Noname manuscript No.
(will be inserted by the editor)

Signaling for Decentralized Routing in a Queueing Network

Yi Ouyang · Demosthenis Teneketzis

Received: date / Accepted: date

Abstract A discrete-time decentralized routing problem in a service system consisting of two service stations and two controllers is investigated. Each controller is affiliated with one station. Each station has an infinite size buffer. Exogenous customer arrivals at each station occur with rate λ. Service times at each station have rate μ. At any time, a controller can route one of the customers waiting in its own station to the other station. Each controller knows perfectly the queue length in its own station and observes the exogenous arrivals to its own station as well as the arrivals of customers sent from the other station. At the beginning, each controller has a probability mass function (PMF) on the number of customers in the other station. These PMFs are common knowledge between the two controllers. At each time a holding cost is incurred at each station due to the customers waiting at that station. The objective is to determine routing policies for the two controllers that minimize either the total expected holding cost over a finite horizon or the average cost per unit time over an infinite horizon. In this problem there is implicit communication between the two controllers; whenever a controller decides to send or not to send a customer from its own station to the other station it communicates information about its queue length to the other station. This implicit communication through control actions is referred to as signaling in decentralized control. Signaling results in complex communication and decision

Preliminary versions of this paper appeared in Ouyang and Teneketzis (2013) and Ouyang and Teneketzis (2014). This work was supported in part by National Science Foundation (NSF) Grant CCF-1111061 and NASA grant NNX12AO54G.

Y. Ouyang
Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI
E-mail: ouyangyi@umich.edu

D. Teneketzis
Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI
E-mail: teneket@umich.edu

problems. In spite of the complexity of signaling involved, it is shown that an optimal signaling strategy is described by a threshold policy which depends on the common information between the two controllers; this threshold policy is explicitly determined.

Keywords Decentralized System · Non-classical Information Structure · Signaling · Queueing Networks · Common Information

1 Introduction

Routing problems to parallel queues arise in many modern technological systems such as communication, transportation and sensor networks. The majority of the literature on optimal routing in parallel queues addresses situations where the information is centralized, either perfect (see Akgun et al (2012); Davis (1977); Ephremides et al (1980); Foley and McDonald (2001); Hajek (1984); Hordijk and Koole (1990, 1992); Lin and Kumar (1984); Menich and Serfozo (1991); Weber (1978); Weber and Stidham Jr (1987); Whitt (1986); Winston (1977) and references therein) or imperfect (see Beutler and Teneketzis (1989); Kuri and Kumar (1995) and references therein). Very few results on optimal routing to parallel queues under decentralized information are currently available. The authors of Cogill et al (2006) present a heuristic approach to decentralized routing in parallel queues. In (Abdollahi and Khorasani (2008); Manfredi (2014); Reddy et al (2012); Si et al (2013); Ying and Shakkottai (2011) and references therein), decentralized routing policies that stablize queueing networks are considered. The work in Pandelis and Teneketzis (1996) presents an optimal policy to a routing problem with a one-unit delay sharing information structure.

In this paper we investigate a decentralized routing problem in discrete time. We consider a system consisting of two service stations/queues, called Q_1 and Q_2 and two controllers, called C_1 and C_2. Controller C_1 (resp. C_2) is affiliated with service station Q_1 (resp. Q_2). Each station has an infinite size buffer. The processes describing exogenous customer arrivals at each station are independent Bernoulli with parameter (λ). The random variables describing the service times at each station are independent geometric with parameter (μ). At any time each controller can route one of the customers waiting in its own queue to the other station. Each controller knows perfectly the queue length in its own station, and observes the exogenous arrivals in its own station as well as the arrivals of customers sent from the other station. At the beginning, controller C_1 (resp. C_2) has a probability mass function (PMF) on the number of customers in station Q_2 (resp. Q_1). These PMFs are common knowledge between the controllers. At each time a holding cost is incurred at each station due to the customers waiting at that station. The objective is to determine decentralized routing policies for the two controllers that minimize either the total expected holding cost over a finite horizon or the average cost per unit time over an infinite horizon. Preliminary versions of this paper appeared in Ouyang and Teneketzis (2013) (for the finite horizon problem) and

Ouyang and Teneketzis (2014) (for the infinite horizon average cost per unit time problem).

In the above described routing problem, each controller has different information. Furthermore, the control actions/routing decisions of one controller affect the information of the other controller. Thus, the information structure of this decentralized routing problem is non-classical with control sharing (see Mahajan (2013) for non-classical control sharing information structures). Non-classical information structures result in challenging signaling problems (see Ho (1980)). Signaling occurs through the routing decisions of the controllers. Signaling is, in essence, a real-time encoding/communication problem within the context of a decision making problem. By sending or not sending a customer from Q_1 (resp. Q_2) to Q_2 (resp. Q_1) controller C_1 (resp. C_2) communicates at each time instant a compressed version of its queue length to C_2 (resp. C_1). For example, by sending a customer from Q_1 to Q_2 at time t, C_1 may signal to C_2 that Q_1's queue length is above a pre-specified threshold l_t. This information allows C_2 to have a better estimate of Q_1's queue length and, therefore, make better routing decisions about the customers in its own queue; the same arguments hold for the signals send (through routing decisions) from C_2 to C_1. Thus, signaling through routing decisions has a triple function: communication, estimation and control.

Within the context of the problems described above, there is enormous number of signaling possibilities. For example, there is an arbitrarily large number of choices of the sequences of pre-specified thresholds $l_1, l_2, \ldots, l_t, \ldots$ and these choices are only a small subset of all the possible sequences of binary partitions of the set of non-negative integers that describe all choices available to C_1 and C_2. All these possibilities result in highly non-trivial decision making problems. It is the presence of signaling that distinguishes the problem formulated in this paper from all other routing problems in parallel queues investigated so far.

Some basic questions associated with the analysis of this problem are:

What is an information state (sufficient statistic) for each controller? How is signaling incorporated in the evolution/update of the information state? Is there an explicit description of an optimal signaling strategy? We will answer these questions in Section 3-6 and will discuss them further in Section 7.

Contribution of the paper

The signaling feature of our problem distinguishes it from all previous routing problems in parallel queues. In spite of the complexity of signaling, we show that an optimal decentralized strategy is described by a single threshold routing policy where the threshold depends on the common information between the two controllers. We explicitly determine this threshold via simple computations.

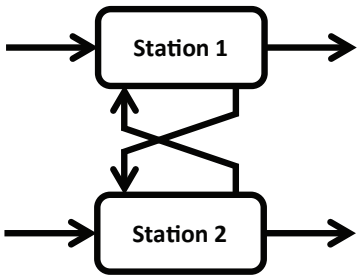

Fig. 1 The Queueing System

Organization

The rest of the paper is organized as follows. In Section 2 we present the model for the queueing system and formulate the finite horizon and infinite horizon decentralized routing problems. In Section 3 we present structural results for optimal policies. In Section 4 we present a specific decentralized routing policy, which we call $\hat{g}$, and state some features associated with its performance. In Section 5, we show that when the initial queue lengths in Q_1 and Q_2 are equal, $\hat{g}$ is an optimal policy for the finite horizon decentralized routing problem. In Section 6, we show that $\hat{g}$ is an optimal decentralized routing policy for the infinite horizon average cost per unit time problem. We conclude in Section 7.

Notation

Random variables (r.v.s) are denoted by upper case letters, their realization by the corresponding lower case letter. In general, subscripts are used as time index while superscripts are used to index service stations. For time indices $t_1 \leq t_2$, $X_{t_1:t_2}$ is the short hand notation for $(X_{t_1}, X_{t_1+1}, ..., X_{t_2})$. For a policy g, we use X^g to denote that the r.v. X^g depends on the choice of policy g. We use vectors in $\mathbb{R}^{\mathbb{Z}_+}$ to denote PMFs (Probability Mass Functions,) where $\mathbb{Z}_+$ denotes the set of non-negative integers. We also use a constant in $\mathbb{Z}_+$ to denote the corner PMF that represents a constant r.v.. i.e. a constant $c \in \mathbb{Z}_+$ denotes the PMF whose entries are all zero except the cth.

2 System Model and Problem Formulation

System Model

The queueing/service system shown in Figure 1, operates in discrete time. The system consists of two service stations/queues, Q_1 and Q_2 with infinite size buffers. Controllers C_1 and C_2 are affiliated with queues Q_1 and Q_2,

respectively. Let X_t^i denote the number of customers waiting, or in service, in $Q_i, i = 1, 2$, at the beginning of time t. Exogenous customer arrivals at $Q_i, i = 1, 2$, occur according to a Bernoulli process $\{A_t^i, t \in \mathbb{Z}_+\}$ with parameter λ. Service times of customers at $Q_i, i = 1, 2$ are described by geometric random variables with parameter μ. We define a Bernoulli process $\{D_t^i, t \in \mathbb{Z}_+\}$ with parameter μ. Then $\{D_t^i 1_{\{X_t^i \neq 0\}}, t \in \mathbb{Z}_+\}$ describes the customer departure process from $Q_i, i = 1, 2$. At any time t, a controller can route one of the customers in its own queue to the other queue. Let U_t^i denote the routing decision of controller C_i at t $(i = 1, 2)$; if $U_t^i = 1$ (resp. 0) one customer (resp. no customer) is routed from Q_i to Q_j $(j \neq i)$. At any time t, each controller $C_i, i = 1, 2$, knows perfectly the number of customers $X_{0:t}^i, i = 1, 2$, in its own queue; furthermore, it observes perfectly the arrival stream $A_{0:t}^i$ to its own queue, and the arrivals due to customers routed to its queue from the other service station up to time $t - 1$, i.e. $U_{0:t-1}^j, j \neq i$. The order of arrivals A_t^i, departures D_t^i and controller decisions U_t^i concerning the routing of customers from one queue to the other is shown in Figure 2. The dynamic evolution of the number of customers $X_t^i, i = 1, 2$ is described by

$$X_{t+1}^1 = \overline{X}_t^1 - U_t^1 + U_t^2, \tag{1}$$

$$X_{t+1}^2 = \overline{X}_t^2 - U_t^2 + U_t^1, \tag{2}$$

where for $i = 1, 2$,

$$\overline{X}_t^i = \left(X_t^i - D_t^i\right)^+ + A_t^i, \tag{3}$$

and $(x)^+ := \max(0, x)$. We assume that the initial queue lengths X_0^1, X_0^2 and the processes $\{A_t^1, t \in \mathbb{Z}_+\}$, $\{A_t^2, t \in \mathbb{Z}_+\}$, $\{D_t^1, t \in \mathbb{Z}_+\}$, $\{D_t^2, t \in \mathbb{Z}_+\}$ are mutually independent and their distributions are known by both controllers C_1 and C_2. Let π_0^1 and π_0^2 be the PMFs on the initial queue lengths X_0^1, X_0^2, respectively. At the beginning of time $t = 0$, C_1 (resp. C_2) knows X_0^1 (resp. X_0^2). Furthermore C_1's (resp. C_2's) knowledge of the queue length X_0^2 (resp. X_0^1) at the other station is described by the PMF π_0^2 (resp. π_0^1). The information of controller $C_i, i = 1, 2$, at the moment it makes the decision $U_t^i, t = 0, 1, \ldots$, is

$$I_t^i := \left\{X_{0:t}^i, A_{0:t}^i, \overline{X}_{0:t}^i, U_{0:t-1}^1, U_{0:t-1}^2, \pi_0^1, \pi_0^2\right\}, i = 1, 2. \tag{4}$$

The controllers' routing decisions/control actions U_t^i are generated according to

$$U_t^i = g_t^i\left(I_t^i\right), i = 1, 2, t \in \mathbb{Z}_+, \tag{5}$$

where

$$\begin{aligned} g_t^i :&(\mathbb{Z}_+)^{t+1} \times \{0,1\}^{t+1} \times (\mathbb{Z}_+)^{t+1} \times \{0,1\}^t \times \\ &\times \{0,1\}^t \times \mathbb{R}^{\mathbb{Z}_+} \times \mathbb{R}^{\mathbb{Z}_+} \mapsto \mathcal{U}_t^i. \end{aligned} \tag{6}$$

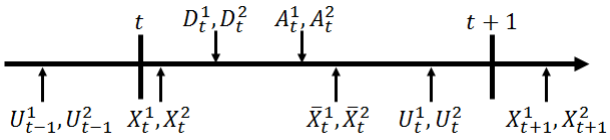

Fig. 2 The order of variables

The control action space $\mathcal{U}_t^i$ at time t depends on $\overline{X}_t^i$. Specifically

$$\mathcal{U}_t^i = \begin{cases} \{0\} & \text{when } \overline{X}_t^i = 0, \\ \{0,1\} & \text{otherwise.} \end{cases} \tag{7}$$

Define $\mathcal{G}_d$ to be the set of feasible decentralized routing policies; that is

$$\begin{aligned} \mathcal{G}_d = \{(g^1, g^2) : \quad & g^i = (g_0^i, g_1^i, \dots, g_t^i, \dots), i = 1, 2 \\ & \text{and } g_t^i \text{ is of form given by (5)-(6)}\}. \end{aligned} \tag{8}$$

We study the operation of the system defined in this section, first over a finite horizon, then over an infinite horizon.

2.1 The finite horizon problem

For the problem with a finite horizon T, we assume the holding cost incurred by the customers present in Q_i at time $t = 0, 1, \dots, T-1$ is $c_t(X_t^i), i = 1, 2$, where $c_t(\cdot)$ is a convex and increasing function. Then, the objective is to determine decentralized routing policies $g \in \mathcal{G}_d$ so as to minimize

$$J_T^g(\pi_0^1, \pi_0^2) := \mathbf{E}\left[\sum_{t=0}^{T-1} \left(c_t\left(X_t^{1,g}\right) + c_t\left(X_t^{2,g}\right)\right) \middle| \pi_0^1, \pi_0^2\right] \text{ [1]} \tag{9}$$

for any PMFs π_0^1, π_0^2 on the initial queue lengths.

2.2 The infinite horizon average cost per unit time problem

For the infinite horizon average cost per unit time problem, we assume the holding cost incurred by the customers present in Q_i at each time is a convex and increasing function $c_t(\cdot) := c(\cdot), i = 1, 2$. Then, the objective is to

[1] The expectation in all equations appearing in this paper is with respect to the probability measure induced by the policy $g \in \mathcal{G}_d$.

determine decentralized routing policies $g = (g^1, g^2) \in \mathcal{G}_d$ so as to minimize

$$\begin{aligned} & J^g(\pi_0^1, \pi_0^2) \\ :=& \limsup_{T\to\infty} \frac{1}{T} J_T^g(\pi_0^1, \pi_0^2) \\ =& \limsup_{T\to\infty} \frac{1}{T} \mathbf{E}\left[\sum_{t=0}^{T-1} \left(c\left(X_t^{1,g}\right) + c\left(X_t^{2,g}\right)\right) \middle| \pi_0^1, \pi_0^2\right] \end{aligned} \tag{10}$$

for any PMFs π_0^1, π_0^2 on the initial queue lengths.

3 Qualitative Properties of Optimal Policies

In this section we present a qualitative property of an optimal routing policy for both the finite horizon and the infinite horizon problem. For that matter we first introduce the following notation.

We denote by Π_t^1 and Π_t^2 the PMFs on X_t^1 and X_t^2, respectively, conditional on all previous decisions $\{U_{0:t-1}^1, U_{0:t-1}^2\}$. $\Pi_t^i, i = 1, 2$ is defined by

$$\Pi_t^i(x) := \mathbf{P}\left(X_t^i = x | U_{0:t-1}^1, U_{0:t-1}^2\right), x \in \mathbb{Z}_+. \tag{11}$$

Similarly, we define the conditional PMFs $\overline{\Pi}_t^1$, $\overline{\Pi}_t^2$ on $\overline{X}_t^1$ and $\overline{X}_t^2$, respectively, as follows.

$$\overline{\Pi}_t^i(x) := \mathbf{P}\left(\overline{X}_t^i = x | U_{0:t-1}^1, U_{0:t-1}^2\right), i = 1, 2, x \in \mathbb{Z}_+. \tag{12}$$

Note that for any policy $g \in \mathcal{G}_d$ all the above defined PMFs are functions of $\{U_{0:t-1}^1, U_{0:t-1}^2\}$. Since both controllers C_1 and C_2 know $\{U_{0:t-1}^1, U_{0:t-1}^2\}$ at time t, the PMFs defined by (11)-(12) are common knowledge Aumann (1976) between the controllers.

We take $\overline{X}_t^i, i = 1, 2$, to be station Q_i's state at time t. Combining (1)-(3) we obtain, for $i = 1, 2$,

$$\begin{aligned} \overline{X}_{t+1}^i =& \left(\overline{X}_t^i - U_t^i + U_t^j - D_{t+1}^i\right)^+ + A_{t+1}^i \\ :=& f_t^i\left(\overline{X}_t^i, U_t^i, U_t^j, W_t^i\right), \end{aligned} \tag{13}$$

where the random variables $W_t^i := (A_{t+1}^i, D_{t+1}^i), i = 1, 2, t = 0, 1, \ldots$ are mutually independent.

The holding cost at time $t, t = 0, 1, \ldots$ can be written as

$$\begin{aligned} & \rho_t\left(\overline{X}_t^1, \overline{X}_t^2, U_t^1, U_t^2\right) \\ :=& c_{t+1}\left(\overline{X}_t^1 - U_t^1 + U_t^2\right) + c_{t+1}\left(\overline{X}_t^2 - U_t^2 + U_t^1\right) \\ =& c_{t+1}\left(X_{t+1}^1\right) + c_{t+1}\left(X_{t+1}^2\right). \end{aligned} \tag{14}$$

Note that for any time horizon T the total expected holding cost due to (14) is equivalent to the total expected holding cost defined by (9) since for any policy $g \in \mathcal{G}_d$

$$\begin{aligned}
&J_T^g(\pi_0^1, \pi_0^2) \\
=&\mathbf{E}\left[\sum_{t=0}^{T-1}\left(c_t\left(X_t^{1,g}\right)+c_t\left(X_t^{2,g}\right)\right)\right] \\
=&\mathbf{E}\left[\sum_{t=0}^{T-2}\left(c_{t+1}\left(X_{t+1}^{1,g}\right)+c_{t+1}\left(X_{t+1}^{2,g}\right)\right)\right] \\
&+\mathbf{E}\left[c_0\left(X_0^1\right)+c_0\left(X_0^2\right)\right] \\
=&\mathbf{E}\left[\sum_{t=0}^{T-2}\rho_t\left(\overline{X}_t^{1,g}, \overline{X}_t^{2,g}, U_t^1, U_t^2\right)\right]+\mathbf{E}\left[c_0\left(X_0^1\right)+c_0\left(X_0^2\right)\right]. \qquad (15)
\end{aligned}$$

With the above notation and definition of system state and instantaneous holding cost, we have a dynamic team problem with non-classical information structure where the common information between the two controllers at any time t is their decisions/control actions up to time $t-1$. This information structure is the control sharing information structure investigated in Mahajan (2013). Furthermore, the independent assumption we made about the exogenous arrivals and the service processes is the same as the assumptions made about the noise variables in Mahajan (2013). Therefore, the following Properties 1-3 hold by the results in Mahajan (2013).

Property 1 *For each t, and any given $g_s^1(.), g_s^2(.), s \leq t$, we have*

$$\begin{aligned}
&\mathbf{P}\left(I_t^1 = i_t^1, I_t^2 = i_t^2 | U_{0:t-1}^1, U_{0:t-1}^2\right) \\
=&\mathbf{P}\left(I_t^1 = i_t^1 | U_{0:t-1}^1, U_{0:t-1}^2\right)\mathbf{P}\left(I_t^2 = i_t^2 | U_{0:t-1}^1, U_{0:t-1}^2\right). \qquad (16)
\end{aligned}$$

Proof Same as that of Proposition 2 in Mahajan (2013).

Property 1 says that the two subsystems are independent conditional on past control actions.

Because of Property 1 and (13), each controller $C_i, i = 1, 2$ can generate its decision at any time t by using only its current local state $\overline{X}_t^i$ and past decisions of both controllers. This assertion is established by the following property.

Property 2 *For the routing problems formulated in Section 2, without loss of optimality we can restrict attention to routing policies of the form*

$$U_t^1 = g_t^1\left(\overline{X}_t^1, U_{0:t-1}^1, U_{0:t-1}^2\right), \qquad (17)$$

$$U_t^2 = g_t^2\left(\overline{X}_t^2, U_{0:t-1}^1, U_{0:t-1}^2\right). \qquad (18)$$

Proof Same as that of Proposition 1 in Mahajan (2013).

Using the common information approach in Nayyar et al (2013), we can refine the result of Property 2 as follows.

Property 3 *For the two routing problems formulated in Section 2, without loss of optimality we can restrict attention to routing policies of the form*

$$U_t^1 = g_t^1\left(\overline{X}_t^1, \overline{\Pi}_t^1, \overline{\Pi}_t^2\right), \tag{19}$$

$$U_t^2 = g_t^2\left(\overline{X}_t^1, \overline{\Pi}_t^1, \overline{\Pi}_t^2\right). \tag{20}$$

Proof Same as that of Theorem 1 in Mahajan (2013).

The result of Property 3 will play a central role in the analysis of the decentralized routing problems formulated in this paper.

4 The Decentralized Policy $\hat{g}$ and Preliminary results

In this section, we specify a decentralized policy $\hat{g}$ and identify an information state for each controller. Furthermore, we develop some preliminary results for both the finite horizon problem and the infinite horizon problem formulated in Section 2.

To specify policy $\hat{g}$, we first define the upper bound and lower bound on the support of the PMF, $\Pi_t^i, i = 1, 2$ as

$$UB_t^i := \max(x : \Pi_t^i(x) \neq 0), \tag{21}$$

$$LB_t^i := \min(x : \Pi_t^i(x) \neq 0). \tag{22}$$

$$UB_t := \max(UB_t^1, UB_t^2), \tag{23}$$

$$LB_t := \min(LB_t^1, LB_t^2). \tag{24}$$

Similarly, we define the bounds on the support of the PMF, $\overline{\Pi}_t^i, i = 1, 2$ as

$$\overline{UB}_t^i := \max(x : \overline{\Pi}_t^i(x) \neq 0), \tag{25}$$

$$\overline{LB}_t^i := \min(x : \overline{\Pi}_t^i(x) \neq 0), \tag{26}$$

$$\overline{UB}_t := \max(\overline{UB}_t^1, \overline{UB}_t^2), \tag{27}$$

$$\overline{LB}_t := \min(\overline{UB}_t^1, \overline{UB}_t^2). \tag{28}$$

Using the above bounds, we specify the policy $\hat{g} := (\hat{g}^1, \hat{g}^2)$ as follows:

$$U_t^i = \hat{g}_t^i\left(\overline{X}_t^i, \overline{UB}_t, \overline{LB}_t\right) = \begin{cases} 1, \text{ when } \overline{X}_t^i \geq TH_t, \\ 0, \text{ when } \overline{X}_t^i < TH_t, \end{cases} \tag{29}$$

where

$$TH_t = \frac{1}{2}\left(\overline{UB}_t + \overline{LB}_t\right). \tag{30}$$

Under $\hat{g}$, each controller routes a customer to the other queue when $\overline{X}_t^i, i = 1, 2$, the queue length of its own station at the time of decision, is greater than or equal to the threshold given by (30).

Note that this decentralized routing policy $\hat{g}$ is indeed of the form asserted by Property 3 since the upper and lower bounds $\overline{UB}_t$ and $\overline{LB}_t$ are both functions of the PMFs $\overline{\Pi}_t^1, \overline{\Pi}_t^2$. Therefore, the threshold TH_t, as a function of $\overline{\Pi}_t^1, \overline{\Pi}_t^2$, is common knowledge between the controllers. Using the common information, each controller can compute the threshold according to (30) individually, and $\hat{g}$ can be implemented in a decentralized manner.

Under policy $\hat{g}$, the evolution of the bounds defined by (23)-(28) are determined by the following lemma.

Lemma 1 *At any time t we have*

$$\overline{UB}_t^{\hat{g}} = UB_t^{\hat{g}} + 1, \quad \overline{LB}_t^{\hat{g}} = \left(LB_t^{\hat{g}} - 1\right)^+. \tag{31}$$

When $(U_t^{1,\hat{g}}, U_t^{2,\hat{g}}) = (0, 0)$

$$UB_{t+1}^{\hat{g}} = \lceil TH_t \rceil - 1, \quad LB_{t+1}^{\hat{g}} = \overline{LB}_t^{\hat{g}} \tag{32}$$

When $(U_t^{1,\hat{g}}, U_t^{2,\hat{g}}) = (1, 1)$

$$UB_{t+1}^{\hat{g}} = \overline{UB}_t^{\hat{g}}, \quad LB_{t+1}^{\hat{g}} = \lceil TH_t \rceil \tag{33}$$

When $(U_t^{i,\hat{g}}, U_t^{j,\hat{g}}) = (1, 0), i = 1, 2, j \neq i$

$$UB_{t+1}^{\hat{g}} = \max\left(\overline{UB}_t^{i,\hat{g}} - 1, \lceil TH_t \rceil\right) \tag{34}$$

$$LB_{t+1}^{\hat{g}} = \min\left(\overline{LB}_t^{j,\hat{g}} + 1, \lceil TH_t \rceil - 1\right) \tag{35}$$

where $\lfloor x \rfloor$ = maximum integer $\leq x$, and $\lceil x \rceil$ = minimum integer $\geq x$.

□

Proof See Appendix A

Corollary 1 below follows directly form (31)-(35) in Lemma 1.

Corollary 1 *Under policy $\hat{g}$,*

$$\begin{aligned} &UB_{t+1}^{\hat{g}} - LB_{t+1}^{\hat{g}} \\ &\leq \begin{cases} \left\lceil \frac{1}{2}\left(UB_t^{\hat{g}} - LB_t^{\hat{g}}\right)\right\rceil & \text{when } (U_t^{1,\hat{g}}, U_t^{2,\hat{g}}) = (0, 0), \\ UB_t^{\hat{g}} - LB_t^{\hat{g}} & \text{otherwise.} \end{cases} \end{aligned} \tag{36}$$

Moreover, if $UB_{t_0}^{\hat{g}} - LB_{t_0}^{\hat{g}} \leq 1$ for some time t_0, then

$$\left(UB_t^{\hat{g}} - LB_t^{\hat{g}}\right) \leq 1 \text{ for all } t \geq t_0. \tag{37}$$

□

Corollary 1 shows that the difference between the highest possible number of customers in Q_1 or Q_2 and the lowest possible number of customers in Q_1 or Q_2 is non-increasing under the policy $\hat{g}$. Furthermore, the difference is reduced by half when there is no customer routed from one queue to another one.

5 The finite horizon problem

In this section, we consider the finite horizon problem formulated in Section 2.1, under the additional condition $X_0^1 = X_0^2 = x_0$, where x_0 is arbitrary but fixed, and is common knowledge between C_1 and C_2.

5.1 Analysis

The main result of this section asserts that the policy $\hat{g}$ defined in Section 4 is optimal.

Theorem 1 *When $X_0^1 = X_0^2 = x_0$ and x_0 is common knowledge between C_1 and C_2, the policy $\hat{g}$ given by (29)-(30) is optimal for the finite horizon decentralized routing problem formulated in Section 2.1, that is*

$$J_T^{\hat{g}}(x_0, x_0) \leq J_T^{g}(x_0, x_0) \tag{38}$$

for any feasible policy $g \in \mathcal{G}_d$ and any initial queue length x_0.

□

Before proving Theorem 1, we note that when $X_0^1 = X_0^2 = x_0$ Corollary 1 implies that

$$UB_t^{\hat{g}} - LB_t^{\hat{g}} \leq 1 \text{ for all } t \geq 0. \tag{39}$$

Equation (39) says that the difference between the highest possible number of customers in Q_1 or Q_2 and the lowest possible number of customers in Q_1 or Q_2 is less than or equal to 1 under policy $\hat{g}$. This property means that $\hat{g}$ controls the length of the joint support of the PMFs $\overline{\Pi}_t^1, \overline{\Pi}_t^2$ and balances the lengths of the two queues. A direct consequence of (39) is the following corollary.

Corollary 2 *At any time t, we have*

$$\left\lfloor \frac{1}{2}(X_t^{1,\hat{g}} + X_t^{2,\hat{g}}) \right\rfloor = \min\left(X_t^{1,\hat{g}}, X_t^{2,\hat{g}}\right), \tag{40}$$

$$\left\lceil \frac{1}{2}(X_t^{1,\hat{g}} + X_t^{2,\hat{g}}) \right\rceil = \max\left(X_t^{1,\hat{g}}, X_t^{2,\hat{g}}\right). \tag{41}$$

□

As pointed out above, the policy $\hat{g}$ balances the lengths of the two queues. This balancing property suggests that the throughput of the system due to $\hat{g}$ is high and the total number of customers in the system is low. This is established by the following lemma.

Lemma 2 *Under the assumption $X_0^1 = X_0^2 = x_0$, where x_0 is common knowledge, for any policy g of the form described by (19)-(20), we have*

$$X_t^{1,\hat{g}} + X_t^{2,\hat{g}} \leq_{st} X_t^{1,g} + X_t^{2,g}, \tag{42}$$

where $Z_1 \leq_{st} Z_2$ means that the r.v. Z_1 is stochastically smaller than the r.v. Z_2, that is, for any $a \in \mathbb{R}$, $\mathbf{P}(Z_1 \geq a) \leq \mathbf{P}(Z_2 \geq a)$ (see Marshall et al (2010)).

□

Proof See Appendix B

Using Lemma 2, we now prove Theorem 1.

Proof (Proof of Theorem 1) For any feasible policy g, since the functions $c_t, t = 0, 1, ..., T$, are convex, we have at any time t

$$\begin{aligned}
&\mathbf{E}\left[c_t\left(X_t^{1,g}\right) + c_t\left(X_t^{2,g}\right)\right] \\
\geq &\mathbf{E}\left[c_t\left(\left\lfloor\frac{1}{2}(X_t^{1,g} + X_t^{2,g})\right\rfloor\right) + c_t\left(\left\lceil\frac{1}{2}(X_t^{1,g} + X_t^{2,g})\right\rceil\right)\right].
\end{aligned} \tag{43}$$

Furthermore, using Lemma 2 and the fact that $c_t(\cdot)$ is increasing, we get

$$\begin{aligned}
&\mathbf{E}\left[c_t\left(\left\lfloor\frac{1}{2}(X_t^{1,g} + X_t^{2,g})\right\rfloor\right) + c_t\left(\left\lceil\frac{1}{2}(X_t^{1,g} + X_t^{2,g})\right\rceil\right)\right] \\
\geq &\mathbf{E}\left[c_t\left(\left\lfloor\frac{1}{2}(X_t^{1,\hat{g}} + X_t^{2,\hat{g}})\right\rfloor\right) + c_t\left(\left\lceil\frac{1}{2}(X_t^{1,\hat{g}} + X_t^{2,\hat{g}})\right\rceil\right)\right] \\
= &\mathbf{E}\left[c_t\left(\min(X_t^{1,\hat{g}}, X_t^{2,\hat{g}})\right) + c_t\left(\max(X_t^{1,\hat{g}}, X_t^{2,\hat{g}})\right)\right] \\
= &\mathbf{E}\left[c_t\left(X_t^{1,\hat{g}}\right) + c_t\left(X_t^{2,\hat{g}}\right)\right].
\end{aligned} \tag{44}$$

The inequality in (44) is true because $X_t^{1,g} + X_t^{2,g} \leq_{st} X_t^{1,\hat{g}} + X_t^{2,\hat{g}}$ (Lemma 2) and $c_t(\cdot)$ is increasing. The first equality in (44) follows from Corollary 2. Combining (43) and (44) we obtain, for any t,

$$\mathbf{E}\left[c_t\left(X_t^{1,g}\right) + c_t\left(X_t^{2,g}\right)\right] \geq \mathbf{E}\left[c_t\left(X_t^{1,\hat{g}}\right) + c_t\left(X_t^{2,\hat{g}}\right)\right]. \tag{45}$$

The optimality of policy $\hat{g}$ follows from (9) and (45).

5.2 Comparison to the performance under centralized information

We compare now the performance of the optimal decentralized policy $\hat{g}$ to the performance of the queueing system under centralized information. The results of this comparison will be useful when we study the infinite horizon problem in Section 6.

Consider a centralized controller who has all the information I_t^1 and I_t^2 at each time t. Then, the set $\mathcal{G}_c$ of feasible routing policies of the centralized controller is

$$\mathcal{G}_c := \{(g^1, g^2) : \quad g^i = (g_0^i, g_1^i, \dots, g_t^i, \dots), i = 1, 2 \\ \text{and } U_t^i = g_t^i(I_t^1, I_t^2)\}. \tag{46}$$

By the definition, $\mathcal{G}_d \subset \mathcal{G}_c$. This means that the centralized controller can simulate any decentralized policy $g \in \mathcal{G}_d$ adopted by controllers C_1 and C_2. Therefore, for any initial PMFs π_0^1, π_0^2

$$\inf_{g \in \mathcal{G}_c} J_T^g(\pi_0^1, \pi_0^2) \leq \inf_{g \in \mathcal{G}_d} J_T^g(\pi_0^1, \pi_0^2) \tag{47}$$

$$\inf_{g \in \mathcal{G}_c} J^g(\pi_0^1, \pi_0^2) \leq \inf_{g \in \mathcal{G}_d} J^g(\pi_0^1, \pi_0^2). \tag{48}$$

When $X_0^1 = X_0^2 = x_0$, Lemma 2 and Theorem 1 show that the cost given by $\hat{g}$ is smaller than the cost given by any policy $g \in \mathcal{G}_d$. Furthermore we have:

Lemma 3 *Under the assumption $X_0^1 = X_0^2 = x_0$, where x_0 is common knowledge, we have*

$$X_t^{1,\hat{g}} + X_t^{2,\hat{g}} \leq_{st} X_t^{1,g} + X_t^{2,g}, \tag{49}$$

for any $g \in \mathcal{G}_c$, and

$$J_T^{\hat{g}}(x_0, x_0) \leq \inf_{g \in \mathcal{G}_c} J_T^g(x_0, x_0). \tag{50}$$

for any $g \in \mathcal{G}_c$.

□

Proof The proof of (49) is the same as the proof of Lemma 2, and the proof of (50) is the same as the proof of Theorem 1.

Since $\hat{g}$ is a decentralized policy, (47) and Lemma 3 imply that

$$J_T^{\hat{g}}(x_0, x_0) = \inf_{g \in \mathcal{G}_d} J_T^g(x_0, x_0) = \inf_{g \in \mathcal{G}_c} J_T^g(x_0, x_0). \tag{51}$$

Equation (51) shows that when $X_0^1 = X_0^2 = x_0$ and x_0 is common knowledge between $C_1$and C_2, policy $\hat{g}$ achieves the same performance as any centralized optimal policy.

5.3 The Case of Different Initial Queue Lengths

When $X_0^1 \neq X_0^2$, the policy $\hat{g}$ is not necessarily optimal for the finite horizon problem.

Consider an example where the horizon $T = 1$ (two-step horizon), $\lambda = 0.1, \mu = 0.5$ and

$$\mathbf{P}\left(X_0^1 = 3\right) = 1, \tag{52}$$

$$\mathbf{P}\left(X_0^2 = 1\right) = 0.9, \quad \mathbf{P}\left(X_0^2 = 5\right) = 0.1, \tag{53}$$

that is,

$$\pi_0^1 = (0, 0, 0, 1, 0, 0, 0, \dots), \tag{54}$$

$$\pi_0^2 = (0, 0.9, 0, 0, 0, 0.1, 0, \dots), \tag{55}$$

where π_0^1, π_0^2 denote the initial PMFs on the lengths of the queues.
Then, $\overline{\Pi}_0^1, \overline{\Pi}_0^2$ and the threshold TH_0 are

$$\overline{\Pi}_0^1 = (0, 0, 0.5, 0.4, 0.1, 0, 0, \dots), \tag{56}$$

$$\overline{\Pi}_0^2 = (0.45, 0.36, 0.09, 0, 0.05, 0.04, 0.01, \dots), \tag{57}$$

$$TH_0 = \frac{1}{2}(6 + 0) = 3. \tag{58}$$

Consider the cost functions $c_0(x) = 0$ and $c_1(x) = x^2$. Then, we have

$$\begin{aligned} & J^g(\pi_0^1, \pi_0^2) \\ =& \mathbf{E}\left[\left(X_1^{1,g}\right)^2 + \left(X_1^{2,g}\right)^2\right] \\ =& \mathbf{E}\left[\left(\overline{X}_0^1 - U_0^{1,g} + U_0^{2,g}\right)^2 + \left(\overline{X}_0^2 - U_0^{2,g} + U_0^{1,g}\right)^2\right]. \end{aligned} \tag{59}$$

Using (56)-(58) and the specification of the policy $\hat{g}$, we can compute the expected cost due to $\hat{g}$. It is

$$J^{\hat{g}}(\pi_0^1, \pi_0^2) = 8.48. \tag{60}$$

Consider now another policy $\tilde{g}$ described below. For $i = 1, 2, i \neq j$,

$$U_t^{i,\tilde{g}} = \tilde{g}_t\left(\overline{X}_t^i, \overline{\Pi}_t^1, \overline{\Pi}_t^2\right) = \begin{cases} 1, \text{ when } \overline{X}_t^i \geq \mathbf{E}\left[\overline{X}_t^j | \overline{\Pi}_t^j\right], \\ 0, \text{ when } \overline{X}_t^i < \mathbf{E}\left[\overline{X}_t^j | \overline{\Pi}_t^j\right], \end{cases} \tag{61}$$

Then, from (56)-(57) and (61) we get

$$U_0^{1,\tilde{g}} = \begin{cases} 1, \text{ when } \overline{X}_0^1 \geq 1, \\ 0, \text{ when } \overline{X}_0^1 < 1, \end{cases} \tag{62}$$

$$U_0^{2,\tilde{g}} = \begin{cases} 1, \text{ when } \overline{X}_0^2 \geq 2.6, \\ 0, \text{ when } \overline{X}_0^2 < 2.6, \end{cases} \tag{63}$$

Therefore, the expected cost due to the policy $\tilde{g}$ is given by

$$J^{\tilde{g}}(\pi_0^1, \pi_0^2) = 8.28 \tag{64}$$

Since $J^{\tilde{g}}(\pi_0^1, \pi_0^2) = 8.28 < 8.48 = J^{\hat{g}}(\pi_0^1, \pi_0^2)$, policy $\hat{g}$ is not optimal.

In this example, each controller has only one decision to make, the decision at time 0. As a result, signaling does not provide any advantages to the controllers, and that is why the policy $\hat{g}$ is not the best policy.

6 Infinite horizon

We consider the infinite horizon decentralized routing problem formulated in Section 2.2, and make the following additional assumptions.

Assumption 1 $\mu > \lambda$.

Assumption 2 *The initial PMFs π_0^1, π_0^2 are finitely supported and common knowledge between controllers C_1 and C_2. i.e. there exists $M < \infty$ such that $\pi_0^1(x) = \pi_0^2(x) = 0$ for all $x > M$.*

Let g_0 denote the open-loop policy that does not do any routing, that is, at any time t

$$U_t^{1,g_0} = U_t^{2,g_0} = 0. \tag{65}$$

Assumption 3

$$\lim_{T\to\infty} \frac{1}{T} J_T^{g_0}(\pi_0^1, \pi_0^2) := J^{g_0} < \infty \quad a.s., \tag{66}$$

where J^{g_0} is a constant that denotes the infinite horizon average cost per unit time due to policy g_0.

Remark 1 Due to policy g_0, the queue length $\{X_t^{g_0,i}, t \in \mathbb{Z}_+\}, i = 1, 2$ is a positive recurrent birth and death chain with arrival rate λ and departure rate $\mu 1_{\{X_t^{g_0,i} \neq 0\}}$. Therefore, as $T \to \infty$, the average cost per unit time converges to a constant a.s. if the expected cost under the stationary distribution of the process is finite (see (Bremaud, 1999, chap. 3)). Assumption 3 is equivalent to the assumption that the expected cost is finite under the stationary distribution of the controlled queue lengths.

We proceed to analyze the infinite horizon average cost per unit time for the model of Section 2 under Assumptions 1-3.

6.1 Analysis

When $X_0^1 \neq X_0^2$, the policy $\hat{g}$, defined in Section 4, is not necessarily optimal for the finite horizon problem (see the example in Section 5.3). Nevertheless, the policy $\hat{g}$ still attempts to balance the queues. Given enough time, policy $\hat{g}$ may be able to balance the queue lengths even if they are not initially balanced. In this section we show that this is indeed the case.

Specifically, we prove the optimality of policy $\hat{g}$ for the infinite horizon average cost per unit time problem, as stated in the following theorem which is the main result of this section.

Theorem 2 *Under Assumptions 1-3, the policy $\hat{g}$, described by (29)-(30), is optimal for the infinite horizon average cost per unit time problem formulated in Section 2.2.*

□

To establish the assertion of Theorem 2 we proceed in four steps. In the first step we show that the infinite horizon average cost per unit time due to policy $\hat{g}$ is bounded above by the cost of the uncontrolled queues (i.e. the cost due to policy g_0). In the second step we show that under policy $\hat{g}$ the queues are eventually balanced, i.e. the queue lengths can differ by at most one. In the third step we derive a result that connects the performance of policy $\hat{g}$ under the initial PMFs $(0,0)$ to the performance of the optimal policy under any arbitrary initial PMFs π_0^1, π_0^2 on queues Q_1 and Q_2. In the forth step we establish the optimality of policy $\hat{g}$ based on the results of steps one, two and three.

Step 1

We prove that $J^{\hat{g}}(\pi_0^1, \pi_0^2) \leq J^{g_0}$. To do this, we first establish some preliminary results that appear in Lemmas 4 and 5.

Lemma 4 *There exists processes $\{Y_t^1, t \in \mathbb{Z}_+\}$ and $\{Y_t^2, t \in \mathbb{Z}_+\}$ such that*

$$\{Y_t^i, t \in \mathbb{Z}_+\} \text{ has the same distribution as } \{X_t^{i,g_0}, t \in \mathbb{Z}_+\} \tag{67}$$

for $i = 1, 2$, and for all times t

$$X_t^{1,\hat{g}} + X_t^{2,\hat{g}} \leq Y_t^1 + Y_t^2 \quad a.s., \tag{68}$$

$$\max_i \left(X_t^{i,\hat{g}}\right) \leq \max_i \left(Y_t^i\right) \quad a.s. \tag{69}$$

□

Proof See Appendix C

Lemma 4 means that the uncontrolled queue lengths are longer than the queue lengths under policy $\hat{g}$ in a stochastic sense. Note that (68) and (69) are not true if $Y_t^i, i = 1, 2$, is replaced by $X_t^{i,g_0}, i = 1, 2$, as the following example shows.

Example

When $X_t^{1,g_0} = 4, X_t^{2,g_0} = 6$ and $X_t^{1,\hat{g}} = X_t^{2,\hat{g}} = 5$, the analogues of (68) and (69) where Y_t^i are replaced by $X_t^{i,g_0}, i = 1, 2$ are

$$X_t^{1,\hat{g}} + X_t^{2,\hat{g}} = X_t^{1,g_0} + X_t^{2,g_0} = 10, \tag{70}$$

$$\max_i \left(X_t^{i,\hat{g}} \right) = 5 \leq 6 = \max_i \left(X_t^{i,g_0} \right). \tag{71}$$

However, if $A_{t+1}^1 = 1, A_{t+1}^2 = 0$ and $D_{t+1}^1 = 0, D_{t+1}^2 = 1$ we get $X_{t+1}^{1,g_0} = X_{t+1}^{2,g_0} = 5$ and $X_{t+1}^{1,\hat{g}} = 6, X_{t+1}^{2,\hat{g}} = 4$, then

$$\max_i \left(X_{t+1}^{i,\hat{g}} \right) = 6 > 5 = \max_i \left(X_{t+1}^{i,g_0} \right), \tag{72}$$

and the analogue of (69), when Y_t^i is replaced by $X_t^{i,g_0}, i = 1, 2$, does not hold.

The stochastic dominance relation asserted by Lemma 4 implies that the instantaneous cost under policy $\hat{g}$ is almost surely no greater than the instantaneous cost due to policy g_0. This implication is made precise by the following lemma.

Lemma 5 *The processes $\{Y_t^1, t \in \mathbb{Z}_+\}$ and $\{Y_t^2, t \in \mathbb{Z}_+\}$ defined in Lemma 4 are such that at any time t*

$$c\left(X_t^{1,\hat{g}} \right) + c\left(X_t^{2,\hat{g}} \right) \leq c\left(Y_t^1 \right) + c\left(Y_t^2 \right) \quad a.s. \tag{73}$$

□

Proof See Appendix C

In order to apply the result of Step 1 as the time horizon goes to infinity, we need the following result on the convergence of the cost due to $\{Y_t^1, t \in \mathbb{Z}_+\}$ and $\{Y_t^2, t \in \mathbb{Z}_+\}$.

Lemma 6 *Let $\{Y_t^1, t \in \mathbb{Z}_+\}$ and $\{Y_t^2, t \in \mathbb{Z}_+\}$ be the processes defined in Lemma 4. Let W_T denote*

$$W_T := \frac{1}{T} \sum_{t=0}^{T-1} \left(c(Y_t^1) + c(Y_t^2) \right). \tag{74}$$

Under Assumptions 2 and 3,

$$\lim_{T \to \infty} W_T = J^{g_0} \quad a.s. \tag{75}$$

Moreover, $\{W_T, T = 1, 2, \dots\}$ is uniformly integrable, so it also converges in expectation.

□

Proof See Appendix C

A direct consequence of Lemmas 4, 5 and 6 is the following.

Corollary 3 *If* $\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \left(c\left(X^{1,\hat{g}}\right) + c\left(X^{2,\hat{g}}\right)\right)$ *converges a.s., then,*

$$\frac{1}{T} \sum_{t=0}^{T-1} \left(c\left(X^{1,\hat{g}}\right) + c\left(X^{2,\hat{g}}\right)\right) \longrightarrow J^{\hat{g}}(\pi_0^1, \pi_0^2) \tag{76}$$

in expectation and a.s. as $T \to \infty$*. Furthermore,*

$$J^{\hat{g}}(\pi_0^1, \pi_0^2) \leq J^{g_0} < \infty. \tag{77}$$

□

Proof See Appendix C

Step 2

We prove that under policy $\hat{g}$ the queues are eventually balanced. For this matter we first establish some preliminary results that appear in Lemmas 7 and 8.

Lemma 7 *Let* T_0 *be a stopping time with respect to the process* $\{X_t^{1,\hat{g}}, X_t^{2,\hat{g}}, t \in \mathbb{Z}_+\}$*. Define the process* $\{S_t = S_t^{\hat{g}}, t \geq T_0 + 1\}$ *as follows.*

$$\begin{aligned} S_{T_0+1} :=& X_{T_0+1}^{1,\hat{g}} + X_{T_0+1}^{2,\hat{g}} && (78)\\ S_{t+1} :=& S_t - D_t^1 - D_t^2 + A_t^1 + A_t^2 \\ &+ 1_{\{S_t=1\}} \left(1_{\left\{X_t^{1,\hat{g}}=0\right\}}(D_t^1 - D_t^2) + D_t^2\right) \\ &+ 1_{\{S_t=0\}} \left(D_t^1 + D_t^2\right) && (79) \end{aligned}$$

If $\mu > \lambda > 0$*, then* $\{S_t, t \geq T_0 + 1\}$ *is an irreducible positive recurrent Markov chain.*

□

Proof See Appendix D

Lemma 7 holds for arbitrary stopping time T_0 with respect to $\{X_t^{1,\hat{g}}, X_t^{2,\hat{g}}, t \in \mathbb{Z}_+\}$. By appropriately selecting T_0 we will show later that S_t is coupled with $X_t^{1,\hat{g}} + X_t^{2,\hat{g}}$, i.e. for all $t > T_0$, $S_t = X_t^{1,\hat{g}} + X_t^{2,\hat{g}}$. This result along with the fact that the process $\{S_t, t \geq T_0 + 1\}$ is an irreducible positive recurrent Markov chain will allow us to analyze the cost due to policy $\hat{g}$.

Lemma 8 *Under policy* $\hat{g}$*,*

$$\mathbf{P}\left(\left(U_t^{1,\hat{g}}, U_t^{2,\hat{g}}\right) = (0,0) \quad i.o.\right) = 1. \tag{80}$$

□

Proof See Appendix D

Lemma 8 means that the event { there exists $t_0 < \infty$ such that at least one of the queue lengths is above the threshold defined by (30) for all $t > t_0$ } can not happen. The idea of Lemma 8 is the following. If one of the queues, say Q_1, has length above the threshold, hence above the lower bound $LB_t^{\hat{g}}$, then, the length of Q_2 does not decrease, because under policy $\hat{g}$ Q_2 receives one customer from Q_1 and has at most one departure at this time. Therefore, both queue lengths at the next time are bounded below by the current lower bound $LB_t^{\hat{g}}$. When at least one of the queue lengths is above the threshold for all $t > t_0$, the queue lengths are bounded below by $LB_{t_0}^{\hat{g}}$ for all $t > t_0$. This kind of lower bound can not exist if the total arrival rate 2λ to the system is less than the total departure rate 2μ from the system.

Lemma 8 and Corollary 1 in Section 4 can be used to establish that under policy $\hat{g}$ the queues are eventually balanced. This is shown in the corollary below.

Corollary 4 *Let*

$$T_0 := \inf\{t : UB_t^{\hat{g}} - LB_t^{\hat{g}} \leq 1\}. \tag{81}$$

Then

$$\mathbf{P}(T_0 < \infty) = 1 \tag{82}$$

and

$$\left(UB_t^{\hat{g}} - LB_t^{\hat{g}}\right) \leq 1 \text{ for all } t \geq T_0. \tag{83}$$

□

Step 3

We compare the finite horizon cost $J_T^{\hat{g}}(0,0)$ (respectively, the infinite horizon cost $J^{\hat{g}}(0,0)$) due to policy $\hat{g}$ under initial PMFs $(0,0)$ to the minimum finite horizon cost $\inf_{g\in\mathcal{G}_d} J_T^g(\pi_0^1,\pi_0^2)$ (respectively, the minimum infinite horizon cost $\inf_{g\in\mathcal{G}_d} J^g(\pi_0^1,\pi_0^2)$) under arbitrary initial PMFs (π_0^1,π_0^2).

Lemma 9 *For any finite time T and any initial PMFs π_0^1,π_0^2.*

$$J_T^{\hat{g}}(0,0) = \inf_{g\in\mathcal{G}_c} J_T^g(0,0) \leq \inf_{g\in\mathcal{G}_c} J_T^g(\pi_0^1,\pi_0^2) \leq \inf_{g\in\mathcal{G}_d} J_T^g(\pi_0^1,\pi_0^2), \tag{84}$$

and

$$J^{\hat{g}}(0,0) = \inf_{g\in\mathcal{G}_c} J^g(0,0) \leq \inf_{g\in\mathcal{G}_c} J^g(\pi_0^1,\pi_0^2) \leq \inf_{g\in\mathcal{G}_d} J^g(\pi_0^1,\pi_0^2). \tag{85}$$

□

Proof See Appendix E.

Lemma 9 states that the minimum cost achieved when the queues are initially empty is smaller than the minimum cost obtained when the system's initial condition is given by arbitrary PMFs on the lengths of queues Q_1 and Q_2. This result is established through the use of the corresponding centralized information system that is discussed in Section 5.2.

Step 4

Based on the results of Steps 1, 2 and 3 we now establish the optimality of policy $\hat{g}$ for the infinite horizon average cost per unit time problem formulated in Section 2.2. First, we outline the key ideas in the proof of Theorem 2, then we present its proof. Step 2 ensures that policy $\hat{g}$ eventually (in finite time) balances the queues. Step 1 ensures that the cost $J^{\hat{g}}(\pi_0^1, \pi_0^2)$ is finite. These two results together imply that the cost due to policy $\hat{g}$ is the same as the cost incurred after the queues are balanced. Furthermore, we show that the cost of policy $\hat{g}$ is independent of the initial PMFs on the queue lengths. Then, the result of Step 3 together with the results on the finite horizon problem establish the optimality of policy $\hat{g}$.

Proof (Proof of Theorem 2) Define T_0 to be the first time when the length of the joint support of PMFs $\Pi_t^{1,\hat{g}}, \Pi_t^{2,\hat{g}}$ is no more than 1. That is

$$T_0 = \inf\{t : UB_t^{\hat{g}} - LB_t^{\hat{g}} \leq 1\}. \tag{86}$$

The random variable T_0 is a stopping time with respect to the process $\{X_t^{1,\hat{g}}, X_t^{2,\hat{g}}, t \in \mathbb{Z}_+\}$. From Corollary 4 we have

$$\mathbf{P}(T_0 < \infty) = 1, \tag{87}$$

$$UB_t^{\hat{g}} - LB_t^{\hat{g}} \leq 1 \text{ for all } t \geq T_0. \tag{88}$$

Furthermore, for all $t \geq T_0$

$$\left| X_t^{1,\hat{g}} - X_t^{2,\hat{g}} \right| \leq UB_t^{\hat{g}} - LB_t^{\hat{g}} \leq 1. \tag{89}$$

Consider the process $\{S_t, t \geq T_0 + 1\}$ defined by (78) and (79) (in Lemma 7). We claim that for all $t \geq T_0 + 1$

$$X_t^{1,\hat{g}} + X_t^{2,\hat{g}} = S_t. \tag{90}$$

We prove the claim in Appendix F. Suppose the claim is true. Since $\left| X_t^{1,\hat{g}} - X_t^{2,\hat{g}} \right| \leq 1$ for all $t \geq T_0 + 1$, the instantaneous cost at time $t \geq T_0 + 1$ is equal to

$$\begin{aligned} & c\left(X_t^{1,\hat{g}}\right) + c\left(X_t^{2,\hat{g}}\right) \\ =& c\left(\left\lfloor \frac{1}{2}(X_t^{1,\hat{g}} + X_t^{2,\hat{g}}) \right\rfloor\right) + c\left(\left\lceil \frac{1}{2}(X_t^{1,\hat{g}} + X_t^{2,\hat{g}}) \right\rceil\right) \\ =& c\left(\left\lfloor \frac{1}{2} S_t^{\hat{g}} \right\rfloor\right) + c\left(\left\lceil \frac{1}{2} S_t^{\hat{g}} \right\rceil\right). \end{aligned} \tag{91}$$

Then, the average cost per unit time due to policy $\hat{g}$ is given by

$$\begin{aligned}
&\frac{1}{T}\sum_{t=0}^{T-1}\left(c\left(X_t^{1,\hat{g}}\right)+c\left(X_t^{2,\hat{g}}\right)\right)\\
=&\frac{1}{T}\sum_{t=0}^{T_0}\left(c\left(X_t^{1,\hat{g}}\right)+c\left(X_t^{2,\hat{g}}\right)\right)\\
&+\frac{1}{T}\sum_{t=T_0+1}^{T-1}\left(c\left(\left\lfloor\frac{1}{2}S_t^{\hat{g}}\right\rfloor\right)+c\left(\left\lceil\frac{1}{2}S_t^{\hat{g}}\right\rceil\right)\right). \qquad (92)
\end{aligned}$$

Since $T_0<\infty\, a.s.$, we obtain

$$\begin{aligned}
&\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\left(c\left(X_t^{1,\hat{g}}\right)+c\left(X_t^{2,\hat{g}}\right)\right)\\
=&\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T_0}\left(c\left(X_t^{1,\hat{g}}\right)+c\left(X_t^{2,\hat{g}}\right)\right)\\
&+\lim_{T\to\infty}\frac{1}{T}\sum_{t=T_0(+1}^{T-1}\left(c\left(\left\lfloor\frac{1}{2}S_t^{\hat{g}}\right\rfloor\right)+c\left(\left\lceil\frac{1}{2}S_t^{\hat{g}}\right\rceil\right)\right)\\
=&\lim_{t\to\infty}\frac{1}{T}\sum_{t=T_0+1}^{T-1}\left(c\left(\left\lfloor\frac{1}{2}S_t^{\hat{g}}\right\rfloor\right)+c\left(\left\lceil\frac{1}{2}S_t^{\hat{g}}\right\rceil\right)\right)\\
=&\sum_{s=0}^{\infty}\pi^{\hat{g}}(s)\left(c\left(\left\lfloor\frac{1}{2}s\right\rfloor\right)+c\left(\left\lceil\frac{1}{2}s\right\rceil\right)\right)a.s. \qquad (93)
\end{aligned}$$

where $\pi^{\hat{g}}(s)$ is the stationary distribution of $\{S_t=S_t^{\hat{g}}, t\geq T_0+1\}$. The second equality in (93) holds because $T_0<\infty\, a.s.$; the last equality in (93) follows by the Ergodic theorem for irreducible positive recurrent Markov chains (Bremaud, 1999, chap. 3).

Since the sum $\frac{1}{T}\sum_{t=0}^{T-1}\left(c\left(X_t^{1,\hat{g}}\right)+c\left(X_t^{2,\hat{g}}\right)\right)$ converges $a.s.$, from Corollary 3 we have

$$\begin{aligned}
J^{\hat{g}}(\pi_0^1,\pi_0^2)&=\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\left(c\left(X^{1,\hat{g}}\right)+c\left(X^{2,\hat{g}}\right)\right)\\
&=\sum_{s=0}^{\infty}\pi^{\hat{g}}(s)\left(c\left(\left\lfloor\frac{1}{2}s\right\rfloor\right)+c\left(\left\lceil\frac{1}{2}s\right\rceil\right)\right). \qquad (94)
\end{aligned}$$

Since the right hand side of equation (94) does not depend on the initial PMFs π_0^1,π_0^2, we obtain

$$J^{\hat{g}}(\pi_0^1,\pi_0^2)=J^{\hat{g}}(0,0). \qquad (95)$$

Combining (95) and Lemma 9 we get

$$J^{\hat{g}}(\pi_0^1, \pi_0^2) = J^{\hat{g}}(0,0) \leq \inf_{g \in \mathcal{G}_d} J^g(\pi_0^1, \pi_0^2). \tag{96}$$

Thus, $\hat{g}$ is an optimal routing policy for the infinite horizon problem.

7 Discussion and Conclusion

Based on the results established in Sections 3-6, we now discuss and answer the questions posed in Section 1.

Controllers C_1 and C_2 communicate with one another through their control actions; thus, each controller's information depends on the decision rule/routing policy of the other controller. Therefore, the queueing system considered in this paper has non-classical information structure Witsenhausen (1971). A key feature of the system's information structure is that at each time instant each controller's information consists of one component that is common knowledge between C_1 and C_2 and another component that is its own private information. The presence of common information allows us to use the common information approach, developed in Nayyar et al (2013), along with specific features of our model to identify an information state/sufficient statistic for the finite and infinite horizon optimization problem. The identification/discovery of an appropriate information state proceeds in two steps: In the first step we use the common information approach (in particular Mahajan (2013)) to identify the general form of an information state (namely $\left(\overline{X}_t^i, \overline{\Pi}_t^1, \overline{\Pi}_t^2\right)$) for controller $C_i, i = 1, 2.$ (and the corresponding structure of an optimal policy, Properties 3). In the second step we take advantage of the features of our system to further refine/simplify the information state; we discover a simpler form of information state, namely, $\left(\overline{X}_t^i, \left\{\overline{UB}_t^j, \overline{LB}_t^j\right\}_{j=1,2}\right)$ for controller $C_i, i = 1, 2.$ The component $\left\{\overline{UB}_t^j, \overline{LB}_t^j\right\}_{j=1,2}$ of the above information state describes the common information between controllers C_1 and C_2 at time $t, t = 1, 2, \ldots.$

Using this common information we established an optimal signaling strategy that is described by the threshold policy $\hat{g}$ specified in Section 4.

The update of $\left\{\overline{UB}_t^j, \overline{LB}_t^j\right\}_{j=1,2}$ is described by (32)-(35) and explicitly depends on the signaling policy $\hat{g}$. Specifically, if a customer is sent from Q_i to Q_j $(i \neq j)$ at time t the lower bound on the queue length of Q_i increases because both controllers know that the length of Q_i is above the threshold TH_t at the time of routing; if no customer is sent from Q_i to Q_j at time t, the upper bound on the length of Q_i decreases because both controllers know that the length of Q_i is below the threshold TH_t at the time of routing. The update of common information incorporates the information about a controller's private information transmitted to the other controller through signaling.

The signaling policy $\hat{g}$ communicates information in such a way that eventually the difference between the upper bound and the lower bound on the

queue lengths is no more than one. Thus, signaling through $\hat{g}$ results in a balanced queueing system.

A Proofs of the Results in Section 4

Proof (Proof of Lemma 1) Since there is one possible arrival to any queue and one possible departure from any queue at each time instant, (31) holds.

When $(U_t^{1,\hat{g}}, U_t^{2,\hat{g}}) = (0,0)$, both $\overline{X}_t^{1,\hat{g}}$ and $\overline{X}_t^{2,\hat{g}}$ are below the threshold and no customers are routed form any queue. Therefore, the upper bound of the queue lengths at $t+1$ is

$$UB_{t+1}^{\hat{g}} = \lceil TH_t \rceil - 1. \tag{97}$$

Moreover, the lower bound of the queue lengths at $t+1$ is the same as the lower bound of $\overline{X}_t^{1,\hat{g}}, \overline{X}_t^{2,\hat{g}}$. That is,

$$LB_{t+1}^{\hat{g}} = \overline{LB}_t^{\hat{g}}. \tag{98}$$

When $(U_t^{1,\hat{g}}, U_t^{2,\hat{g}}) = (1,1)$, both $\overline{X}_t^{1,\hat{g}}$ and $\overline{X}_t^{2,\hat{g}}$ are greater than or equal to the threshold. Since the routing only exchanges two customers between the two queues, the queue lengths remain the same as the queue lengths before routing. As a result, the upper bound and lower bound of the queue lengths at $t+1$ are given by

$$UB_{t+1}^{\hat{g}} = \overline{UB}_t^{\hat{g}}. \tag{99}$$

$$LB_{t+1}^{\hat{g}} = \lceil TH_t \rceil. \tag{100}$$

When $(U_t^{i,\hat{g}}, U_t^{j,\hat{g}}) = (1,0), i \neq j$, $\overline{X}_t^{i,\hat{g}}$ is greater than or equal to the threshold; $\overline{X}_t^{j,\hat{g}}$ is below the threshold. Since one customer is routed from Q_i to Q_j,

$$X_{t+1}^{i,\hat{g}} = \overline{X}_t^{i,\hat{g}} - 1, \tag{101}$$

$$X_{t+1}^{j,\hat{g}} = \overline{X}_t^{j,\hat{g}} + 1. \tag{102}$$

Therefore, the upper bound of the queue lengths at $t+1$ becomes

$$\begin{aligned} UB_{t+1}^{\hat{g}} &= \max\left\{\overline{UB}_t^{i,\hat{g}} - 1, \lceil TH_t \rceil - 1 + 1\right\} \\ &= \max\left\{\overline{UB}_t^{i,\hat{g}} - 1, \lceil TH_t \rceil\right\}, \end{aligned} \tag{103}$$

and lower bound of the queue lengths at $t+1$ is given by

$$LB_{t+1}^{\hat{g}} = \min\left\{\lceil TH_t \rceil - 1, \overline{LB}_t^{j,\hat{g}} + 1\right\}. \tag{104}$$

B Proofs of the Results in Section 5

Proof (Proof of Lemma 2) The proof is done by induction.
At time $t = 0$, $X_0^{1,\hat{g}} + X_0^{2,\hat{g}} = X_0^{1,g} + X_0^{2,g} = x_0$.
Suppose the lemma is true at time t.
At time $t+1$, from the system dynamics (1)-(3) we get, for any g,

$$\begin{aligned} &X_{t+1}^{1,g} + X_{t+1}^{2,g} \\ &= \left(X_t^{1,g} - D_t^1\right)^+ + \left(X_t^{2,g} - D_t^2\right)^+ + A_t^1 + A_t^2. \end{aligned} \tag{105}$$

Therefore, it suffices to show that

$$\left(X_t^{1,\hat{g}} - D_t^1\right)^+ + \left(X_t^{2,\hat{g}} - D_t^2\right)^+ \leq_{st} \left(X_t^{1,g} - D_t^1\right)^+ + \left(X_t^{2,g} - D_t^2\right)^+. \tag{106}$$

Consider any realization $(X_t^{1,g}, X_t^{2,g}) = (x^1, x^2)$.
If $x^1, x^2 > 0$, then $\left\lfloor \frac{1}{2}(x^1+x^2) \right\rfloor, \left\lceil \frac{1}{2}(x^1+x^2) \right\rceil > 0$. Therefore,

$$\begin{aligned}
&\left(X_t^{1,g} - D_t^1\right)^+ + \left(X_t^{2,g} - D_t^2\right)^+ \\
=&x^1 + x^2 - D_t^1 - D_t^2 \\
=&\left(\left\lfloor \frac{1}{2}(x^1+x^2) \right\rfloor - D_t^1\right)^+ + \left(\left\lceil \frac{1}{2}(x^1+x^2) \right\rceil - D_t^2\right)^+.
\end{aligned} \tag{107}$$

If $x^i = 0$ and $x^j \geq 2$ $(i \neq j)$, then $\left\lfloor \frac{1}{2}(x^1+x^2) \right\rfloor > 0$ and $\left\lceil \frac{1}{2}(x^1+x^2) \right\rceil > 0$. Therefore,

$$\begin{aligned}
&\left(X_t^{1,g} - D_t^1\right)^+ + \left(X_t^{2,g} - D_t^2\right)^+ \\
=&x^j - D_t^j \\
\geq&x^1 + x^2 - D_t^1 - D_t^2 \\
=&\left(\left\lfloor \frac{1}{2}(x^1+x^2) \right\rfloor - D_t^1\right)^+ + \left(\left\lceil \frac{1}{2}(x^1+x^2) \right\rceil - D_t^2\right)^+.
\end{aligned} \tag{108}$$

If $x^i = 0$ and $x^j = 1$ $(i \neq j)$, then $\left\lfloor \frac{1}{2}(x^1+x^2) \right\rfloor = 0$ and $\left\lceil \frac{1}{2}(x^1+x^2) \right\rceil = 1$. Therefore,

$$\begin{aligned}
&\left(X_t^{1,g} - D_t^1\right)^+ + \left(X_t^{2,g} - D_t^2\right)^+ \\
=&1 - D_t^j \\
\geq_{st}&1 - D_t^2 \\
=&\left(\left\lfloor \frac{1}{2}(x^1+x^2) \right\rfloor - D_t^1\right)^+ + \left(\left\lceil \frac{1}{2}(x^1+x^2) \right\rceil - D_t^2\right)^+,
\end{aligned} \tag{109}$$

If $x^1, x^2 = 0$, then $\left\lfloor \frac{1}{2}(x^1+x^2) \right\rfloor, \left\lceil \frac{1}{2}(x^1+x^2) \right\rceil = 0$. Therefore,

$$\begin{aligned}
&\left(X_t^{1,g} - D_t^1\right)^+ + \left(X_t^{2,g} - D_t^2\right)^+ \\
=&0 \\
=&\left(\left\lfloor \frac{1}{2}(x^1+x^2) \right\rfloor - D_t^1\right)^+ + \left(\left\lceil \frac{1}{2}(x^1+x^2) \right\rceil - D_t^2\right)^+.
\end{aligned} \tag{110}$$

As a result of (107)-(110), we obtain

$$\begin{aligned}
&\left(X_t^{1,g} - D_t^1\right)^+ + \left(X_t^{2,g} - D_t^2\right)^+ \\
\geq_{st}&\left(\left\lfloor \frac{1}{2}(X_t^{1,g}+X_t^{2,g}) \right\rfloor - D_t^1\right)^+ + \left(\left\lceil \frac{1}{2}(X_t^{1,g}+X_t^{2,g}) \right\rceil - D_t^2\right)^+.
\end{aligned} \tag{111}$$

Then, from (111), the induction hypothesis and Corollary 2 we obtain

$$\begin{aligned}
&\left(X_t^{1,g} - D_t^1\right)^+ + \left(X_t^{2,g} - D_t^2\right)^+ \\
\geq_{st} &\left(\left\lfloor \frac{1}{2}(X_t^{1,g} + X_t^{2,g}) \right\rfloor - D_t^1\right)^+ + \left(\left\lceil \frac{1}{2}(X_t^{1,g} + X_t^{2,g}) \right\rceil - D_t^2\right)^+ \\
\geq_{st} &\left(\left\lfloor \frac{1}{2}(X_t^{1,\hat{g}} + X_t^{2,\hat{g}}) \right\rfloor - D_t^1\right)^+ + \left(\left\lceil \frac{1}{2}(X_t^{1,\hat{g}} + X_t^{2,\hat{g}}) \right\rceil - D_t^2\right)^+ \\
= &\left(\min(X_t^{1,\hat{g}}, X_t^{2,\hat{g}}) - D_t^1\right)^+ + \left(\max(X_t^{1,\hat{g}}, X_t^{2,\hat{g}}) - D_t^2\right)^+ \\
\geq_{st} &\left(X_t^{1,\hat{g}} - D_t^1\right)^+ + \left(X_t^{2,\hat{g}} - D_t^2\right)^+ . \qquad (112)
\end{aligned}$$

The first and second stochastic inequalities in (112) follow from (111) and the induction hypothesis, respectively. The equality in (112) follows from Corollary 2. The last stochastic inequality in (112) is true because D_t^1, D_t^2 are i.i.d. and independent of $X_t^{1,\hat{g}}, X_t^{2,\hat{g}}$.
Thus, inequality (106) is true, and the proof of the lemma is complete.

C Proofs of the Results Associated with Step 1 of the Proof of Theorem 2

Proof (Proof of Lemma 4) The proof is done by induction. At $t = 0$, (67), (68) and (69) hold if we let $Y_0^i = X_0^{i,g_0}$ for $i = 1, 2$.
Assume the assertion of this lemma is true at time t; we want to show that the assertion is also true at time $t + 1$.
For that matter we claim the following.
Claim 1

$$X_{t+1}^{1,\hat{g}} + X_{t+1}^{2,\hat{g}} = \overline{X}_t^{1,\hat{g}} + \overline{X}_t^{2,\hat{g}} \quad a.s., \qquad (113)$$

$$\max_i \left(X_{t+1}^{i,\hat{g}}\right) \leq \max_i \left(\overline{X}_t^{i,\hat{g}}\right) \quad a.s. \qquad (114)$$

Claim 2
There exists $Y_{t+1}^i, i = 1, 2$ such that

$$\mathbf{P}\left(Y_{t+1}^i = y_{t+1} | Y_{0:t}^i = y_{0:t}\right) = \mathbf{P}\left(X_{t+1}^{i,g0} = y_{t+1} | X_{0:t}^{i,g_0} = y_{0:t}\right) \text{ for all } y_{0:t}, \qquad (115)$$

$$\overline{X}_t^{1,\hat{g}} + \overline{X}_t^{2,\hat{g}} \leq Y_{t+1}^1 + Y_{t+1}^2 \quad a.s., \qquad (116)$$

$$\max_i \left(\overline{X}_t^{i,\hat{g}}\right) \leq \max_i \left(Y_{t+1}^i\right) \quad a.s. \qquad (117)$$

We assume the above claims to be true and prove them after the completion of the proof of the induction step.
For all $y_{0:t+1}$, from (115) and the induction hypothesis for (67) we get for $i = 1, 2$

$$\begin{aligned}
&\mathbf{P}\left(Y_{0:t+1}^i = y_{0:t+1}\right) \\
=&\mathbf{P}\left(Y_{t+1}^i = y_{t+1} | Y_{0:t}^i = y_{0:t}\right) \mathbf{P}\left(Y_t^i = y_t, \ldots, Y_0^i = y_0\right) \\
=&\mathbf{P}\left(X_{t+1}^{i,g0} = y_{t+1} | X_{0:t}^{i,g_0} = y_{0:t}\right) \mathbf{P}\left(X_{0:t}^{i,g_0} = y_{0:t}\right) \\
=&\mathbf{P}\left(X_{0:t+1}^{i,g0} = y_{0:t+1}\right). \qquad (118)
\end{aligned}$$

From (113) and (116) we obtain

$$\begin{aligned}
X_{t+1}^{1,\hat{g}} + X_{t+1}^{2,\hat{g}} =& \overline{X}_t^{1,\hat{g}} + \overline{X}_t^{2,\hat{g}} \\
\leq& Y_{t+1}^1 + Y_{t+1}^2 \quad a.s. \qquad (119)
\end{aligned}$$

Furthermore, combination of (114) and (117) gives

$$\max_i \left(X_{t+1}^{i,\hat{g}}\right) \leq \max_i \left(\overline{X}_t^{i,\hat{g}}\right) = \max_i \left(Y_{t+1}^i\right) \quad a.s. \tag{120}$$

Therefore, the assertions (67), (68) and (69) of the lemma are true at $t+1$ by (118), (119) and (120), respectively.
We now prove claims 1 and 2.
Proof of Claim 1
From the system dynamics (1)-(2)

$$X_{t+1}^{1,\hat{g}} = \overline{X}_t^{i,\hat{g}} - U_t^{i,\hat{g}} + U_t^{j,\hat{g}}, \tag{121}$$

$$X_{t+1}^{2,\hat{g}} = \overline{X}_t^{i,\hat{g}} - U_t^{i,\hat{g}} + U_t^{j,\hat{g}}. \tag{122}$$

Therefore, (113) follows by summing (121) and (122).
For (114), consider $X_{t+1}^{1,\hat{g}}$ (the case of $X_{t+1}^{2,\hat{g}}$ follows from similar arguments).
When $U_t^{2,\hat{g}} = 0$,

$$X_{t+1}^{1,\hat{g}} = \overline{X}_t^{1,\hat{g}} - U_t^{1,\hat{g}} \leq \max_i \left(\overline{X}_t^{i,\hat{g}}\right). \tag{123}$$

When $U_t^{1,\hat{g}} = U_t^{2,\hat{g}} = 1$,

$$X_{t+1}^{1,\hat{g}} = \overline{X}_t^{1,\hat{g}} \leq \max_i \left(\overline{X}_t^{i,\hat{g}}\right). \tag{124}$$

When $U_t^{1,\hat{g}} = 0, U_t^{2,\hat{g}} = 1$, $\overline{X}_t^{1,\hat{g}}$ is less than the threshold and $\overline{X}_t^{2,\hat{g}}$ is greater than or equal to the threshold. Therefore, by (121),

$$\begin{aligned} X_{t+1}^{1,\hat{g}} = \overline{X}_t^{1,\hat{g}} + 1 \leq & \lceil TH_t \rceil \\ \leq & \overline{X}_t^{2,\hat{g}} \leq \max_i \left(\overline{X}_t^{i,\hat{g}}\right). \end{aligned} \tag{125}$$

Therefore, (114) follows from (123)-(125).
Proof of Claim 2
We set

$$Y_{t+1}^i := \left(Y_t^i - \tilde{D}_t^i\right)^+ + \tilde{A}_t^i \tag{126}$$

where Y_t^i satisfy the induction hypothesis, and $\tilde{A}_t^i, \tilde{D}_t^i, i = 1, 2$ are specified as follows. Let

$$M_x = \text{argmax}_i\{X_t^{i,\hat{g}}\}, \quad m_x = \text{argmin}_i\{X_t^{i,\hat{g}}\} \tag{127}$$

$$M_y = \text{argmax}_i\{Y_t^i\}, \quad m_y = \text{argmin}_i\{Y_t^i\}, \tag{128}$$

where $M_x = 1, m_x = 2$ (resp. $M_y = 1, m_y = 2$) when $\{X_t^{1,\hat{g}} = X_t^{2,\hat{g}}\}$ (resp. $\{Y_t^1 = Y_t^2\}$); define

$$\left(\tilde{A}_t^{M_y}, \tilde{D}_t^{M_y}, \tilde{A}_t^{m_y}, \tilde{D}_t^{m_y}\right) := \begin{cases} \left(A_t^{M_x}, D_t^{m_x}, A_t^{m_x}, D_t^{M_x}\right) & \text{in case 1,} \\ \left(A_t^{M_x}, D_t^{M_x}, A_t^{m_x}, D_t^{m_x}\right) & \text{in case 2,} \end{cases} \tag{129}$$

where the two cases are :
Case 1: $\{Y_t^{M_y} - 1 = X_t^{M_x,\hat{g}} = X_t^{m_x,\hat{g}}$ and $\left(A_t^{M_x}, D_t^{M_x}, A_t^{m_x}, D_t^{m_x}\right) = (0,1,1,0)$ or $(0,0,1,1)\}$.
Case 2: All other instances.

Assertion: The random variables Y_{t+1}^1, Y_{t+1}^2, defined by (126)-(129) satisfy (115)-(117).

As the proof of this assertion is long, we first provide a sketch of its proof and then we provide a full proof.
Sketch of the proof of the assertion

– Equation (129) implies the following: In case 2 we associate the arrival to and the departure from the longer queue M_x to those of the longer queue M_y, i.e. we set $\tilde{A}_t^{M_y} = A_t^{M_x}, \tilde{D}_t^{M_y} = D_t^{M_x}$. We do the same for the shorter queue m_x, m_y, i.e. $\tilde{A}_t^{m_y} = A_t^{m_x}, \tilde{D}_t^{m_y} = D_t^{m_x}$.
In case 1, we have the same association for the arrivals as in case 2, that is $\tilde{A}_t^{M_y} = A_t^{M_x}, \tilde{A}_t^{m_y} = A_t^{m_x}$, but we reverse the association of the departures, that is $\tilde{D}_t^{M_y} = D_t^{m_x}, \tilde{D}_t^{m_y} = D_t^{M_x}$. Therefore the arrivals $\tilde{A}_t^i$, and departures $\tilde{D}_t^i$, have the same distribution as the original A_t^i, D_t^i, respectively, $i = 1, 2$. Then (115) follows from (126).
– To establish (116), we note that, because of (129), the sum of arrivals to (respectively, departures from) queues M_y and m_y equals to the sum of arrivals to (respectively, departures from) queues M_x and m_x.
When $X_t^{i,\hat{g}}, Y_t^i \neq 0$, $i = 1, 2$, the function $(x-d)^+ + a$ is linear x, as $(x-d)^+ + a = x - d + a$. Then from (126), (129) and the induction hypothesis we obtain

$$\begin{aligned} & Y_{t+1}^1 + Y_{t+1}^2 - \overline{X}_t^{1,\hat{g}} - \overline{X}_t^{2,\hat{g}} \\ =& Y_t^1 + Y_t^2 - X_t^{1,\hat{g}} - X_t^{2,\hat{g}} \geq 0 \end{aligned} \tag{130}$$

and this establish (116) when $X_t^{i,\hat{g}}, Y_t^i \neq 0$, $i = 1, 2$. In the full proof of the assertion, we show that show that (116) is also true when $X_t^{i,\hat{g}}, Y_t^i$ are not all non-zero.
– To establish (117) we consider the maximum of the queue lengths. In case 2, we show that (126)-(129) ensure that

$$Y_{t+1}^{M_y} \geq \overline{X}_t^{M_x,\hat{g}}, \tag{131}$$

$$\max\left(Y_{t+1}^{M_y}, Y_{t+1}^{m_y}\right) \geq \overline{X}_t^{m_x,\hat{g}}; \tag{132}$$

then (117) follows from (131)-(132).
In case 1 (117) is verified by direct computation in the full proof.

Proof of the assertion
For all $y_{0:t}$, we denote by $E_{y_{0:t}}$ the event $\{Y_{0:t}^i = y_{0:t}\}$.
Let $\tilde{Z}_t = \left(\tilde{A}_t^{M_y}, \tilde{D}_t^{M_y}, \tilde{A}_t^{m_y}, \tilde{D}_t^{m_y}\right)$, then for any realization $z_t \in \{0,1\}^4$ of $\tilde{Z}_t$ we have

$$\begin{aligned} & \mathbf{P}\left(\tilde{Z}_t = z_t | E_{y_{0:t}}\right) \\ =& \mathbf{P}\left(\tilde{Z}_t = z_t, \text{case } 1 | E_{y_{0:t}}\right) + \mathbf{P}\left(\tilde{Z}_t = z_t, \text{case } 2 | E_{y_{0:t}}\right). \end{aligned} \tag{133}$$

When $z_t \neq (0,1,1,0)$ or $(0,0,1,1)$, we get

$$\mathbf{P}\left(\tilde{Z}_t = z_t, \text{case } 1 | E_{y_{0:t}}\right) = 0, \tag{134}$$

and

$$\begin{aligned} & \mathbf{P}\left(\tilde{Z}_t = z_t, \text{case } 2 | E_{y_{0:t}}\right) \\ =& \mathbf{P}\left(\left(A_t^{M_x}, D_t^{M_x}, A_t^{m_x}, D_t^{m_x}\right) = z_t | E_{y_{0:t}}\right) \\ =& \mathbf{P}\left(\left(A_t^1, D_t^1, A_t^2, D_t^2\right) = z_t\right), \end{aligned} \tag{135}$$

where the last equality in (135) holds because the random variables $A_t^{M_x}, D_t^{M_x}, A_t^{m_x}, D_t^{m_x}$ are independent of $Y_0, Y_1, \ldots, Y_t$ and have the same distribution as $A_t^1, D_t^1, A_t^2, D_t^2$.
Therefore, combining (134) and (135) we obtain for $z_t \neq (0,1,1,0)$ or $(0,0,1,1)$

$$\mathbf{P}\left(\tilde{Z}_t = z_t | E_{y_{0:t}}\right) = \mathbf{P}\left(\left(A_t^1, D_t^1, A_t^2, D_t^2\right) = z_t\right) \tag{136}$$

When $z_t = (0,1,1,0)$ or $(0,0,1,1)$, let E denote the event $\{Y_t^{M_y} - 1 = X_t^{M_x,\hat{g}} = X_t^{m_x,\hat{g}}\}$; then we obtain

$$\begin{aligned}
&\mathbf{P}\left(\tilde{Z}_t = z_t, \text{case } 1 | E_{y_{0:t}}\right)\\
=&\mathbf{P}\left(\left(A_t^{M_x}, D_t^{m_x}, A_t^{m_x}, D_t^{M_x}\right) = z_t, E | E_{y_{0:t}}\right)\\
=&\mathbf{P}\left(\left(A_t^1, D_t^2, A_t^2, D_t^1\right) = z_t\right)\mathbf{P}\left(E | E_{y_{0:t}}\right),
\end{aligned} \tag{137}$$

and

$$\begin{aligned}
&\mathbf{P}\left(\tilde{Z}_t = z_t, \text{case } 2 | E_{y_{0:t}}\right)\\
=&\mathbf{P}\left(\left(A_t^{M_x}, D_t^{M_x}, A_t^{m_x}, D_t^{m_x}\right) = z_t, E^c | E_{y_{0:t}}\right)\\
=&\mathbf{P}\left(\left(A_t^1, D_t^1, A_t^2, D_t^2\right) = z_t\right)\mathbf{P}\left(E^c | E_{y_{0:t}}\right),
\end{aligned} \tag{138}$$

where the last equality in (137) and (138) follow by the fact that the random variables $A_t^{M_x}, D_t^{M_x}, A_t^{m_x}, D_t^{m_x}$ are independent of $Y_0, Y_1, \ldots, Y_t$ (hence, the event E which is generated by $Y_0, Y_1, \ldots, Y_t$) and have the same distribution as $A_t^1, D_t^1, A_t^2, D_t^2$.
Therefore, combining (137) and (138) we obtain for $z_t = (0,1,1,0)$ or $(0,0,1,1)$

$$\begin{aligned}
&\mathbf{P}\left(\tilde{Z}_t = z_t | E_{y_{0:t}}\right)\\
=&\mathbf{P}\left(\left(A_t^1, D_t^2, A_t^2, D_t^1\right) = z_t\right)\mathbf{P}\left(E | E_{y_{0:t}}\right)\\
&+\mathbf{P}\left(\left(A_t^1, D_t^1, A_t^2, D_t^2\right) = z_t\right)\mathbf{P}\left(E^c | E_{y_{0:t}}\right)\\
=&\mathbf{P}\left(\left(A_t^1, D_t^1, A_t^2, D_t^2\right) = z_t\right),
\end{aligned} \tag{139}$$

where the last equality in (139) is true because $A_t^1, D_t^1, A_t^2, D_t^2$ are independent and D_t^1 has the same distribution as D_t^2.
As a result of (136) and (139), for any $z_t \in \{0,1\}^4$ we have

$$\mathbf{P}\left(\tilde{Z}_t = z_t | E_{y_{0:t}}\right) = \mathbf{P}\left(\left(A_t^1, D_t^1, A_t^2, D_t^2\right) = z_t\right). \tag{140}$$

Now consider any $y_{0:t+1}$. By (140) we have for $i = M_y$ or m_y

$$\begin{aligned}
&\mathbf{P}\left(Y_{t+1}^i = y_{t+1} | E_{y_{0:t}}\right)\\
=&\mathbf{P}\left(\left(y_t^i - \tilde{D}_t^i\right)^+ + \tilde{A}_t^i = y_{t+1} | E_{y_{0:t}}\right)\\
=&\mathbf{P}\left(\left(y_t^i - D_t^i\right)^+ + A_t^i = y_{t+1}\right)\\
=&\mathbf{P}\left(X_{t+1}^{i,g0} = y_{t+1} | X_{0:t}^{i,g_0} = y_{0:t}\right).
\end{aligned} \tag{141}$$

which is (115).
Now consider the sum $Y_{t+1}^1 + Y_{t+1}^2$.
From (129), we know that

$$\tilde{A}_t^{M_y} + \tilde{A}_t^{m_y} = A_t^{M_x} + A_t^{m_x} \quad a.s., \tag{142}$$

$$\tilde{D}_t^{M_y} + \tilde{D}_t^{m_y} = D_t^{M_x} + D_t^{m_x} \quad a.s. \tag{143}$$

Therefore, (142) implies

$$\begin{aligned}
&Y_{t+1}^1 + Y_{t+1}^2 - \overline{X}_{t+1}^{1,\hat{g}} - \overline{X}_{t+1}^{1,\hat{g}}\\
=&\left(Y_t^{M_y} - \tilde{D}_t^{M_y}\right)^+ + \left(Y_t^{m_y} - \tilde{D}_t^{m_y}\right)^+\\
&-\left(X_t^{M_x,\hat{g}} - D_t^{M_x}\right)^+ - \left(X_t^{m_x,\hat{g}} - D_t^{m_x}\right)^+.
\end{aligned} \tag{144}$$

We proceed to show that the right hand side of (144) is positive. From the induction hypothesis for (69)-(68) we have

$$Y_t^{m_y} + Y_t^{M_y} \geq X_t^{m_x,\hat{g}} + X_t^{M_x,\hat{g}} \quad a.s., \tag{145}$$

$$Y_t^{M_y} \geq X_t^{M_x,\hat{g}} \quad a.s. \tag{146}$$

There are three possibilities: $\{Y_t^{M_y} = X_t^{M_x,\hat{g}}\}$, $\{Y_t^{M_y} > X_t^{M_x,\hat{g}}, X_t^{m_x,\hat{g}} = 0\}$ and $\{Y_t^{M_y} > X_t^{M_x,\hat{g}}, X_t^{m_x} > 0\}$.
First consider $\{Y_t^{M_y} = X_t^{M_x,\hat{g}}\}$. By (145) we have

$$Y_t^{m_y} \geq X_t^{m_x,\hat{g}} \quad a.s. \tag{147}$$

Note that $\{Y_t^{M_y} = X_t^{M_x,\hat{g}}\}$ belongs to case 2 in (129). From case 2 of (129) we also know that

$$D_t^{M_x} = \tilde{D}_t^{M_y}, \quad D_t^{m_x} = \tilde{D}_t^{m_y}. \tag{148}$$

Then, because of (146)-(148) we get

$$\begin{aligned}
&\left(X_t^{M_x,\hat{g}} - D_t^{M_x}\right)^+ + \left(X_t^{m_x,\hat{g}} - D_t^{m_x}\right)^+ \\
\leq &\left(Y_t^{M_y} - D_t^{M_x}\right)^+ + \left(Y_t^{m_y} - D_t^{m_x}\right)^+ \\
= &\left(Y_t^{M_y} - \tilde{D}_t^{M_y}\right)^+ + \left(Y_t^{m_y} - \tilde{D}_t^{m_y}\right)^+ \quad a.s.
\end{aligned} \tag{149}$$

If $Y_t^{M_y} > X_t^{M_x,\hat{g}}$ and $X_t^{m_x,\hat{g}} = 0$

$$\begin{aligned}
&\left(X_t^{M_x,\hat{g}} - D_t^{M_x}\right)^+ + \left(X_t^{m_x,\hat{g}} - D_t^{m_x}\right)^+ \\
= &\left(X_t^{M_x,\hat{g}} - D_t^{M_x}\right)^+ \\
\leq &X_t^{M_x,\hat{g}} \leq Y_t^{M_y} - 1 \\
\leq &\left(Y_t^{M_y} - \tilde{D}_t^{M_y}\right)^+ + \left(Y_t^{m_y} - \tilde{D}_t^{m_y}\right)^+
\end{aligned} \tag{150}$$

If $Y_t^{M_y} > X_t^{M_x,\hat{g}}$ and $X_t^{m_x} > 0$, then

$$\begin{aligned}
&\left(X_t^{M_x,\hat{g}} - D_t^{M_x}\right)^+ + \left(X_t^{m_x,\hat{g}} - D_t^{m_x}\right)^+ \\
= &X_t^{M_x,\hat{g}} - D_t^{M_x} + X_t^{m_x,\hat{g}} - D_t^{m_x} \\
= &X_t^{M_x,\hat{g}} + X_t^{m_x,\hat{g}} - \tilde{D}_t^{M_y} - \tilde{D}_t^{m_y} \\
\leq &Y_t^{M_y} + Y_t^{m_y} - \tilde{D}_t^{M_y} - \tilde{D}_t^{m_y} \\
\leq &\left(Y_t^{M_y} - \tilde{D}_t^{M_y}\right)^+ + \left(Y_t^{m_y} - \tilde{D}_t^{m_y}\right)^+
\end{aligned} \tag{151}$$

where the second equality in (151) follows from (143) and the first inequality in (151) follows from the induction hypothesis for (68).
The above results, namely (149)-(151), show that the right hand side of (144) is positive, and the proof for (116) is complete.
It remains to show that (117) is true.
We first consider case 2.
In case 2, we know from (129) that

$$\left(\tilde{A}_t^{M_y}, \tilde{D}_t^{M_y}, \tilde{A}_t^{m_y}, \tilde{D}_t^{m_y}\right) = \left(A_t^{M_x}, D_t^{M_x}, A_t^{m_x}, D_t^{m_x}\right). \tag{152}$$

Then,

$$\begin{aligned}\overline{X}_t^{M_x,\hat{g}} =& \left(X_t^{M_x,\hat{g}} - D_t^{M_x}\right)^+ + A_t^{M_x}\\ =& \left(X_t^{M_x,\hat{g}} - \tilde{D}_t^{M_y}\right)^+ + \tilde{A}_t^{M_y}\\ \leq& \left(Y_t^{M_y} - \tilde{D}_t^{M_y}\right)^+ + \tilde{A}_t^{M_y}\\ =& Y_{t+1}^{M_y},\end{aligned} \tag{153}$$

where the second equality is a consequence of (152) and the inequality follows from the induction hypothesis for (69).
To proceed further we note that in case 2 there are three possibilities: $\{Y_t^{M_y} = X_t^{M_x,\hat{g}}\}$, $\{Y_t^{M_y} - 2 \geq X_t^{m_x,\hat{g}}\}$ and $\{Y_t^{M_y} > X_t^{M_x,\hat{g}}, Y_t^{M_y} - 2 < X_t^{m_x,\hat{g}}\}$
If $Y_t^{M_y} = X_t^{M_x,\hat{g}}$, (147) is also true. Following similar arguments as in (153) we obtain

$$\overline{X}_t^{m_x,\hat{g}} \leq Y_{t+1}^{m_y}. \tag{154}$$

If $Y_t^{M_y} - 2 \geq X_t^{m_x,\hat{g}}$

$$\overline{X}_t^{m_x,\hat{g}} \leq X_t^{m_x,\hat{g}} + 1 \leq Y_t^{M_y} - 1 \leq Y_{t+1}^{M_y}. \tag{155}$$

If $Y_t^{M_y} > X_t^{M_x,\hat{g}}$ and $Y_t^{M_y} - 2 < X_t^{m_x,\hat{g}}$ it can only be $Y_t^{M_y} - 1 = X_t^{M_x,\hat{g}} = X_t^{m_x,\hat{g}}$. Since we are in case 2, $\left(A_t^{M_x}, D_t^{M_x}, A_t^{m_x}, D_t^{m_x}\right) \neq (0,1,1,0)$. Therefore,

$$A_t^{m_x} - D_t^{m_x} \leq A_t^{M_x} - D_t^{M_x} + 1. \tag{156}$$

Then we get

$$\begin{aligned}\overline{X}_t^{m_x,\hat{g}} =& \left(Y_t^{M_y} - 1 - D_t^{m_x}\right)^+ + A_t^{m_x}\\ =& \max\left(A_t^{m_x}, Y_t^{M_y} - 1 - D_t^{m_x} + A_t^{m_x}\right)\\ \leq& \max\left(A_t^{m_x}, Y_t^{M_y} - D_t^{M_x} + A_t^{M_x}\right)\\ \leq& \max\left(A_t^{m_x}, Y_{t+1}^{M_y}\right)\\ \leq& \max\left(Y_{t+1}^{m_y}, Y_{t+1}^{M_y}\right).\end{aligned} \tag{157}$$

Combining (153), (154), (155) and (157) we get (117) when case 2 is true.
Now consider case 1. We have $Y_t^{M_y} - 1 = X_t^{M_x,\hat{g}} = X_t^{m_x,\hat{g}}$.
When $\left(A_t^{M_x}, D_t^{M_x}, A_t^{m_x}, D_t^{m_x}\right) = (0,1,1,0)$, then

$$\begin{aligned}\overline{X}_t^{M_x,\hat{g}} =& \left(X_t^{M_x,\hat{g}} - 1\right)^+\\ \leq& \overline{X}_t^{m_x}\\ =& X_t^{m_x} + 1\\ =& \left(Y_t^{M_y} - D_t^{m_x}\right)^+ + A_t^{M_x}\\ =& Y_{t+1}^{M_y}\end{aligned} \tag{158}$$

When $\left(A_t^{M_x}, D_t^{M_x}, A_t^{m_x}, D_t^{m_x}\right) = (0,0,1,1)$ we get

$$\begin{aligned}
\overline{X}_t^{M_x,\hat{g}} =& X_t^{M_x,\hat{g}} \\
\leq& \overline{X}_t^{m_x,\hat{g}} \\
=& \max\left(X_t^{m_x,\hat{g}}, 1\right) \\
=& \max\left(\left(Y_t^{M_y} - D_t^{m_x}\right)^+ + A_t^{M_x}, A_t^{m_x}\right) \\
=& \max\left(Y_{t+1}^{M_y}, A_t^{m_x}\right) \\
\leq& \max\left(Y_{t+1}^{M_y}, Y_{t+1}^{m_y}\right). \qquad (159)
\end{aligned}$$

Combining (158) and (159) we obtain (117) for case 1.
As a result, (117) holds for both cases 1 and 2.
Remark:

We note that we need the two cases described in (129) for the following reasons. If we eliminate case 1 and always associate $\left(\tilde{A}_t^{M_y}, \tilde{D}_t^{M_y}, \tilde{A}_t^{m_y}, \tilde{D}_t^{m_y}\right)$ with $\left(A_t^{M_x}, D_t^{M_x}, A_t^{m_x}, D_t^{m_x}\right)$ as in case 2, then when $\{Y_t^{M_y} - 1 = X_t^{m_x,\hat{g}}$ and $\left(A_t^{M_x}, D_t^{M_x}, A_t^{m_x}, D_t^{m_x}\right) = (0,1,1,0)\}$, the shorter queue m_x increases by one customer, and the longer queue M_y decreases by one customer; therefore $\overline{X}_t^{m_x,\hat{g}} = Y_{t+1}^{M_y} + 1$ and (117) is not satisfied.

Proof (Proof of Lemma 5) From Lemma 4, at any time t there exists Y_t^i such that such that (67)-(69) hold.
Adopting the notations M_x, m_x and M_y, m_y in the proof of Lemma 4, we have at every time t

$$X_t^{m_x,\hat{g}} \leq X_t^{M_X,\hat{g}} \quad a.s., \qquad (160)$$

$$Y_t^{m_y} \leq Y_t^{M_y} \quad a.s. \qquad (161)$$

Furthermore, from (69) we have

$$X_t^{M_x,\hat{g}} \leq Y_t^{M_y} \quad a.s. \qquad (162)$$

If $X_t^{m_x,\hat{g}} \leq Y_t^{m_y}$, (162) and the fact that $c(\cdot)$ is increasing give

$$c\left(X_t^{M_X,\hat{g}}\right) + c\left(X_t^{m_X,\hat{g}}\right) \leq c\left(Y_t^{M_y}\right) + c\left(Y_t^{m_y}\right). \qquad (163)$$

If $X_t^{m_x,\hat{g}} > Y_t^{m_y}$, then

$$Y_t^{m_y} < X_t^{m_x,\hat{g}} \leq X_t^{M_x,\hat{g}} \leq Y_t^{M_y}. \qquad (164)$$

Since $c(\cdot)$ is convex, it follows from (164) that

$$\frac{c\left(Y_t^{M_y}\right) - c\left(X_t^{M_x,\hat{g}}\right)}{Y_t^{M_y} - X_t^{M_x,\hat{g}}} \geq \frac{c\left(X_t^{m_x,\hat{g}}\right) - c\left(Y_t^{m_y}\right)}{X_t^{m_x,\hat{g}} - Y_t^{m_y}}. \qquad (165)$$

From (68) in Lemma 4 we know that

$$Y_t^{M_y} - X_t^{M_x,\hat{g}} \geq X_t^{m_x,\hat{g}} - Y_t^{m_y}. \qquad (166)$$

Combining (165) and (166) we get

$$c\left(Y_t^{M_y}\right) + c\left(Y_t^{m_y}\right) \geq c\left(X_t^{M_x,\hat{g}}\right) + c\left(X_t^{m_x,\hat{g}}\right). \qquad (167)$$

Proof (Proof of Lemma 6) Let $\{Y_t^1, t \in \mathbb{Z}_+\}$ and $\{Y_t^2, t \in \mathbb{Z}_+\}$ be the processes defined in Lemma 4. Then $\{Y_t^i, t \in \mathbb{Z}_+\}$ has the same distribution as $\{X_t^{i,g_0}, t \in \mathbb{Z}_+\}$ for $i = 1, 2$.
Since $\mu > \lambda$, the processes $\{Y_t^i, t \in \mathbb{Z}_+\}, i = 1, 2$ are irreducible positive recurrent Markov chains. Moreover, the two processes $\{Y_t^1, t \in \mathbb{Z}_+\}$ and $\{Y_t^2, t \in \mathbb{Z}_+\}$ have the same stationary distribution, denoted by π^{g_0}. Under Assumption 2, by Ergodic theorem of Markov chains (see (Bremaud, 1999, chap. 3)) we get

$$\begin{aligned}\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} c(Y_t^1) &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} c(Y_t^2) \\ &= \sum_{x=0}^{\infty} \pi^{g_0}(x) c(x) \quad a.s. \end{aligned} \tag{168}$$

Let $W_T^i(Y_{0:T-1}) := \frac{1}{T} \sum_{t=0}^{T-1} c(Y_t^i), i = 1, 2$.
We show that $\{W_T^i(Y_{0:T-1}), T = 1, 2, \dots\}$ is uniformly integrable for $i = 1, 2$. That is,

$$\sup_T \mathbf{E}\left[W_T^i(Y_{0:T-1}) 1_{\{W_T^i(Y_{0:T-1}) > N\}}\right] \to 0 \tag{169}$$

as $N \to \infty$.
Let $p^{g_0}(x, y), x, y \in \mathbb{Z}_+$ be the transition probabilities of the Markov chain. Note that the initial PMF of the process $\{Y_t^i, t \in \mathbb{Z}_+\}, i = 1, 2$ is π_0^i. From Assumption 2 we know that $\pi_0^i(x) = 0, i = 1, 2$ for all $x > M$.
Letting $R := \max_{x \le M} \frac{\pi_0^i(x)}{\pi^{g_0}(x)} < \infty$, we obtain for $i = 1, 2$

$$\begin{aligned}&\mathbf{E}\left[W_T^i(Y_{0:T-1}) 1_{\{W_T^i(Y_{0:T-1}) > N\}}\right] \\ =& \sum_{y_{0:T-1}} W_T^i(y_{0:T-1}) 1_{\{W_T^i(y_{0:T-1}) > N\}} \mathbf{P}(Y_{0:T-1} = y_{0:T-1}) \\ =& \sum_{y_{0:T-1}} W_T^i(y_{0:T-1}) 1_{\{W_T^i(y_{0:T-1}) > N\}} \pi_0^i(y_0) \Pi_{t=1}^{T-1} p^{g_0}(y_{t-1}, y_t) \\ \le& R \sum_{y_{0:T-1}} W_T^i(y_{0:T-1}) 1_{\{W_T^i(y_{0:T-1}) > N\}} \pi^{g_0}(y_0) \Pi_{t=1}^{T-1} p^{g_0}(y_{t-1}, y_t) \\ =& R\mathbf{E}\left[W_T^{\pi^{g_0}} 1_{\{W_T^{\pi^{g_0}} > N\}}\right], \end{aligned} \tag{170}$$

where $W_T^{\pi^{g_0}} = \frac{1}{T} \sum_{t=0}^{T-1} c(Y_t^{\pi^{g_0}})$ and $\{Y_t^{\pi^{g_0}}, t \in \mathbb{Z}_+\}$ is the chain with transition probabilities $p^{g_0}(x, y)$ and initial PMF π^{g_0}.
Note that $\{Y_t^{\pi^{g_0}}, t \in \mathbb{Z}_+\}$ is stationary because the initial PMF is the stationary distribution π^{g_0}. From Birkhoff's Ergodic theorem we know that $\{W_T^{\pi^{g_0}}, T = 1, 2, \dots\}$ converges $a.s.$ and in expectation (see (Petersen and Petersen, 1989, chap. 2)). Therefore, $\{W_T^{\pi^{g_0}}, T = 1, 2, \dots\}$ is uniformly integrable, and the right hand side of (170) goes to zeros uniformly as $N \to \infty$. Consequently, $\{W_T^i(Y_{0:T-1}), T = 1, 2, \dots\}$ is also uniformly integrable for $i = 1, 2$.
Since $W_T = W_T^1(Y_{0:T-1}) + W_T^2(Y_{0:T-1})$ for all $T = 1, 2, \dots$, $\{W_T, T = 1, 2, \dots\}$ is uniformly integrable.

Proof (Proof of Corollary 3) From Lemma 5, there exists $\{Y_t^1, Y_t^2, t \in \mathbb{Z}_+\}$ such that (67) holds and

$$c\left(X_t^{1,\hat{g}}\right) + c\left(X_t^{2,\hat{g}}\right) \le c\left(Y_t^1\right) + c\left(Y_t^2\right) \quad a.s. \tag{171}$$

Let

$$W_T := \frac{1}{T}\sum_{t=0}^{T-1}\left(c\left(Y_t^1\right)+c\left(Y_t^2\right)\right), \tag{172}$$

$$V_T := \frac{1}{T}\sum_{t=0}^{T-1}\left(c\left(X_t^{1,\hat{g}}\right)+c\left(X_t^{2,\hat{g}}\right)\right). \tag{173}$$

From (171) it follows that

$$V_T \le W_T, T = 1, 2, \dots \tag{174}$$

From Lemmas 6, $\{W_T, T = 1, 2, \dots\}$ is uniformly integrable, therefore $\{V_T, T = 1, 2, \dots\}$, which is bounded above by $\{W_T, T = 1, 2, \dots\}$ is also uniformly integrable.
From the property of uniformly integrability, if $\{V_T, T = 1, 2, \dots\}$ converges a.s., we know that $\{V_T, T = 1, 2, \dots\}$ also converges in expectation. Furthermore,

$$\begin{aligned} J^{\hat{g}}\left(\pi_0^1, \pi_0^2\right) =& \limsup_{T\to\infty}\frac{1}{T}\mathbf{E}\left[\sum_{t=0}^{T-1}\left(c\left(Y_t^1\right)+c\left(Y_t^2\right)\right)\right] \\ =& \limsup_{T\to\infty}\mathbf{E}\left[V_T\right] \\ \le& \limsup_{T\to\infty}\mathbf{E}\left[W_T\right] = J^{g_0}. \end{aligned} \tag{175}$$

D Proofs of the Results Associated with Step 2 of the Proof of Theorem 2

Proof (Proof of Lemma 7) First we show that $\{S_t, t \ge T_0 + 1\}$ is a Markov chain.
For $s_t \ge 2$,

$$\begin{aligned} &\mathbf{P}\left(S_{t+1} = s_{t+1} | S_{T_0+1:t} = s_{T_0+1:t}\right) \\ =&\mathbf{P}\left(\left(s_t - D_t^1 - D_t^2 + A_t^1 + A_t^2\right) = s_{t+1}\right. \\ &\qquad\left. | S_{T_0+1:t} = s_{T_0+1:t}\right) \\ =&\mathbf{P}\left(\left(s_t - D_t^1 - D_t^2 + A_t^1 + A_t^2\right) = s_{t+1} | S_t = s_t\right) \\ =&\mathbf{P}\left(S_{t+1} = s_{t+1} | S_t = s_t\right). \end{aligned} \tag{176}$$

The first and last equalities in (176) follow from the construction of the process $\{S_t, t \ge T_0 + 1\}$. The second equality in (176) is true because T_0 is a stopping time with respect to $\{X_t^{1,\hat{g}}, X_t^{2,\hat{g}}, t \in \mathbb{Z}_+\}$, and $A_t^i, D_t^i, i = 1, 2$ are independent of all random variables before t. Similarly, for $s_t = 0$ we have, by arguments similar to the above,

$$\begin{aligned} &\mathbf{P}\left(S_{t+1} = s_{t+1} | S_{T_0+1:t} = s_{T_0+1:t}\right) \\ =&\mathbf{P}\left(A_t^1 + A_t^2 = s_{t+1} | S_{T_0+1:t-1} = s_{T_0+1:t-1}, S_t = 0\right) \\ =&\mathbf{P}\left(A_t^1 + A_t^2 = s_{t+1} | S_t = 0\right) \\ =&\mathbf{P}\left(S_{t+1} = s_{t+1} | S_t = 0\right). \end{aligned} \tag{177}$$

The first and last equality in (177) follow from the construction of the process $\{S_t, t \ge T_0 + 1\}$. The second equality in (177) is true because $A_t^i, D_t^i, i = 1, 2$ are independent of all

variables before t. For $s_t = 1$,

$$\begin{aligned}
&\mathbf{P}\left(S_{t+1} = s_{t+1} | S_{T_0+1:t} = s_{T_0+1:t}\right)\\
=&\mathbf{P}\left(s_t + 1_{\left\{X_t^{1,\hat{g}}=0\right\}}(D_t^1 - D_t^2) - D_t^1 + A_t^1 + A_t^2 = s_{t+1} | S_{T_0+1:t} = s_{T_0+1:t}\right)\\
=&\mathbf{P}\left(1 - D_t^2 + A_t^1 + A_t^2 = s_{t+1}, X_t^{1,\hat{g}} = 0 | S_{T_0+1:t-1} = s_{T_0+1:t-1}, S_t = 1\right)\\
&+\mathbf{P}\left(1 - D_t^1 + A_t^1 + A_t^2 = s_{t+1}, X_t^{1,\hat{g}} = 1 | S_{T_0+1:t-1} = s_{T_0+1:t-1}, S_t = 1\right)\\
=&\mathbf{P}\left(1 - D_t^1 + A_t^1 + A_t^2 = s_{t+1}, X_t^{1,\hat{g}} = 0 | S_{T_0+1:t-1} = s_{T_0+1:t-1}, S_t = 1\right)\\
&+\mathbf{P}\left(1 - D_t^1 + A_t^1 + A_t^2 = s_{t+1}, X_t^{1,\hat{g}} = 1 | S_{T_0+1:t-1} = s_{T_0+1:t-1}, S_t = 1\right)\\
=&\mathbf{P}\left(1 - D_t^1 + A_t^1 + A_t^2 = s_{t+1} | S_{T_0+1:t-1} = s_{T_0+1:t-1}, S_t = 1\right)\\
=&\mathbf{P}\left(1 - D_t^1 + A_t^1 + A_t^2 = s_{t+1} | S_t = 1\right)\\
=&\mathbf{P}\left(S_{t+1} = s_{t+1} | S_t = s_t\right). \qquad (178)
\end{aligned}$$

The first equality in (178) follows from the construction of the process $\{S_t, t \geq T_0 + 1\}$. The second and forth equalities follow from the fact that $X_t^{1,\hat{g}}$ can be either 0 or 1. In the third equality, D_t^2 is replaced by D_t^1 in the first term; this is true because D_t^1 and D_t^2 are identically distributed and independent of $X_t^{1,\hat{g}}$ and all past random variables. The fifth equality holds because T_0 is a stopping time with respect to $\{X_t^{1,\hat{g}}, X_t^{2,\hat{g}}, t \in \mathbb{Z}_+\}$ and $A_t^i, D_t^i, i = 1, 2$ are independent of all past random variables. The last equality follows from the same arguments that lead to the first through the fifth equalities.

Therefore, the process $\{S_t, t \geq T_0 + 1\}$ is a Markov chain.
Since $\lambda, \mu > 0$, the Markov chain is irreducible.
We prove that the process $\{S_t, t \geq T_0 + 1\}$ is positive recurrent. Note that, for all $s = 0, 1, 2, \ldots$, because of the construction of $\{S_t, t \geq T_0 + 1\}$

$$\begin{aligned}
&\mathbf{E}\left[S_{t+1} | S_t = s\right]\\
\leq&\mathbf{E}\left[S_t + A_t^1 + A_t^2 | S_t = s\right]\\
=&s + 2\lambda < \infty. \qquad (179)
\end{aligned}$$

Moreover, for all $s \geq 2$,

$$\begin{aligned}
&\mathbf{E}\left[S_{t+1} | S_t = s\right]\\
=&\mathbf{E}\left[s - D_t^1 - D_t^2 + A_t^1 + A_t^2 | S_t = s\right]\\
=&s - 2\mu + 2\lambda < s. \qquad (180)
\end{aligned}$$

Using Foster's theorem (see (Bremaud, 1999, chap. 5)), we conclude that the Markov chain $\{S_t, t \geq T_0 + 1\}$ is positive recurrent.

Proof (Proof of Lemma 8) Let $(\Omega, \mathcal{F}, \mathbf{P})$ denote the basic probability space for our problem. Define events $E_t \in \mathcal{F}, t = 0, 1, \ldots$ to be

$$E_t = \{\omega \in \Omega : \left(U_{t'}^{1,\hat{g}}(\omega), U_{t'}^{2,\hat{g}}(\omega)\right) \neq (0, 0) \quad \forall t' \geq t\} \qquad (181)$$

If the claim of this lemma is not true, we get

$$\mathbf{P}\left(\bigcup_{t=0}^{\infty} E_t\right) = 1 - \mathbf{P}\left(\left(U_t^{1,\hat{g}}, U_t^{2,\hat{g}}\right) = (0, 0) \quad i.o.\right) > 0. \qquad (182)$$

Therefore, there exist some t_0 such that $\mathbf{P}(E_{t_0}) > 0$. Since t_0 is a constant, it is a stopping time with respect to $\{X_t^{1,\hat{g}}, X_t^{2,\hat{g}}, t \in \mathbb{Z}_+\}$.

Consider the process $\{S_t, t = t_0 + 1, t_0 + 2, ...\}$ defined in Lemma 7 with the stopping time t_0. From Lemma 7 we know that $\{S_t, t \geq t_0 + 1\}$ is an irreducible positive recurrent Markov chain. Furthermore, along the sample path induced by any $\omega \in E_{t_0}$, we claim that for all $t \geq t_0 + 1$

$$\begin{aligned} S_t(\omega) =& X_t^{1,\hat{g}}(\omega) + X_t^{2,\hat{g}}(\omega) \\ =& \overline{X}_{t-1}^{1,\hat{g}}(\omega) + \overline{X}_{t-1}^{2,\hat{g}}(\omega). \end{aligned} \tag{183}$$

The claim is shown by induction below.
By the definition of $\{S_t, t \geq t_0 + 1\}$ in Lemma 7, we have at time $t_0 + 1$ for any $\omega \in E_{t_0}$

$$\begin{aligned} S_{t_0+1}(\omega) =& X_{t_0+1}^{1,\hat{g}}(\omega) + X_{t_0+1}^{2,\hat{g}}(\omega) \\ =& \overline{X}_{t_0}^{1,\hat{g}}(\omega) + \overline{X}_{t_0}^{2,\hat{g}}(\omega), \end{aligned} \tag{184}$$

where the last inequality in (184) follows from the system dynamics (1)-(3).
Assume equation (183) is true at time t ($t \geq t_0 + 1$). At time $t + 1$ we have, by (1)-(3),

$$\begin{aligned} & X_{t+1}^{1,\hat{g}} + X_{t+1}^{2,\hat{g}} \\ =& (X_t^{1,\hat{g}} - D_t^1)^+ + (X_t^{2,\hat{g}} - D_t^2)^+ + A_t^1 + A_t^2 \\ =& X_t^{1,\hat{g}} + X_t^{2,\hat{g}} - D_t^1 - D_t^2 + A_t^1 + A_t^2 \\ & + D_t^1 1_{\left\{X_t^{1,\hat{g}}=0\right\}} + D_t^2 1_{\left\{X_t^{2,\hat{g}}=0\right\}}. \end{aligned} \tag{185}$$

Since along the sample path induced by $\omega \in E_{t_0}$, $\left(U_{t-1}^{1,\hat{g}}(\omega), U_{t-1}^{2,\hat{g}}(\omega)\right) \neq (0,0)$ and $X_t^{i,\hat{g}} = \overline{X}_{t-1}^{i,\hat{g}} - U_{t-1}^{i,\hat{g}} + U_{t-1}^{j,\hat{g}}$, the event $\{X_t^{i,\hat{g}} = 0\} \bigcap E_{t_0}$ ($i = 1$ or 2) implies that $\overline{X}_{t-1}^{i,\hat{g}} = 1$, $U_{t-1}^{i,\hat{g}} = 1$ and $U_{t-1}^{j,\hat{g}} = 0$. For this case, $\overline{X}_{t-1}^{i,\hat{g}} = 1$ and $U_{t-1}^{i,\hat{g}} = 1$ further imply that the threshold is smaller than one. Then, the only possibility for $U_{t-1}^{j,\hat{g}} = 0$ is $\overline{X}_{t-1}^{j,\hat{g}} = 0$. Therefore,

$$\begin{aligned} & \left\{X_t^{i,\hat{g}} = 0\right\} \bigcap E_{t_0} \\ \subseteq & \left\{\overline{X}_{t-1}^{i,\hat{g}} = 1,\ U_{t-1}^{i,\hat{g}} = 1,\ \overline{X}_{t-1}^{j,\hat{g}} = 0 \text{ and } U_{t-1}^{j,\hat{g}} = 0\right\} \\ \subseteq & \{S_t = 1\}. \end{aligned} \tag{186}$$

Consequently, from (186), for any $\omega \in E_{t_0}$

$$\begin{aligned} & D_t^1(\omega) 1_{\left\{X_t^{1,\hat{g}}(\omega)=0\right\}} + D_t^2(\omega) 1_{\left\{X_t^{2,\hat{g}}(\omega)=0\right\}} \\ =& 1_{\{S_t(\omega)=1\}} \left(D_t^1(\omega) 1_{\left\{X_t^{1,\hat{g}}(\omega)=0\right\}} + D_t^2(\omega) 1_{\left\{X_t^{2,\hat{g}}(\omega)=0\right\}} \right). \\ =& 1_{\{S_t(\omega)=1\}} \left(1_{\left\{X_t^{1,\hat{g}}(\omega)=0\right\}} (D_t^1(\omega) - D_t^2(\omega)) + D_t^2(\omega) \right). \end{aligned} \tag{187}$$

Moreover, $\left(U_{t-1}^{1,\hat{g}}(\omega), U_{t-1}^{2,\hat{g}}(\omega)\right) \neq (0,0)$ implies that $\left(\overline{X}_{t-1}^{1,\hat{g}}(\omega), \overline{X}_{t-1}^{2,\hat{g}}(\omega)\right) \neq (0,0)$. Hence,

$$S_t(\omega) = \overline{X}_{t-1}^{1,\hat{g}}(\omega) + \overline{X}_{t-1}^{2,\hat{g}}(\omega) \neq 0, \tag{188}$$

and

$$\begin{aligned}
&X_{t+1}^{1,\hat{g}}(\omega) + X_{t+1}^{2,\hat{g}}(\omega) \\
=&X_t^{1,\hat{g}}(\omega) + X_t^{2,\hat{g}}(\omega) - D_t^1(\omega) - D_t^2(\omega) + A_t^1(\omega) + A_t^2(\omega) \\
&+ 1_{\{S_t(\omega)=1\}} \left(1_{\left\{X_t^{1,\hat{g}}(\omega)=0\right\}} (D_t^1(\omega) - D_t^2(\omega)) + D_t^2(\omega) \right) \\
=&X_t^{1,\hat{g}}(\omega) + X_t^{2,\hat{g}}(\omega) - D_t^1(\omega) - D_t^2(\omega) + A_t^1(\omega) + A_t^2(\omega) \\
&+ 1_{\{S_t(\omega)=1\}} \left(1_{\left\{X_t^{1,\hat{g}}(\omega)=0\right\}} (D_t^1(\omega) - D_t^2(\omega)) + D_t^2(\omega) \right) \\
&+ 1_{\{S_t(\omega)=0\}} \left(D_t^1(\omega) + D_t^2(\omega) \right) \\
=&S_{t+1}(\omega), \qquad (189)
\end{aligned}$$

where the first and second equalities in (189) follow from (187) and (188), respectively. The last equality in (189) follows from the construction of $\{S_t, t \geq t_0 + 1\}$.
Furthermore, by the system dynamics (1)-(3) we have

$$\begin{aligned}
\overline{X}_t^{1,\hat{g}}(\omega) + \overline{X}_t^{2,\hat{g}}(\omega) =& X_{t+1}^{1,\hat{g}}(\omega) + X_{t+1}^{2,\hat{g}}(\omega) \\
=& S_{t+1}(\omega). \qquad (190)
\end{aligned}$$

Thus, equation (183) is true for any $\omega \in E_{t_0}$ for all $t \geq t_0 + 1$.
Then, for any $\omega \in E_{t_0}$

$$S_t(\omega) = \overline{X}_{t-1}^{1,\hat{g}}(\omega) + \overline{X}_{t-1}^{2,\hat{g}}(\omega) \neq 0 \text{ for all } t \geq t_0 + 1 \qquad (191)$$

because $\left(U_{t-1}^{1,\hat{g}}(\omega), U_{t-1}^{2,\hat{g}}(\omega)\right) \neq (0,0)$ for all $t \geq t_0 + 1$. Since $\mathbf{P}(E_{t_0}) > 0$, (191) contradicts the fact that $\{S_t, t \geq t_0 + 1\}$ is recurrent.
Therefore, no such event $E_{t_0} \in \mathcal{F}$ with positive probability exists, and the proof of this lemma is complete.

E Proofs of the Results Associated with Step 3 of the Proof of Theorem 2

Proof (Proof of Lemma 9) For any fixed centralized policy $g \in \mathcal{G}_c$, the information I_t^1, I_t^2 available to the centralized controller includes all primitive random variables $X_0^i, A_{0:t}^i, D_{0:t}^i, i = 1, 2$ up to time t. Since all other random variables are functions of these primitive random variables and g, we have

$$\begin{aligned}
U_t^{i,g} =& g_t^i(I_t^1, I_t^2) \\
=& g_t^i(X_0^1, X_0^2, A_{0:t}^1, A_{0:t}^2, D_{0:t}^1, D_{0:t}^2), \qquad (192)
\end{aligned}$$

for $i = 1, 2$. For any initial queue lengths x_0^1, x_0^2, we now define a policy $\tilde{g}$ from g for the case when both queues are initially empty. Let $\tilde{g}$ be the policy such that for $i = 1, 2$

$$\begin{aligned}
U_t^{i,\tilde{g}} =& \tilde{g}_t^i(I_t^1, I_t^2) \\
:=& \begin{cases} g_t^i(x_0^1, x_0^2, A_{0:t}^1, A_{0:t}^2, D_{0:t}^i, D_{0:t}^2) & \text{if } \overline{X}_t^{i,\tilde{g}} > 0 \\ 0 & \text{if } \overline{X}_t^{i,\tilde{g}} = 0 \end{cases} \\
=& \min\left(U_t^{i,g}, \overline{X}_t^{i,\tilde{g}}\right) \leq U_t^{i,g}, \qquad (193)
\end{aligned}$$

where $X_t^{1,\tilde{g}}$ and $X_t^{2,\tilde{g}}$ denote the queue lengths at time t due to policy $\tilde{g}$ with initial queue lengths $X_0^{1,\tilde{g}} = X_0^{2,\tilde{g}} = 0$.

At time 0 we have $X_0^{i,g} = x_0^i \geq 0 = X_0^{i,\tilde{g}}$ for $i = 1, 2$. We now prove by induction that for all time t

$$X_t^{i,g} \geq X_t^{i,\tilde{g}}, \quad i = 1, 2. \tag{194}$$

Suppose the claim is true at time t. Then, from the system dynamics (1)-(2) and (194) we obtain, for $i = 1, 2$,

$$\begin{aligned} \overline{X}_t^{i,g} &= \left(X_t^{i,g} - D_t^i\right)^+ + A_t^i \\ &\geq \left(X_t^{i,\tilde{g}} - D_t^i\right)^+ + A_t^i = \overline{X}_t^{i,\tilde{g}}. \end{aligned} \tag{195}$$

Furthermore from (1)-(2) and (193)

$$\begin{aligned} X_{t+1}^{i,g} =& \overline{X}_t^{i,g} - U_t^{i,g} + U_t^{j,g} \\ \geq& \overline{X}_t^{i,g} - U_t^{i,g} + U_t^{j,\tilde{g}} \end{aligned} \tag{196}$$

If $\overline{X}_t^{i,\tilde{g}} > 0$, then, because of (193) and (195)

$$\begin{aligned} \overline{X}_t^{i,g} - U_t^{i,g} =& \overline{X}_t^{i,g} - \min\left(U_t^{i,g}, \overline{X}_t^{i,\tilde{g}}\right) \\ =& \overline{X}_t^{i,g} - U_t^{i,\tilde{g}} \geq \overline{X}_t^{i,\tilde{g}} - U_t^{i,\tilde{g}}. \end{aligned} \tag{197}$$

If $\overline{X}_t^{i,\tilde{g}} = 0$, since $\overline{X}_t^{i,g} - U_t^{i,g} \geq 0$, (193) implies

$$\overline{X}_t^{i,g} - U_t^{i,g} \geq 0 = \overline{X}_t^{i,\tilde{g}} - U_t^{i,\tilde{g}}. \tag{198}$$

Combining (196)-(198) and (1)-(2) we get

$$\begin{aligned} X_{t+1}^{i,g} \geq& \overline{X}_t^{i,g} - U_t^{i,g} + U_t^{j,\tilde{g}} \\ \geq& \overline{X}_t^{i,\tilde{g}} - U_t^{i,\tilde{g}} + U_t^{j,\tilde{g}} = X_{t+1}^{i,\tilde{g}}. \end{aligned} \tag{199}$$

Therefore, we complete the proof of the claim (194).
Since the cost function is increasing, (194) implies that for all $g \in \mathcal{G}_c$ and any initial condition $X_0^1 = x_0^1, X_0^2 = x_0^2$,

$$\inf_{g \in \mathcal{G}_c} J_T^g(0,0) \leq J_T^{\tilde{g}}(0,0) \leq J_T^g(x_0^1, x_0^2). \tag{200}$$

Consequently, for any PMFs π_0^1, π_0^2

$$\inf_{g \in \mathcal{G}_c} J_T^g(0,0) \leq \inf_{g \in \mathcal{G}_c} J_T^g(\pi_0^1, \pi_0^2). \tag{201}$$

Moreover, the result of Lemma 3 ensures that $\hat{g}$ gives the smallest expected cost among policies in $\mathcal{G}_c$ for any finite horizon when $X_0^1 = X_0^2 = 0$. It follows that, for any finite T,

$$J_T^{\hat{g}}(0,0) = \inf_{g \in \mathcal{G}_c} J_T^g(0,0) \leq J_T^{\tilde{g}}(0,0) \leq J_T^g(x_0^1, x_0^2). \tag{202}$$

For infinite horizon cost, we divide each term in (202) by T and let T to infinity, and we obtain, for any π_0^1, π_0^2,

$$J^{\hat{g}}(0,0) = \inf_{g \in \mathcal{G}_c} J^g(0,0) \leq J^{\tilde{g}}(0,0) \leq J^g(x_0^1, x_0^2). \tag{203}$$

F Proofs of the Results Associated with Step 4 of the Proof of Theorem 2

Proof (Proof of the claim in the proof of Theorem 2)
We prove here our claim expressed by equation (90) to complete the proof of Theorem 2. By (78),

$$S_{T_0+1} = X^{1,\hat{g}}_{T_0+1} + X^{2,\hat{g}}_{T_0+1}. \tag{204}$$

We prove by induction that $X^{1,\hat{g}}_t + X^{2,\hat{g}}_t = S_t$ for all $t \geq T_0 + 1$.
Assume that $X^{1,\hat{g}}_t + X^{2,\hat{g}}_t = S_t$ at time t, $t \geq T_0 + 1$. Then for time $t+1$, because of the systems dynamics (1)-(3),

$$\begin{aligned}
&X^{1,\hat{g}}_{t+1} + X^{2,\hat{g}}_{t+1} \\
=&(X^{1,\hat{g}}_t - D^1_t)^+ + (X^{2,\hat{g}}_t - D^2_t)^+ + A^1_t + A^2_t \\
=&X^{1,\hat{g}}_t + X^{2,\hat{g}}_t - D^1_t - D^2_t + A^1_t + A^2_t \\
&+ D^1_t 1_{\left\{X^{1,\hat{g}}_t = 0\right\}} + D^2_t 1_{\left\{X^{2,\hat{g}}_t = 0\right\}}.
\end{aligned} \tag{205}$$

When $X^{i,\hat{g}}_t = 0$ ($i = 1$ or 2), $U^{j,\hat{g}}_{t-1}$ should be 0 because

$$0 = X^{i,\hat{g}}_t = \overline{X}^{i,\hat{g}}_{t-1} - U^{i,\hat{g}}_{t-1} + U^{j,\hat{g}}_{t-1} \tag{206}$$

and $\overline{X}^{i,\hat{g}}_{t-1} - U^{i,\hat{g}}_{t-1} \geq 0$.
We consider the following two cases separately:

Case 1 $U^{i,\hat{g}}_{t-1} = 0$.

Case 2 $U^{i,\hat{g}}_{t-1} = 1$.

Case 1 When $U^{i,\hat{g}}_{t-1} = 0$, we must have $\overline{X}^{i,\hat{g}}_{t-1} = 0$ by (206). Then $\overline{X}^{j,\hat{g}}_{t-1} \in \{0, 1\}$ for the following reason. When $U^{i,\hat{g}}_{t-1} = U^{j,\hat{g}}_{t-1} = 0$, the sizes of both queues are between the lower bound and the threshold. That is

$$\overline{LB}^{\hat{g}}_{t-1} \leq \overline{X}^{i,\hat{g}}_{t-1} \leq \lceil TH_t \rceil - 1, \tag{207}$$

$$\overline{LB}^{\hat{g}}_{t-1} \leq \overline{X}^{j,\hat{g}}_{t-1} \leq \lceil TH_t \rceil - 1. \tag{208}$$

Combining (207), (208) with $\overline{X}^{i,\hat{g}}_{t-1} = 0$ we obtain

$$\begin{aligned}
\overline{X}^{j,\hat{g}}_{t-1} =& \left|\overline{X}^{j,\hat{g}}_{t-1} - \overline{X}^{i,\hat{g}}_{t-1}\right| \\
\leq& \lceil TH_t \rceil - 1 - \overline{LB}^{\hat{g}}_{t-1} \\
\leq& \frac{1}{2}\left(\overline{UB}^{\hat{g}}_{t-1} - \overline{LB}^{\hat{g}}_{t-1}\right) \leq 1.5,
\end{aligned} \tag{209}$$

where the last inequality in (209) is true because of (31) in Lemma 1, (89), and

$$\overline{UB}^{\hat{g}}_{t-1} - \overline{LB}^{\hat{g}}_{t-1} \leq UB^{\hat{g}}_t + 1 - LB^{\hat{g}}_t + 1 \leq 3.$$

Therefore, $\overline{X}^{j,\hat{g}}_{t-1} \leq 1$ because $\overline{X}^{j,\hat{g}}_{t-1}$ takes integer values.

Case 2 When $U^{i,\hat{g}}_{t-1} = 1$, we must have $\overline{X}^{i,\hat{g}}_{t-1} = 1$ by (206). This implies that the threshold is not more than 1, and the only possible value of $\overline{X}^{j,\hat{g}}_{t-1}$ less than the threshold is 0.

As a consequence of the above analysis for the cases 1 and 2, $\{X_t^{i,\hat{g}} = 0\}$ implies

$$S_t = \overline{X}_{t-1}^{i,\hat{g}} + \overline{X}_{t-1}^{j,\hat{g}} \leq 1. \tag{210}$$

Thus, for $i = 1, 2$,

$$\left\{X_t^{i,\hat{g}} = 0\right\} = \left\{X_t^{i,\hat{g}} = 0, S_t \leq 1\right\}. \tag{211}$$

Then,

$$\begin{aligned}
&D_t^1 1_{\left\{X_t^{1,\hat{g}}=0\right\}} + D_t^2 1_{\left\{X_t^{2,\hat{g}}=0\right\}} \\
=&D_t^1 1_{\left\{X_t^{1,\hat{g}}=0, S_t\leq 1\right\}} + D_{t+1}^2 1_{\left\{X_t^{2,\hat{g}}=0, S_t\leq 1\right\}} \\
=&D_t^1 1_{\left\{X_t^{1,\hat{g}}=0, S_t=1\right\}} + D_t^2 1_{\left\{X_t^{1,\hat{g}}\neq 0, S_t=1\right\}} \\
&+ D_t^1 1_{\{S_t=0\}} + D_t^2 1_{\{S_t=0\}}.
\end{aligned} \tag{212}$$

Combining (205) and (212) we obtain

$$\begin{aligned}
&X_{t+1}^{1,\hat{g}} + X_{t+1}^{2,\hat{g}} \\
=&X_t^{1,\hat{g}} + X_t^{2,\hat{g}} - D_t^1 - D_t^2 + A_t^1 + A_t^2 \\
&+ D_t^1 1_{\{X_t^1=0, S_t=1\}} + D_t^2 1_{\{X_t^1\neq 0, S_t=1\}} \\
&+ D_t^1 1_{\{S_t=0\}} + D_t^2 1_{\{S_t=0\}} \\
=&S_{t+1},
\end{aligned} \tag{213}$$

where the last equality follows by the definition of S_{t+1}.
Therefore, at any time $t \geq T_0 + 1$ we have

$$X_t^{1,\hat{g}} + X_t^{2,\hat{g}} = S_t. \tag{214}$$

The proof of claim (90), and consequently, the proof of Theorem 2 is complete.

Acknowledgements This work was partially supported by National Science Foundation (NSF) Grant CCF-1111061 and NASA grant NNX12AO54G. The authors thank Mark Rudelson and Aditya Mahajan for helpful discussions.

References

Abdollahi F, Khorasani K (2008) A novel H_∞ control strategy for design of a robust dynamic routing algorithm in traffic networks. IEEE Journal on Selected Areas in Communications 26(4):706–718

Akgun OT, Righter R, Wolff R (2012) Understanding the marginal impact of customer flexibility. Queueing Systems 71(1-2):5–23

Aumann RJ (1976) Agreeing to disagree. The Annals of Statistics pp 1236–1239

Beutler FJ, Teneketzis D (1989) Routing in queueing networks under imperfect information: Stochastic dominance and thresholds. Stochastics and Stochastic Reports 26(2):81–100

Bremaud P (1999) Markov chains: Gibbs fields, Monte Carlo simulation, and queues, vol 31. springer

Cogill R, Rotkowitz M, Van Roy B, Lall S (2006) An approximate dynamic programming approach to decentralized control of stochastic systems. In: Control of Uncertain Systems: Modelling, Approximation, and Design, Springer, pp 243–256

Davis E (1977) Optimal control of arrivals to a two-server queueing system with separate queues. PhD thesis, PhD dissertation, Program in Operations Research, North Carolina State University, Raleigh, NC

Ephremides A, Varaiya P, Walrand J (1980) A simple dynamic routing problem. IEEE Transactions on Automatic Control 25(4):690–693

Foley RD, McDonald D (2001) Join the shortest queue: stability and exact asymptotics. Annals of Applied Probability pp 569–607

Hajek B (1984) Optimal control of two interacting service stations. IEEE Transactions on Automatic Control 29(6):491–499

Ho YC (1980) Team decision theory and information structures. Proceedings of the IEEE 68(6):644–654

Hordijk A, Koole G (1990) On the optimality of the generalized shortest queue policy. Probability in the Engineering and Informational Sciences 4(4):477–487

Hordijk A, Koole G (1992) On the assignment of customers to parallel queues. Probability in the Engineering and Informational Sciences 6(04):495–511

Kuri J, Kumar A (1995) Optimal control of arrivals to queues with delayed queue length information. IEEE Transactions on Automatic Control 40(8):1444–1450

Lin W, Kumar P (1984) Optimal control of a queueing system with two heterogeneous servers. IEEE Transactions on Automatic Control 29(8):696–703

Mahajan A (2013) Optimal decentralized control of coupled subsystems with control sharing. IEEE Transactions on Automatic Control 58(9):2377–2382, DOI 10.1109/TAC.2013.2251807

Manfredi S (2014) Decentralized queue balancing and differentiated service scheme based on cooperative control concept. IEEE Transactions on Industrial Informatics 10(1):586–593

Marshall A, Olkin I, Arnold B (2010) Inequalities: theory of majorization and its applications. Springer Verlag

Menich R, Serfozo RF (1991) Optimality of routing and servicing in dependent parallel processing systems. Queueing Systems 9(4):403–418

Nayyar A, Mahajan A, Teneketzis D (2013) Decentralized stochastic control with partial history sharing: A common information approach. IEEE Transactions on Automatic Control 58(7):1644–1658, DOI 10.1109/TAC.2013.2239000

Ouyang Y, Teneketzis D (2013) A routing problem in a simple queueing system with non-classical information structure. In: Proc. 51th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, pp 1278 – 1284

Ouyang Y, Teneketzis D (2014) Balancing through signaling in decentralized routing. In: Proc. 53rd Conference on Decision and Control, Los Angeles, CA, accepted

Pandelis DG, Teneketzis D (1996) A simple load balancing problem with decentralized information. Mathematical Methods of Operations Research 44(1):97–113

Petersen KE, Petersen K (1989) Ergodic theory, vol 2. Cambridge University Press

Reddy AA, Banerjee S, Gopalan A, Shakkottai S, Ying L (2012) On distributed scheduling with heterogeneously delayed network-state information. Queueing Systems 72(3-4):193–218

Si X, Zhu XL, Du X, Xie X (2013) A decentralized routing control scheme for data communication networks. Mathematical Problems in Engineering 2013, article ID 648267

Weber RR (1978) On the optimal assignment of customers to parallel servers. Journal of Applied Probability pp 406–413

Weber RR, Stidham Jr S (1987) Optimal control of service rates in networks of queues. Advances in applied probability pp 202–218

Whitt W (1986) Deciding which queue to join: Some counterexamples. Operations research 34(1):55–62

Winston W (1977) Optimality of the shortest line discipline. Journal of Applied Probability pp 181–189

Witsenhausen HS (1971) Separation of estimation and control for discrete time systems. Proceedings of the IEEE 59(11):1557–1566

Ying L, Shakkottai S (2011) On throughput optimality with delayed network-state information. IEEE Transactions on Information Theory 57(8):5116–5132