example 10.6.1]). Hence the optimal policy is similar in form to the h-myopic policy illustrated in Figure 3.

2.2.2   Perturbed value function

We now consider a function h_0 that serves as a Lyapunov function for the fluid model. 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-MaxWeight policy with $h_0 = J^*$, the optimal fluid value function defined below (42). In simple examples it is approximated by a switching curve with logarithmic growth.

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,

$$J^*(x) = \tfrac{1}{2} \frac{c_1}{\mu_2 - \alpha_1}(x_1 + x_2)^2 + \tfrac{1}{2} \frac{c_2 - c_1}{\mu_2} x_2^2, \qquad x \in \mathbb{R}_+^2. \tag{65}$$

Letting $h_0 = J^*$, the dynamic programing inequality (15) 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 (65) as

$$h_0(x) = J^*(x) = \tfrac{1}{2} d_1 (x_1 + x_2)^2 + \tfrac{1}{2} d_2 x_2^2, \qquad x \in \mathbb{R}_+^2,$$

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

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

21

Writing $Bu + \alpha = (-\mu_1 u_1 + \alpha_1, \mu_1 u_1 + \mu_2 u_2)^{\mathrm{T}}$, we obtain for any $x \in \mathbb{Z}_+^\ell$, $u \in \mathsf{U}(x)$,

$$\langle \nabla h\,(x), Bu + \alpha \rangle = \mu_1 u_1 \big[d_1 (e^{-x_1/\theta} - e^{-x_2/\theta})(\tilde{x}_1 + \tilde{x}_2) + d_2(1 - e^{-x_2/\theta})\tilde{x}_2 \big]$$

$$- \mu_2 u_2 \big[d_1(1 - e^{-x_2/\theta})(\tilde{x}_1 + \tilde{x}_2) + d_2(1 - e^{-x_2/\theta})\tilde{x}_2 \big]$$

$$+ \alpha_1 d_1(1 - e^{-x_1/\theta})(\tilde{x}_1 + \tilde{x}_2)$$

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,

$$d_1(e^{-x_1/\theta} - e^{-x_2/\theta})(\tilde{x}_1 + \tilde{x}_2) + d_2(1 - e^{-x_2/\theta})\tilde{x}_2 = 0. \tag{66}$$

When x_1 is large we obtain the approximation,

$$x_2 \approx s(x_1) := \theta \log\Big(1 + \frac{d_1}{d_2} x_1\Big), \tag{67}$$

where by (65),

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

This is an approximation to (66) 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 (66), 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 (67) is similar to the policy introduced in [37] to obtain HTAO (see eq. (32) and the surrounding discussion.)
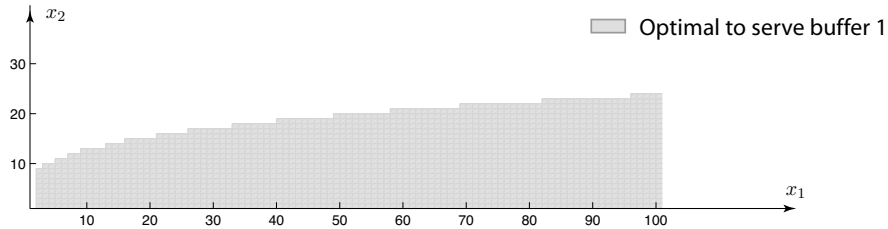

Figure 5: Average cost optimal policy for the tandem queues with $c_1 = 1$, $c_2 = 3$, $\rho_1 = 9/11$, $\rho_2 = 9/10$.

Consider now the average-cost optimal policy for the CRW model with statistics defined in (64), 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 [5] and [39, Chapter 9].) Moreover, Theorem 7.2 of [34] 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 (67).

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

22

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

$$\langle \nabla h\left(x\right), v^{\mathrm{MW}}\rangle \leq -c(x) + k_{2.1}\log(1 + \|x\|), \qquad x \in \mathbb{Z}_+^\ell,$$

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

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

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

To prove the lemma we compare v^{MW} with another velocity vector $v \in \mathsf{V}$, subject to the following constraints:

$$v_i \geq 0 \quad \text{whenever} \quad x_i < s_-(\|x\|), \qquad i = 1, \ldots, \ell. \tag{68}$$

The minimum of $\langle \nabla h\left(x\right), v\rangle$ 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 as we search for bounds on this inner product.

The purpose of (68) is to obtain the following bound:

$$- e^{-x_i/\theta}v_i \leq |v_i|\left(1 + \|x\|/\beta_-\right)^{-\beta_-/\theta}, \qquad i = 1, \ldots, \ell. \tag{69}$$

Since h_0 is assumed monotone we have $\nabla h_0 \colon \mathbb{R}_+^\ell \to \mathbb{R}_+^\ell$, and applying (54) we obtain,

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

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

$$\langle \nabla h\left(x\right), v\rangle \leq \langle \nabla h_0\left(\tilde{x}\right), v\rangle + k_0 \tag{70}$$

To bound (70) we shift $\tilde{x}$ as follows: Let $\tilde{x}^- \in \mathbb{Z}_+^\ell$ denote the vector with components

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

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

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

Moreover, we have $\tilde{x}_i^- = 0$ whenever the constraint on x_i in (68) is active. Since $u \in \mathsf{U}(x)$, this implies that the vector $v = Bu + \alpha$ satisfies $v_i \geq 0$. That is, v satisfies the constraint (68).

Using this v in (70) gives

$$\langle \nabla h\left(x\right), v\rangle \leq \langle \nabla h_0\left(\tilde{x}^-\right), v\rangle + \langle \nabla h_0\left(\tilde{x}\right) - \nabla h_0\left(\tilde{x}^-\right), v\rangle + k_0$$

$$\leq -c(\tilde{x}^-) + \|v\|\|\nabla h_0\left(\tilde{x}\right) - \nabla h_0\left(\tilde{x}^-\right)\| + k_0$$

$$\leq -c(x) + |c(x) - c(\tilde{x}^-)| + \|v\|\|\nabla h_0\left(\tilde{x}\right) - \nabla h_0\left(\tilde{x}^-\right)\| + k_0.$$

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 $k_{2.2}$,*

$$\mathcal{D}h(x) \le \langle \nabla h(x), v^{\mathrm{MW}} \rangle + k_{2.2}(1 + \theta^{-1}\|x\|), \qquad x \in \mathbb{Z}_+^\ell,$$

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

Proof. The first order Mean Value Theorem (37) results in the representation (38) for $\mathcal{D}h$ with b_h defined in (39). The Cauchy-Schwartz inequality gives,

$$b_h(x) \le \mathsf{E}\big[\|\nabla h(\bar{Q}) - \nabla h(Q(t))\|^2 \mid Q(t) = x\big]^{\frac{1}{2}} \mathsf{E}\big[\|\Delta(t+1)\|^2 \mid Q(t) = x\big]^{\frac{1}{2}} \qquad (71)$$

It remains to bound the right hand side.

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

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

where $\tilde{\bar{Q}}$ is obtained from $\bar{Q}$ via the pointwise transformation (16). The first expectation at right in (71) is bounded through an application of the triangle inequality,

$$\mathsf{E}\big[\|\nabla h(\bar{Q}) - \nabla h(Q(t))\|^2 \mid Q(t) = x\big]^{\frac{1}{2}} \le \mathsf{E}\big[\|[I - M_\theta(\bar{Q})](\nabla h_0(\tilde{\bar{Q}}) - \nabla h_0(\tilde{x}))\|^2 \mid Q(t) = x\big]^{\frac{1}{2}}$$
$$+ \mathsf{E}\big[\|[M_\theta(x) - M_\theta(\bar{Q})]\nabla h_0(\tilde{x})\|^2 \mid Q(t) = x\big]^{\frac{1}{2}} \tag{72}$$

To bound the first term on the right hand side of (72) we apply the Lipschitz condition on h_0: For some constant k_0,

$$\|\nabla h_0(\tilde{\bar{Q}}) - \nabla h_0(\tilde{x})\| \le k_0\|\tilde{\bar{Q}} - \tilde{x}\| \le \|\Delta(t+1)\|$$

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,

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

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 $k_3 < \infty$,

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

This combined with (19), (71) and (72) completes the proof. ☐

2.3 Universally stabilizing policies

The policies described in the previous sections are stabilizing, provided the parameter $\theta > 0$ is chosen sufficiently large. With a different change of variables we can construct a family of policies that are stabilizing regardless of the parameter.

In this subsection only we redefine $\tilde{x}$ as follows,

$$\tilde{x}_i := x_i \log(1 + x_i/\theta), \tag{73}$$

where $\theta > 0$ is fixed, but arbitrary. Theorem 1.1 and Proposition 2.5 can be generalized using this new definition of $\tilde{x}$. First we require derivative formulae.

Proposition 2.6.   *For any C^1 function h_0, the function h defined in (17) with $\tilde{x}$ defined component-wise in (73) satisfies the derivative conditions (14):*

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

$$\nabla h\,(x) = L_\theta \nabla h_0\,(\tilde{x}), \tag{74}$$

   where,
$$L_\theta(x) = \operatorname{diag}\left(x_i/(\theta + x_i) + \log(1 + x_i/\theta)\right), \qquad x \in \mathbb{R}_+^\ell. \tag{75}$$

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

$$\nabla^2 h\,(x) = L_\theta \nabla^2 h_0\,(\tilde{x}) L_\theta + \theta^{-1} \operatorname{diag}\left(N_\theta \nabla h_0\,(\tilde{x})\right), \tag{76}$$

   where $N_\theta(x) = \operatorname{diag}\left(\theta/(\theta + x_i) + \theta^2/(\theta + x_i)^2\right)$. In particular, h is convex provided h_0 is both convex and monotone.

$\square$

Proposition 2.7 establishes stability of the h-MaxWeight policy when h_0 is linear. We omit the proof since it is so similar to the proof of Proposition 2.5 (i), applying Proposition 2.6 instead of Proposition 2.3.

Proposition 2.7.   *Suppose that Assumptions (i) and (ii) of Theorem 1.1 hold, along with the stabilizability condition (41). Then, under the h-MaxWeight policy with h_0 linear, and $\tilde{x}$ defined in (73) with $\theta > 0$, there exists $\varepsilon = \varepsilon(\theta) > 0$ such that Condition (V3) holds with $f(x) = 1 + \varepsilon \log(1 + \|x\|)$.* $\square$

We now consider a version of Theorem 1.1. It is necessary to strengthen the L_2 condition on the arrival process to a second moment bound on $\widetilde{A}$,

$$\mathsf{E}[\|\tilde{A}(t)\|^2] := \sum_{i=1}^{\ell} \mathsf{E}[(A_i(t) \log(1 + A_i(t)/\theta))^2] < \infty \tag{77}$$

Moreover, it appears that the monotonicity assumption must be strengthened to the following,

25

$$\frac{\partial}{\partial x_i} h_0(x) \geq \varepsilon_{78} x_i, \qquad x \in \mathbb{R}_+^\ell, \ i = 1, \ldots, \ell, \tag{78}$$

where $\varepsilon_{78} > 0$ is constant. For example, (78) holds for a quadratic (1) in which $D_{ij} \geq 0$ and $D_{ii} > 0$ for each i, j.

Theorem 2.5. *Consider the h-MaxWeight policy in which the function h is based on the change of variables $\tilde{x}$ defined in (73). The constant $\theta > 0$ is fixed, but arbitrary. The network model (2) and function h_0 satisfy all of the assumptions of Theorem 1.1. In addition, the arrival process satisfies (77), and h_0 satisfies the uniform bound (78) with $\varepsilon_{78} > 0$. Then, there exists $\delta_h > 0$ and $\overline{\eta}_h < \infty$ such that the following version of the Poisson inequality (43) holds under the h-MaxWeight policy:*

$$\mathcal{D}h(x) = \mathsf{E}[h(Q(t+1)) - h(Q(t)) \mid Q(t) = x] \leq -\delta_h \|x\| \big(\log(1 + \|x\|)\big)^2 + \overline{\eta}_h. \tag{79}$$

Proof. We first obtain an extension of Lemma 2.1: There exists $\delta_h^0 > 0$ such that,

$$\langle \nabla h(x), v^{\mathrm{MW}} \rangle \leq -\delta_h^0 \|x\| \big(\log(1 + \|x\|)\big)^2, \qquad x \in \mathbb{Z}_+^\ell. \tag{80}$$

As in the proof of Proposition 2.4 we have the bound $\langle \nabla h(x), v^{\mathrm{MW}} \rangle \leq \langle \nabla h(x), v \rangle$ with $v_i = -\varepsilon$, $i \geq 1$, and $\varepsilon > 0$ chosen so that $v \in \mathsf{V}$. Hence,

$$\langle \nabla h(x), v^{\mathrm{MW}} \rangle \leq -\varepsilon \sum_{i=1}^\ell \frac{\partial}{\partial x_i} h(x) = -\varepsilon \sum_{i=1}^\ell (x_i/(\theta + x_i) + \log(1 + x_i/\theta)) \frac{\partial}{\partial x_i} h_0(\tilde{x}).$$

This combined with the bound (78) gives,

$$\langle \nabla h(x), v^{\mathrm{MW}} \rangle \leq -\varepsilon \varepsilon_{78} \sum_{i=1}^\ell \log(1 + x_i/\theta) \tilde{x}_i$$

From the definition of $\tilde{x}$ we obtain (80) for some $\delta_h^0 > 0$.

Based on (80) we complete the proof using an extension of Lemma 2.2. Combining (71) and (80) gives,

$$\begin{aligned}
\mathcal{D}h(x) \leq{} & -\delta_h^0 \|x\| \big(\log(1 + \|x\|)\big)^2 \\
& + \mathsf{E}\big[\|\nabla h(\bar{Q}) - \nabla h(Q(t))\|^2 \mid Q(t) = x\big]^{\frac{1}{2}} \mathsf{E}\big[\|\Delta(t+1)\|^2 \mid Q(t) = x\big]^{\frac{1}{2}}
\end{aligned} \tag{81}$$

Following the arguments used in the proof of Lemma 2.2, based on Proposition 2.6 in place of Proposition 2.3, we obtain the bound, for some constant k_0,

$$\mathsf{E}\big[\|\nabla h(\bar{Q}) - \nabla h(Q(t))\|^2 \mid Q(t) = x\big]^{\frac{1}{2}} \leq k_0 \|x\| \big(\log(1 + \|x\|)\big).$$

This together with (81) completes the proof of (79), where the constant δ_h can be chosen arbitrarily in the open interval $(0, \delta_h^0)$. $\qquad\square$

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 (33) subject to the following conventions. The load parameters defined in (34) 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,

$$W(t+1) = W(t) - S_1(t+1) + S_1(t+1)\iota(t) + L_1(t+1), \qquad t \ge 0, \tag{82}$$

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

The one-dimensional relaxation is defined on the same probability space with Q, and evolves as a controlled random walk analogous to (82):

$$\widehat{W}(t+1) = \widehat{W}(t) - S_1(t+1) + S_1(t+1)\widehat{\iota}(t) + L_1(t+1), \qquad t \ge 0. \tag{83}$$

The idleness process $\{\widehat{\iota}(t)\}$ is assumed to be non-negative, and adapted to $\{\widehat{W}(t), S_1(t), L_1(t)\}$. The relaxation is denoted $\widehat{W}^*$ when controlled using the non-idling policy, $\widehat{\iota}^*(t) = \mathbb{1}\{\widehat{W}^*(t) \ge 1\}$. In this case we have (35), 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 $c \colon \mathbb{R}_+^\ell \to \mathbb{R}_+$ the effective cost $\overline{c} \colon \mathbb{R}_+ \to \mathbb{R}_+$ is defined in (24). 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 (24):

$$\mathcal{X}^*(w) = \underset{x \in \mathbb{R}_+^\ell}{\arg\min}\Big(c(x) : \xi^{\mathrm{T}}x = w\Big). \tag{84}$$

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

$$c(Q(t)) \ge \overline{c}(W(t)) \ge \overline{c}(\widehat{W}^*(t)). \tag{85}$$

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

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

where the index i^* is any solution to $c_{i^*}/\xi_{i^*} = \min_{1 \le i \le \ell}(c_i/\xi_i)$. 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 $c(x) = \frac{1}{2}x^{\mathrm{T}}Dx$, $x \in \mathbb{R}_+^\ell$ 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,

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

and the effective cost is the one-dimensional quadratic,

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

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 (31) 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 $i \in \{1, \dots, \ell\}$. 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,

$$\overline{c}(w) = c(\mathcal{X}^*(w)) = \frac{c_1}{\xi_1}w, \qquad w \in \mathbb{R}_+. \tag{86}$$

We assume moreover that the solution to (24) is unique, which amounts to the following strict bound,

$$\frac{c_i}{\xi_i} > \frac{c_1}{\xi_1}, \qquad i = 2, \dots, \ell. \tag{87}$$

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

$$\widehat{\eta}^* = \tfrac{1}{2}\frac{\sigma_\kappa^2}{\mu_1 - \lambda_1}\frac{c_1}{\xi_1}, \tag{88}$$

where $\sigma_\kappa^2 := \rho_\bullet \mathsf{E}[(S_1(1) - L_1(1))^2] + (1 - \rho_\bullet)\mathsf{E}[(L_1(1))^2]$. From (85) we evidently have $\widehat{\eta}^* \leq \eta^*$. To establish (31) 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 $\kappa_0 \geq 1$, $\bar{\rho} < 1$ we have

$$\begin{aligned} \rho_\bullet := \rho_1 = 1 - 1/\kappa \qquad &\text{for each } \kappa \in [\kappa_0, \infty), \\ \text{and} \quad \rho_i \leq \bar{\rho} \qquad &\text{for each } \kappa \text{ and } i \geq 2. \end{aligned} \tag{89}$$

The one-dimensional workload process $\boldsymbol{W}$ is given in (82), 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 $\iota(t) \geq 0$. Given the cost function $\bar{c}$ given in (86), the fluid value function is given by,

$$\widehat{J}^*(w) = \tfrac{1}{2}\frac{c_1}{\xi_1}\frac{w^2}{\mu_1}\kappa, \qquad w \geq 0. \tag{90}$$

The solution to Poisson's equation (51) 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

$$h_0(x) = \widehat{J}^*(\xi^{\mathrm{T}}x) + \tfrac{1}{2}b\big(c(x) - \bar{c}(\xi^{\mathrm{T}}x)\big)^2. \tag{91}$$

This function is monotone since $c(x) \geq \bar{c}(\xi^{\mathrm{T}}x)$ for each x. We then take $h(x) = h_0(\tilde{x})$, or

$$h(x) := \widehat{J}^*(\widetilde{w}) + \tfrac{1}{2}b\big(c(\tilde{x}) - \bar{c}(\widetilde{w})\big)^2, \qquad x \in \mathbb{R}_+^\ell, \tag{92}$$

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 (89), the representation (88) becomes,

$$\widehat{\eta}^* = \tfrac{1}{2}\frac{\sigma_\kappa^2}{\mu_1}\frac{c_1}{\xi_1}\kappa. \tag{93}$$

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 (31) implies the bounds,

$$\eta^* \leq \eta \leq \eta^* + k_{94}\log(\kappa), \qquad\qquad \kappa \geq \kappa_0. \tag{94}$$

Theorem 3.1. *Suppose that the following hold for the parameterized family of networks.*

(HTAO 1) For each κ, the network is a CRW scheduling model with deterministic routing.

(HTAO 2) The network is homogeneous for each κ. Moreover,

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

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

29

(b) *The distribution of $S_s^\kappa(t)$ is Bernoulli for each s, and $\mathsf{E}[\|A^\infty(t)\|^2] < \infty$.*

(c) *For some $\varepsilon > 0$ independent of κ and s,*

$$\mathsf{P}\{S_s^\kappa(t) = 1 \text{ and } A^\kappa(t) = \mathbf{0}\} \geq \varepsilon. \tag{95}$$

(HTAO 3) Linear cost.

(HTAO 4) The load parameters satisfy (89). Hence $\rho^\kappa \to \rho^\infty$ as $\kappa \to \infty$, where $\rho_1^\infty = 1$ and $\rho_i^\infty < 1$ for $i \geq 2$.

Moreover, the effective state is unique: Equation (87) holds.

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 (92), 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 $\mathbf{0}$-irreducible and (V3) holds, for each $\kappa \geq \kappa_0$. Moreover, the family of controlled networks satisfies the bound (94) for some fixed $k_{94} < \infty$. 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),

$$\mathsf{E}_x[V(Q(1))] \leq V(x) - c(x) + \widehat{\eta}^* + \mathcal{E}(x), \qquad x \in \mathbb{Z}_+^\ell, \tag{96}$$

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

$$\mathcal{E}(x) = k_{97}\Big(\log(\kappa + c(x)) + \kappa/(1 + (\xi^{\mathsf{T}}x)^2)\Big), \qquad x \in \mathbb{R}_+^\ell,\ \kappa \geq 1. \tag{97}$$

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 (38). 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 (97). The constant k_{97} may differ in each appearance.

We first establish irreducibility:

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

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

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

Combining this with (95) we can conclude that for some $\delta > 0$, and any non-zero $x \in \mathsf{X}_\diamond$,

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

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

Aperiodicity also follows from (95) since $P(0, 0) = \mathsf{P}\{A(t) = 0\} > 0$. $\qquad\square$

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 $\varepsilon_{3.2} > 0$, independent of b and x,*

$$\langle \nabla h\,(x), v^{\mathrm{MW}} \rangle \leq -\overline{c}(w) - \varepsilon\Big(b|c(x) - \overline{c}(w)| + \|M_\theta(x)\xi\|\kappa w \Big) + \mathcal{E}(x), \tag{98}$$

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

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

$$\mathcal{D}h\,(x) \leq \langle \nabla h\,(x), v^{\mathrm{MW}} \rangle + \widehat{\eta}^*(x) + k_{3.3}\theta^{-1}\Big(b|c(x) - \overline{c}(w)| + \|M_\theta(x)\xi\|\kappa w \Big) + \mathcal{E}(x), \tag{99}$$

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

$$\widehat{\eta}^*(x) := \tfrac{1}{2}\frac{\kappa}{\mu_1}\frac{c_1}{\xi_1}\sigma_\kappa^2(x), \quad \text{with} \quad \sigma_\kappa^2(x) := \mathsf{E}[(\xi^{\mathrm{T}}\Delta(t+1))^2 \mid Q(t) = x] \tag{100}$$

$\qquad\square$

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

$$c(x) - \overline{c}(w) = \sum_{i=2}^{\ell}\big(c_i - (c_1/\xi_1)\xi_i\big)x_i, \tag{101}$$

which is non-negative by (87). 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,

$$\nabla h\,(x) = \nabla\widehat{J}^*(\widetilde{w})[I - M_\theta]\xi + b(c(\tilde{x}) - \overline{c}(\widetilde{w}))[I - M_\theta]\Big(c - \frac{c_1}{\xi_1}\xi\Big), \tag{102}$$

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,*

$$\langle \nabla h\,(x), v \rangle = \nabla \widehat{J}^*(\widetilde{w}) \langle \xi, v \rangle - \nabla \widehat{J}^*(\widetilde{w}) \sum_{i=1}^{\ell} e^{-x_i/\theta} \xi_i v_i$$
$$+ [c(\tilde{x}) - \overline{c}(\widetilde{w})] \sum_{i=2}^{\ell} (1 - e^{-x_i/\theta}) b_i v_i. \tag{103}$$

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

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

$$s_-(r) = \beta_- \log(1 + r/\beta_-), \quad s_+(r) = \beta_+ \log(1 + r/\beta_+). \tag{104}$$

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

Similar to the proof of Lemma 2.1, to bound (103) we impose the following constraints on the velocity vector v.

$$v_i \geq 0 \quad \text{if } x_i \leq s_-(\xi^{\mathrm{T}}x) - \beta_- \log(|\xi|) \text{ and } \xi_i > 0, \tag{105}$$

where $|\xi| := \sum \xi_i$. The minimum of $\langle \nabla h\,(x), v \rangle$ over v satisfying (105) provides a bound under the h-MaxWeight policy. The following two results imply that (105) is feasible for a policy that is non-idling at Station 1.

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

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

$$s_-(w) = \beta_- \log(1 + w/\beta_-) \leq \beta_- \log(1 + x^\infty |\xi|/\beta_-)$$
$$\leq \beta_- \Big(\log(|\xi|) + (1 + x^\infty |\xi|/\beta_- - |\xi|)/|\xi| \Big).$$

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

We now establish a set of feasible values for v.

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

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

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. To avoid trivialities we assume that $\ell \geq 3$ (no less than three buffers), and that $\xi_i > 0$ for at least three values of i.

The drift obtained with any $v \in \mathsf{V}$ provides an upper bound on the drift obtained using the h-MaxWeight policy. 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 $\xi^{\mathrm{T}}v = \xi^{\mathrm{T}}v^* = -(\mu_1 - \lambda_1)$. 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.

Lemma 3.3 implies that we can find $\varepsilon_{3.3} > 0$ such that the following inclusion holds for each $\kappa \geq \kappa_v$,

$$\big\{ v = v^* + v^0 \in \mathbb{R}^\ell : \xi^{\mathrm{T}}v^0 = 0,\ \text{and}\ |v_i^0| \leq \varepsilon_{3.3}\ \text{for each}\ i \geq 1 \big\} \subset \mathsf{V}^\kappa.$$

With this value of $\varepsilon_{3.3}$ fixed, we set $v_i^0 = -\varepsilon_{3.3}$ for each i satisfying $\xi_i = 0$, and $v_i^0 = 0$ 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 $v_{i_\ominus}^0 < 0$ and $v_{i_\oplus}^0 > 0$.

To complete the specification of v^0 we introduce the index sets,

$$I_\oplus = \{i \geq 1 : \xi_i > 0,\ x_i \leq s_-(\xi^{\mathrm{T}}x) - \beta_- \log(|\xi|)\}$$
$$I_\ominus = \{i \geq 2 : \xi_i > 0,\ x_i > s_+(\xi^{\mathrm{T}}x)\}$$

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

(i) If $I_\oplus = \emptyset$ and $I_\ominus = \emptyset$ then $v_i^0 = 0$ for each 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 $v_{i_\ominus}^0 \xi_{i_\ominus} = -\varepsilon_{3.3}$, and set $v_1^0 \xi_1 = \varepsilon_{3.3}$.

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

$$i_\oplus \in \arg\min\{x_i : i \geq 2,\ \xi_i > 0\},\quad i_\ominus = 1,$$

and define $v_{i_\oplus}^0 \xi_{i_\oplus} = \varepsilon_{3.3}^2$, $v_{i_\ominus}^0 \xi_{i_\ominus} = -\varepsilon_{3.3}^2$, and $v_i^0 = 0$ for all other i satisfying $\xi_i > 0$.

(iv) If $I_\oplus \neq \emptyset$ and $I_\ominus \neq \emptyset$ then $i_\ominus \in \arg\max\{x_i : i \geq 2,\ \xi_i > 0\}$. To determine $i_\oplus$ there are two sub-cases to consider.

   a) If $I_\oplus = \{1\}$ then $v_{i_\ominus}^0 \xi_{i_\ominus} = -\varepsilon_{3.3}$, and $i_\oplus = 1$ with $v_1^0 \xi_1 = \varepsilon_{3.3}$.

   b) Otherwise, $i_\oplus \in \arg\min\{x_i : i \geq 2,\ \xi_i > 0\}$, and we take,

$$v_{i_\oplus}^0 \xi_{i_\oplus} = \varepsilon_{3.3}^2,\quad v_{i_\ominus}^0 \xi_{i_\ominus} = -\varepsilon_{3.3},\quad \text{and}\quad v_1^0 \xi_1 = \varepsilon_{3.3} - \varepsilon_{3.3}^2$$

   where again $v_i = 0$ for all other i satisfying $\xi_i > 0$.

The added complexity in cases (iii) and (iv) is due to the positive drift induced by $v_{i_\oplus}$. By imposing the constraint that this is of order $\varepsilon_{3.3}^2$ rather than $\varepsilon_{3.3}$ we can maintain a negative overall drift.

This choice of v satisfies (105). Moreover, under the assumption that $\beta_- \geq 3\theta^{-1}$ and $\beta_+ > \beta_-$ we have for some constants k_{106}, ε_{106}, and all x,

$$\left|\nabla\widehat{J}^*(\widetilde{w})e^{-x_i/\theta}\xi_i v_i\right| \leq k_{106}\kappa(1+w^2)^{-1} \leq \mathcal{E}(x), \qquad i \notin I_{\oplus}.$$

$$-\nabla\widehat{J}^*(\widetilde{w})e^{-x_{i_\oplus}/\theta}\xi_{i_\oplus} v_{i_\oplus} \leq -\varepsilon_{106}\|M_\theta(x)\xi\|\kappa w, \qquad \text{if } I_\oplus \neq \emptyset. \tag{106}$$

$$\|M_\theta(x)\xi\|\kappa w \leq \mathcal{E}(x), \qquad \text{if } I_\oplus = \emptyset.$$

Combining the bounds in (106) with Lemma 3.1 we obtain,

$$\langle\nabla h(x), v\rangle \leq -(\mu_1 - \lambda_1)\nabla\widehat{J}^*(\widetilde{w}) - \varepsilon_{106}\|M_\theta(x)\xi\|\kappa w + \mathcal{E}(x)$$
$$+ [c(\tilde{x}) - \overline{c}(\widetilde{w})]\sum_{i=2}^{\ell}(1 - e^{-x_i/\theta})b_i v_i.$$

To complete the proof we argue that the following bound holds: For some $\varepsilon_{107} > 0$,

$$[c(\tilde{x}) - \overline{c}(\widetilde{w})]\sum_{i=2}^{\ell}(1 - e^{-x_i/\theta})b_i v_i \leq -\varepsilon_{107}b[c(\tilde{x}) - \overline{c}(\widetilde{w})] + \mathcal{E}(x). \tag{107}$$

It is here that we require the the fact that at most one value of v_i is positive for $i \geq 2$, and that for all such i we have the bound $v_i\xi_i \leq \varepsilon_{3.3}^2$.

If $x_i > s_+(\xi^{\mathrm{T}}x)$ for some $i \geq 2$ (not necessarily satisfying $\xi_i > 0$) then $v_i = -\varepsilon_{3.3}$ for some $i \geq 2$. In fact, with $i_- \in \arg\max\{x_i : i \geq 2\}$ we have $v_{i_-} = -\varepsilon_{3.3}$, and from (101) we obtain $c(x) - \overline{c}(w) \leq |c|x_{i_-}$. Consequently,

$$\sum_{i=2}^{\ell}(1 - e^{-x_i/\theta})b_i v_i \leq -(b_{i_-}\varepsilon_{3.3} - b_{i_\oplus}\varepsilon_{3.3}^2) + \varepsilon_{3.3}b_{i_-}e^{-[c(x)-\overline{c}(w)]/(|c|\theta)}$$

The bound (107) follows for $\varepsilon_{3.3} > 0$ sufficiently small: fix $\varepsilon_{3.3} < \min_{i,j\geq 2} b_i/b_j$ and set $\varepsilon_{107} = \min_{i,j\geq 2}(b_i\varepsilon_{3.3} - b_j\varepsilon_{3.3}^2)/b$. Note that the positive term $b_{i_\oplus}\varepsilon_{3.3}^2$ 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 $v_i = -\varepsilon_{3.3}$ 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 (107) follows trivially. □

Proof of Proposition 3.3. We begin with a representation of the form (38) based on a second-order Mean Value Theorem of the form (61). 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,

$$\mathcal{D}h(x) = \langle\nabla h(x), v\rangle + \tfrac{1}{2}\mathsf{E}\left[\Delta(t+1)^{\mathrm{T}}H_0\Delta(t+1) \mid Q(t) = x\right]$$
$$+ \tfrac{1}{2}\mathsf{E}\left[\Delta(t+1)^{\mathrm{T}}(\nabla^2 h(\bar{Q}) - H_0)\Delta(t+1) \mid Q(t) = x\right]. \tag{108}$$

We also have by definition of $\widehat{\eta}^*(x)$ in (100),

34

$$\mathsf{E}\big[\Delta(t+1)^{\mathrm{T}}H_0\Delta(t+1) \mid Q(t) = x\big] = \widehat{\eta}^*(x) + b\mathsf{E}\big[\big((c-(c_1/\xi_1)\xi)^{\mathrm{T}}\Delta(t+1)\big)^2 \mid Q(t) = x\big]. \quad (109)$$

We apply Proposition 2.3 to bound the final term in (108):

$$\nabla^2 h\,(x) - H_0 = -[M_\theta H_0 + M_\theta H_0] + M_\theta H_0 M_\theta + \theta^{-1}\mathrm{diag}\,(M_\theta \nabla h_0(\tilde{x})). \qquad (110)$$

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

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

where terms that are independent of κ are suppressed using the "big O" notation. Applying the Mean Value Theorem as in the proof of Lemma 2.2 we obtain the crude bound, $\|(I - M_\theta)\xi\| \leq \theta^{-1}w$, and hence for some $k_0 < \infty$,

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

Also, for a possibly larger constant k_0,

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

Consequently, for some $k_0 < \infty$,

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

This combined with (108) and (109) 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 (96) for a function V derived from h. We define,

$$V(x) = h(x) + \tfrac{1}{2}\frac{c_1}{\xi_1}\mu^{-1}\Big(\frac{m^2 - m_L^2}{\mu - \alpha}\Big)\xi^{\mathrm{T}}x, \qquad x \in \mathbb{Z}_+^\ell,$$

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. Based on the definitions of $\widehat{\eta}^*$ and $\widehat{\eta}^*(x)$ in (88) and (100) we obtain the following identity *for any policy*,

$$\mathsf{E}\Big[\tfrac{1}{2}\frac{c_1}{\xi_1}\mu^{-1}\Big(\frac{m^2 - m_L^2}{\mu - \alpha}\Big)\big(W(t+1) - W(t)\big) \mid Q(t) = x\Big] = \widehat{\eta}^* - \widehat{\eta}^*(x).$$

Hence the function V does satisfy (96).

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

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

35

From the form of $\mathcal{E}$ given in (97), it follows that $\pi(c)$ is bounded by a constant times κ. In fact, by the bound above, (97), and Jensen's inequality we obtain,

$$\pi(c) \leq \widehat{\eta}^* + k_{97} \log(\kappa + \pi(c)) + k_{97}\kappa \mathsf{E}_\pi[(1 + (\xi^{\mathrm{T}} Q(t)))^2)^{-1}]$$

so that for a possibly larger constant,

$$\pi(c) \leq \widehat{\eta}^* + k_{97} \log(\kappa) + k_{97}\kappa \mathsf{E}_\pi[(1 + (\xi^{\mathrm{T}} Q(t)))^2)^{-1}].$$

Moreover, applying (35) we obtain,

$$\pi(c) \leq \widehat{\eta}^* + k_{97} \log(\kappa) + k_{97}\kappa \mathsf{E}[(1 + (\widehat{W}^*(t))^2)^{-1}],$$

where the expectation is taken for the steady-state relaxation. Lemma A.2 of [37] implies that $\kappa \mathsf{E}[(1 + (\widehat{W}^*(t))^2)^{-1}]$ is uniformly bounded in κ, so this final bound completes the proof. $\square$

4   Extensions, Conjectures, and 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.

There remain many questions.

Statistics   Generalizations of Theorems 1.1 and 2.5 to network models with renewal arrivals and service is straightforward by applying fluid limit techniques [12, 15, 13]. It would be of interest to develop alternative methods to cope with these more complex models to obtain sharp performance estimates in heavy traffic. In particular, to-date logarithmic regret has been established only for homogeneous Markovian models. To relax the homogeneity (Kelly-type) assumption, the workload process should be constructed in units of time rather than inventory as done here. It would be much more interesting to find a counterexample within the class of renewal models with finite second moment, though this is not likely to exist.

Of greater practical importance is the issue of memory: Long-range dependent models remain a frontier. It may be possible to extend fluid limit techniques to establish stability. New techniques are required to obtain performance bounds.

Information   Design of the function h_0 requires stability considerations, e.g., monotonicity has been imposed as a blanket assumption in each of the main results. A second consideration is the amount of information required for implementation.

Consider the special case of the quadratic function (1). Monotonicity holds provided $D_{ij} \geq 0$ for each i, j, and the dynamic programing inequality holds if in addition $D_{ii} > 0$ for each i. In typical scheduling and routing models the MaxWeight policy in which D is diagonal requires only local information [52]. The main result of this paper shows that this can be relaxed through the introduction of a state transformation. If the matrix D that defines h_0

is sparse, then the amount of information required for implementation of the h-MaxWeight policy will be limited.

Suppose that D is a band matrix with width $n_0 \geq 1$ ($D_{ij} = 0$ for each i, j satisfying $|i - j| > n_0$). For the same scheduling and routing models considered in earlier work, the resulting h-MaxWeight policy will require "$(n_0 + 1)$-hop" local information. For example, for a network consisting of a sequence of N queues in tandem, for each integer $i \in [n_0, N - n_0 - 1]$ the policy at Station i will require queue length information at buffers $\{i \pm k, |k| \leq n_0\}$ and buffer $i + n_0 + 1$.

Another important class of quadratics is those with low rank. Suppose that for linearly independent vectors $\{d^i : i = 1, \ldots, n_1\} \subset \mathbb{R}_+^\ell$ we have,

$$h_0(x) = \tfrac{1}{2} \sum_{i=1}^{n_1} d^i d^{i\mathrm{T}}.$$

For example, the function (91) is of this form with $n_1 = 2$. This matrix is not banded in general, but the h-MaxWeight policy will again require limited information. For scheduling and routing models the required information is the same "one-hop" data required in the usual MaxWeight algorithm, along with values of the inner products $\{d^{i\mathrm{T}} Q(t) : i = 1, \ldots, n_1\}$. For small values of n_1 this information might be distributed through message passing.

Performance bounds for universally stabilizing policies   The h-MaxWeight policy considered in Theorem 2.5 using the change of variables (73) is universally stabilizing, but its performance is not understood. It will be interesting to evaluate its performance in the setting of Section 3. It will not yield logarithmic regret in general under the assumptions of Theorem 3.1.

Conjecture 1:   Under the assumptions of Theorem 3.1, with $\tilde{x}$ redefined via (73) and the L_2 bound (77) satisfied for $A^\infty(t)$, the h-MaxWeight policy will result in asymptotic minimality of workload, with logarithmic regret, in the sense that

$$\mathsf{E}[W(t; x)] \leq \mathsf{E}[\widehat{W}^*(t; x)] + k_1 \log(\kappa)$$

for some k_1, each κ sufficiently large, and each t and initial condition.   □

It is likely that minimality is not difficult to establish. It is also likely that state space collapse (25) will hold in some form, perhaps with $\mathcal{X}^*$ defined with respect to the perturbed cost function,

$$\mathcal{X}^*(w) = \arg\min_{x \in \mathbb{R}_+^\ell} \big(c(\tilde{x}) : \xi^\mathrm{T} x = w\big).$$

Linearity and logarithmic regret   If c is an arbitrary norm on $\mathbb{R}^\ell$ then the average cost for the relaxation becomes,

$$\widehat{\eta}^* = \tfrac{1}{2}\overline{c}(1)\frac{\sigma_\kappa^2}{\mu_1}\kappa,$$

where the definition of the effective cost is given in (24). This generalizes (93) since $\overline{c}(1) = c_1/\xi_1$ under the assumptions of Theorem 3.1. A stability proof is simplified if the function h_0 in (91) used to define the h-MaxWeight policy is redefined via,

$$h_0(x) = \widehat{J}^*(\xi^{\mathrm{T}}x) + \tfrac{1}{2}b\|x - x^*\|^2$$

Conjecture 2:   The conclusions of Theorem 3.1 remain valid when c is a general norm, h_0 redefined as above, and with the remaining assumptions of the theorem maintained.   □

Multiple bottlenecks   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.

To formulate a conjecture we again restrict to the homogeneous scheduling model with linear cost. If there are n bottlenecks in heavy traffic we consider an n-dimensional relaxation, as formulated in [9]. Analysis of the usual MaxWeight policy based on a workload relaxation of dimension $n \geq 2$ is contained in [39, Chapter 6].

The workload relaxation evolves in a positive cone $\mathsf{W} \subset \mathbb{R}^n$ according to a multi-dimensional recursion of the form (83). Let $\widehat{\eta}^*$ denote the optimal average cost, and suppose that $\hat{h}_* \colon \mathsf{W} \to \mathbb{R}$ solves the average-cost optimality equations for the relaxation, perhaps approximately: For some finite constant k_3 and all κ,

$$\min \mathsf{E}[\hat{h}_*(\widehat{W}(t+1)) \mid \widehat{W}(t) = w, \ \iota(t) = \iota] \leq \hat{h}_*(w) - \overline{c}(w) + \widehat{\eta}^* + k_3 \log(\kappa)$$

where $\iota(t)$ denotes the idleness at time t, and the minimum is over all $\iota \geq 0$ such that $\widehat{W}(t+1) \in \mathsf{W}$ with probability one. The integer constraints are relaxed so that $\hat{h}_*$ is defined on all of W. Based on the relative value function for the relaxation, the function h_0 in (91) is redefined via,

$$h_0(x) := \hat{h}_*(\xi^{\mathrm{T}}x) + \tfrac{1}{2}b[c(x) - \overline{c}(\xi^{\mathrm{T}}x)]^2, \qquad x \in \mathbb{R}_+^\ell. \tag{111}$$

The function $\hat{h}_*$ will depend upon κ in a parameterized model, but the constant b will be fixed as in Theorem 3.1.

Conjecture 3:   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_0 given in (111).   □

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

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 [3, 50, 53, 16, 45]. It is hoped that specialized algorithms can be constructed for networks based on the techniques introduced here.

References

[1] 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.

[2] 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.

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

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

[5] 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.

[6] M. Bramson and R. J. Williams. Two workload properties for Brownian networks. *Queueing Syst. Theory Appl.*, 45(3):191–221, 2003.

[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] J. G. Dai. On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. *Ann. Appl. Probab.*, 5(1):49–77, 1995.

[13] J. G. Dai and Wuqin Lin. Maximum pressure policies in stochastic processing networks. *Operations Res.*, 53(2):197–218, 2005.

[14] J. G. Dai and Wuqin Lin. Asymptotic optimality of maximum pressure policies in stochastic processing networks. To appear in *The Annals of Applied Probability*, 2008.

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

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

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

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

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

[20] 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.

[21] 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.

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

[23] S. G. Henderson, S. P. Meyn, and V. B. Tadić. Performance evaluation and policy selection in multiclass networks. *Discrete Event Dynamic Systems: Theory and Applications*, 13(1-2):149–189, 2003. Special issue on learning, optimization and decision making (invited).

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

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

[26] 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.

[27] 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.

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

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

[32] C. Maglaras. Discrete-review policies for scheduling stochastic networks: trajectory tracking and fluid-scale asymptotic optimality. *Ann. Appl. Probab.*, 10(3):897–929, 2000.

[33] 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.

[34] 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.

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

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

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

[38] S. P. Meyn. Workload models for stochastic networks: Value functions and performance evaluation. *IEEE Trans. Automat. Control*, 50(8):1106– 1122, August 2005.

[39] S. P. Meyn. *Control techniques for complex networks*. Cambridge University Press, Cambridge, 2008.

[40] S. P. Meyn and D. G. Down. Stability of generalized Jackson networks. *Ann. Appl. Probab.*, 4:124–148, 1994.

[41] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, second edition, 1993. 2008 Edition to appear, Cambride University Press, Cambridge Mathematical Library. 1993 edition online: `http://black.csl.uiuc.edu/~meyn/pages/book.html`.

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

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

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

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] V. Subramanian and D. Leith. Draining time based scheduling algorithm. In *Proceedings of the 46th IEEE Conf. on Decision and Control*, pages 1162–1167, 2007.

[50] R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction.* MIT Press, online edition at http://www.cs.ualberta.ca/~sutton/book/the-book.html edition, 1998.

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

[52] 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.

[53] 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.

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