Dynamic control of a tandem system with abandonments

Gabriel Zayas-Cabán[1], Jingui Xie[2], Linda V. Green[3], and Mark E. Lewis[4]

[1]*Center for Healthcare Engineering and Patient Safety*

University of Michigan

Ann Arbor, MI

gzayasca@umich.edu

[2]*The School of Management*

University of Science and Technology of China

Hefei, China

xiej@ustc.edfu.cn

[3]*Graduate School of Business*

Columbia University

New York, NY

lvg1@columbia.edu

[4]*School of Operations Research and Information Engineering*

Cornell University

Ithaca, NY

mark.lewis@cornell.edu

January 4, 2016

Abstract

The goal of this paper is to provide a model that is an extension of classic scheduling problems for a tandem queueing system by including customer impatience. In such scenarios, the server(s) must balance abandonments from each phase of service with the need to prioritize higher reward customers. This presents an interesting challenge since the trade-off between the cost of abandonments and revenue maximization are not at all clear.

As a result of customer abandonments, uniformization is not possible since the transition rates are unbounded. We do our analysis in continuous time, using the continuous time Markov decision process framework to discern simple relationships of the value functions depending on the starting state. We then use sample path arguments to analyze the structure of the optimal policy. We provide conditions under which an optimal policy is non-idling and never splits servers (except to avoid idling). We then consider the single-server model as a proxy for the decision problem when there are more customers at each station than there are servers. In this case we provide conditions under which it is optimal to prioritize each station.

1 Introduction

For many systems, service consists of two or more phases by one or more servers. A fundamental decision that a flexible server (or servers) must make is at which phase should the server direct its effort? This question has been considered by many authors for different cost or reward criterion but without abandonments (c.f [10, 5, 1]). The goal of this paper is to provide a model that is an extension of these classic scheduling problems by including customer impatience. In such scenarios, the server(s) must now also balance abandonments from each phase of service. This presents an interesting challenge since the trade-off between the cost of abandonments and revenue maximization are not at all clear.

The present study is motivated by the third author's work with the emergency department (ED) at the Lutheran Medical Center (LMC) in Brooklyn, New York where a new model of care known as the "Triage-Treat-and-Release" (TTR) program has recently been implemented. Traditionally, nurses in the ED are solely responsible for triaging patients while doctors are solely responsible for treating patients. In contrast to this traditional set-up, the TTR program makes physician assistants or nurse practitioners responsible for both phases of service for low-acuity patients. Providers in this setting must decide between prioritizing triage versus treatment in order to balance initial delays for care with the need to discharge patients in a timely fashion. Alternative interpretations of the tandem queueing system with abandonments in practice are in manufacturing with order cancellations and in call centers (including web chat) with customer impatience.

We formulate the server allocation problem as a continuous-time Markov decision process (CTMDP) and provide sufficient conditions for when it is optimal to prioritize phase 1 and phase 2 service. To the best our best knowledge, this is the first paper to develop a framework for studying allocation policies for flexible servers in a two-stage tandem queue with abandonments from both phases of service, and see our work as having the following contributions in this area:

1. As a result of customer abandonments, uniformization (c.f. [16]) is not possible since the transition rates are unbounded. We do our analysis in continuous time, using the continuous time Markov decision process framework to discern simple relationships of the value functions depending on the starting state. We then use sample path arguments to analyze the structure of the optimal policy for each reward criterion.

2. In the multi-server setting we show that the optimal control ensures that all servers stay busy. That is to say, when there are not enough customers to serve at one station, they work at the other.

3. We show that when there is enough work to do, servers need not be split between stations. This justifies the single-server proxy when there are enough jobs in the system to keep servers busy.

4. In the single server proxy, we provide sufficient conditions for when it is optimal to prioritize phase 1 and phase 2 service. In the discounted case, these sufficient conditions are akin to the classic c-μ rule (cf. [3]).

From an analytical perspective, our model is closely related to the analysis and control of service policies for flexible servers in tandem queueing systems. Among those that analyze the performance of single server tandem queues, [17] considers one with Poisson arrivals and general service times under a non-zero switching rule for phase-one and a zero switching rule for phase two. A non-zero switching rule is one in which the server continues to serve in a phase until some specified number of consecutive services have been completed and then switches to the other phase, while a zero switching rule continues to serve until the phase is empty before switching to the other phase. Taube-Netto [20] considers the same model but with a zero switching rule at each phase. Katayama [12] analyzes the system under a zero switching rule at each phase but with non-zero switchover times. A K-limited policy has the server visit a phase and continue serving that phase until either it is empty or K customers are served, whichever occurs first. A gated service policy is one in which, once the server switches phases, (s)he serves only customers who are in that phase at the time immediately following the switch. In [13] and [14], Katayama extends his previous work to consider gated service and K-limited service disciplines, respectively. In both papers, intermediate finite waiting room is allowed. Katayama [15] analyze the sojourn time under general K-decrementing service policies; policies in which once the server visits phase one, (s)he continues serving that phase until either this phase becomes empty or K customers are served, whichever occurs first, and then serves at phase-two until it is empty.

Work on optimal service disciplines for tandem queueing systems is too plentiful to provide a complete review here. For earlier work the readers are pointed to [18, 11, 10, 5] and references therein. More recently, [1] consider optimal service disciplines for two flexible servers in a two-stage tandem queueing system assuming Poisson arrivals and exponential service times. They consider a collaborative scenario and a non-collaborative scenario. In the former, servers can collaborate to work on the same job, which is similar to the single server case. They provide sufficient conditions under which it is optimal to allocate both servers to phase 1 or 2 in the collaborative case. In [2] the authors consider policies that maximize throughput of a tandem queue with a finite intermediate buffer. Motivated by ED patient flow problems similar to ours, [9], [4], and [19] use queueing models, control techniques, and/or simulation to determine when physicians in EDs should prioritize patients at different stages of service. To the best of our knowledge, our work is the first to consider optimal service policies for servers in a two-phase stochastic service system in which customers can abandon before receiving service.

The rest of this paper is organized as follows. Section 2 describes the queueing dynamics in detail and provides a Markov decision process formulation of the problem under both the discounted expected reward and the long run average reward optimality criteria. We also show that it is enough to consider policies that do not idle the servers whenever there are customers waiting to be served. Section 3 contains our main theoretical results. We give sufficient conditions for when to prioritize each phase of service under each reward criteria. We discuss possible extensions to our work and conclude in Section 4.

2 Model Description and Preliminaries

Suppose customers arrive to a tandem service system according to a Poisson process of rate λ and immediately join the first queue. Upon arrival, a customer's station 1 (random and hidden) patience time is generated. After receiving service at station 1 a reward of $R_1 > 0$ is accrued. Independent of the service time and arrival process, with probability p the customer joins the queue at station 2. With probability $q := 1 - p$, the customer leaves the system forever. If the customer joins station 2 his/her patience time is generated. Assume that the service times at station i are exponential with rate $\mu_i > 0$ and the abandonment time is exponential with rate β_i, $i = 1, 2$. If the customer does not complete service before the abandonment time ends, the customer leaves the system without receiving service at the second station. If service is completed at station 2, a reward $R_2 > 0$ is accrued. There are $N \geq 1$ servers, each of which can be assigned to either station. After each event (arrival, service completion or abandonment) the server(s) view the number of customers at each station and decides where to serve next.

To model the two-phase decision problem we consider a *Markov decision process* (MDP) formulation. Let $\{t_n, n \geq 1\}$ denote the sequence of event times that includes arrivals, abandonments and potential service completions. Define the state space $\mathbb{X} := \{(i, j) | i, j \in \mathbb{Z}^+\}$, where i (j) represents the number of customers at station 1 (2). The available actions in state $x = (i, j)$ are

$$A(x) = \{(n_1, n_2) | n_1, n_2 \in \mathbb{Z}^+, n_1 + n_2 \leq N\},$$

where n_1 (n_2) represents the number of servers assigned to station 1 (2). A policy prescribes how many servers should be allocated to stations 1 and 2 for all states for all time. The function

$r(x,a)$ (where $a \in A(x)$) is the expected reward function

$$r((i,j),(n_1,n_2)) = \frac{\min\{i,n_1\}\mu_1 R_1}{\lambda + \min\{i,n_1\}\mu_1 + i\beta_1 + j\beta_2} + \frac{\min\{j,n_2\}\mu_2 R_2}{\lambda + \min\{j,n_2\}\mu_2 + i\beta_1 + j\beta_2}.$$

Under the α-discounted expected reward criterion the value of the policy f given that the system starts in state (i,j) over the horizon of length t is given by

$$v^f_{\alpha,t}(i,j) = \mathbb{E}_{(i,j)} \sum_{n=1}^{N(t)} e^{-\alpha t_n} \big[r(X(t_n-), f(X(t_n-))) \big],$$

where $N(t)$ is the counting process that counts the number of decision epochs in the first t time units, and $\{X(s), s \geq 0\}$ is the continous-time Markov chain denoting the state of the system (i.e. the number of customers at each queue) at time s. The infinite horizon $\alpha-$discounted reward of the policy is $v^f_\alpha(i,j) := \lim_{t\to\infty} v^f_{\alpha,t}(i,j)$ and the optimal reward is $v_\alpha(i,j) := \max_{\pi\in\Pi} v^\pi_\alpha(i,j)$, where Π is the set of all non-anticipating policies. Similarly, for the average reward case, the average reward of the policy f is $\rho^f(i,j) := \liminf_{t\to\infty} \frac{v^f_{0,t}(i,j)}{t}$.

Note that when the system is in state (i,j) the abandonment rate is $i\beta_1 + j\beta_2$. In short, customers that are in service are allowed to abandon. This is a realistic assumption for "in-service" models where the server can leave and return to the customer during service (web chats, placing the customer on hold, etc.). It also serves as a simpler formulation that approximates the case when customers do not abandon during service.

Next we note that ignoring the abandonments in either case can lead to making the completely wrong decision in terms of where the server should allocate its time. Consider the following two examples.

Example 2.1 *Suppose we have the following inputs for the model:* $N = 1$*;* $p = 1$*;* $\lambda = 3$*;* $\mu_1 = 60/7$*;* $\mu_2 = 60/13$*;* $\beta_1 = 0$*;* $\beta_2 = 0.3$*;* $R_1 = 15$*;* $R_2 = 20$*. In this case, the average rewards for prioritizing station* 1 *(2) are given by 92 (104).*

Example 2.2 *Suppose we have the following inputs for the model:* $N = 1$*;* $p = 1$*;* $\lambda = 3$*;* $\mu_1 = 60/7$*;* $\mu_2 = 60/13$*;* $\beta_1 = 0.3$*;* $\beta_2 = 0$*;* $R_1 = 20$*;* $R_2 = 10$*. In this case, the average rewards for prioritizing station* 1 *(2) are given by 84 (75).*

In Example 2.1 we note that $\mu_1 R_1 > \mu_2 R_2$ so that if $\beta_2 = 0$ we should prioritize station 1. However, doing so results in average rewards that are more than 11.5% below optimal. Similarly, in Example 2.2 we have that $\mu_1 R_1 < \mu_2 R_2$ so that $\beta_1 = 0$ implies that the second station would be prioritized. Prioritizing station 2 yields average rewards 10.6% below optimal. The examples illustrate that ignoring the abandonments, can result in policies that prioritize precisely the wrong station, and lead to a significant loss in average rewards.

Our first result states that there is an optimal policy that does not idle the server whenever there are customers waiting. This is used to simplify the optimality equations that follow.

Proposition 2.3 *Under the $\alpha-$discounted reward (finite or infinite horizon) or the average reward criterion, there exists a (Markovian) non-idling policy that is optimal.*

Proof. We prove the result using a sample path argument. Suppose we start two processes in the same state (i, j), and on the same probability space so that all of the (potential) events coincide. Moreover, each customer that enters the system has attached to it patience times and a mark indicating whether or not it will leave the system after completing service at station 1 or continue on to station 2. Of course, they are all unbeknownst to the decision-maker at the time of the arrival. Fix time $t \geq 0$ and assume $\alpha > 0$. Process 1 uses a policy π_1 that idles servers unnecessarily until the first event and uses an optimal policy thereafter. To make matters concrete, assume π_1 idles worker(s) at station 1 (the case where it idles workers at station 2 is analogous) when $i + j \geq N$. In what follows, we show how to construct a (potentially sub-optimal) policy for Process 2, that we denote by π_2, that assigns the extra servers to station 1 until the first event and satisfies

$$v_{\alpha,t}^{\pi_2}(i, j) \geq v_{\alpha,t}^{\pi_1}(i, j). \tag{2.1}$$

Since π_2 is not required to be optimal and π_1 uses the optimal actions after the first event, the result follows.

Suppose that in the first event both processes see an arrival, a service completion, or an abandonment. In all cases, the two processes transition to the same state. At this point let $\pi_2 = \pi_1$ so that the two processes couple. It follows that if an arrival, a service completion seen by the two processes, or an abandonment occurs first, then (2.1) holds with equality.

Suppose now that the first event is a service completion at station 1 in Process 2 that is not seen by Process 1. In this case, Process 1 remains in state (i, j) while Process 2 transitions to either state $(i-1, j+1)$ or state $(i-1, j)$ (depending on the mark of customer 1) and accrues a reward of R_1. From this point forward, if we reach time t, before another event occurs, (2.1) holds strictly since Process 2 has seen an extra reward. Otherwise, let policy π_2 choose exactly the same action that policy π_1 chooses at every subsequent decision epoch except in the case that it cannot assign the same number of servers to station 1 (since it now has one less customer at station 1 than Process 1). In this case, it assigns the same number to station 2 as Process 1 and one less at station 1 than Process 1 (idling the extra server). Since each customer in queue for Process 1 is a replica of each customer in Process 2 (including the one just served), it follows from the description of π_2 that Process 1 can see at most one service completion at station 1 that is not seen at station 2, at which point, the two processes couple. Discounting in this case implies that (2.1) holds strictly. Continuing in this fashion for all initial states yields the result for fixed t.

Note that the results hold *pathwise* for any t. This implies that the result holds as $t \to \infty$, and hence, for the infinite horizon discounted case. Moreover, the discounting is only used to show that (2.1) holds strictly. Setting $\alpha = 0$ and repeating the proof yields the result for the average case. ■

A typical method for obtaining the optimal value and control policy under either the discounted or average reward criterion is to use the dynamic programming optimality equations. In theory (since the state space is countably infinite) this is still possible. On the other hand, it is also quite common to use the optimality equations to obtain the structure of an optimal control in hopes that said structure is simple enough to be easily implementable. We pursue this direction with the caveat that none of the typical methods for using the optimality equations (successive approximations, action elimination) work directly since the transition rates are not bounded. This means that the problem is not uniformizable and no discrete-time equivalent exists. Instead, we rely on the optimality equations to simplify our search for optimal policies. Proposition 2.3 implies that we may restrict our attention to non-idling policies thereby simplifying the search for the optimal control policy. We next provide conditions under which the optimality equations have a solution. To simplify notation for a function h on $\mathbb{X}$ define the following mapping

$$\begin{aligned} Th(i,j) = {} & \lambda h(i+1,j) + i\beta_1 h(i-1,j) + j\beta_2 h(i,j-1) - (\lambda + i\beta_1 + j\beta_2)h(i,j) \\ & + \max_{a \in \{0,1,\cdots,N\}} \Big(\min\{i,a\}\mu_1\big(R_1 + ph(i-1,j+1) + qh(i-1,j) \\ & - h(i,j)\big) + \min\{j, N-a\}\mu_2\big(R_2 + h(i,j-1) - h(i,j)\big)\Big). \end{aligned} \tag{2.2}$$

The next result provides the discounted and average reward optimality equations and conditions under which they have a solution. The proof for the average case is provided in the Appendix.

Theorem 2.4 *The MDP has the following properties:*

1. *Fix $\alpha > 0$. Under the $\alpha-$discounted expected reward criterion:*
   - *(a) The value function v_α satisfies the discounted reward optimality equations (DROE),*

     $$\alpha v_\alpha = T v_\alpha. \tag{2.3}$$

   - *(b) There exists a deterministic stationary optimal policy.*
   - *(c) Any f satisfying the maximum in the DROE defines a stationary optimal policy.*

2. *Under the average reward criterion:*

   (a) *There exists a constant g and function w on the state space such that (g, w) satisfies the average reward optimality equations (AROE),*

   $$g\mathbf{1} = Tw,$$

   where $\mathbf{1}$ is the vector of ones.

   (b) *There exists a deterministic stationary optimal policy.*

   (c) *Any f satisfying the maximum in the AROE defines a stationary optimal policy and $g = \rho(x)$ for all x.*

Proof. See Appendix A.1. ∎

3 Dynamic Control

The next result shows that we need only consider policies that never divide the servers between queues whenever the number of customers at each queue exceeds the number of servers.

Proposition 3.1 *If the number of customers at each queue exceeds the number of servers, there exists a discounted reward optimal control policy that does not split the servers. Similarly in the average reward case.*

Proof. We show the result in the discounted expected reward case. The average reward case is analogous. Assume that $i, j \geq N$, and hence, $\min\{i, a\} = a$ and $\min\{j, N-a\} = N-a$ for any a in the action set $A(i, j) = \{0, 1, \cdots, N\}$. Let

$$\begin{aligned} G_\alpha(i, j, a) := a\mu_1\big(R_1 + pv_\alpha(i-1, j+1) + qv_\alpha(i-1, j) - v_\alpha(i, j)\big) \\ + (N-a)\mu_2\big(R_2 + v_\alpha(i, j-1) - v_\alpha(i, j)\big), \end{aligned}$$

and note that $G_\alpha(i, j, a)$ is linear in a. Since $\max_{a\in\{0,1,\ldots,N\}}[G_\alpha(i, j, a)]$ defines the optimal action in (i, j) and a linear function achieves its maximum (and minimum) at the extreme points, it follows that it is optimal to set a equal to 0 or N, the two extreme points in $\{0, 1, \cdots, N\}$. ∎

Proposition 3.1 implies that we may restrict attention to policies that always allocate all workers to one station or the other in those states where $i, j \geq N$ (making the service rate $N\mu_1$ or $N\mu_2$). Similarly, Proposition 2.3 implies that we should keep as many servers busy as possible in states such that $i + j \leq N$. We are left without any guidance in states with i

or j (but not both) greater than N, nor how to choose the priorities when $i, j \geq N$. We point out that solving the problem generally is quite difficult and we simply were unable to do so. Instead, we discuss under what conditions simple priority policies are optimal in the single server model (see Section 3.1) and use it as a proxy for the multi-server model.

3.1 The single server proxy

In this section, we restrict attention to the single server model ($N = 1$). Using this model, in the next section we discuss when priority rules are optimal. To simplify notation in the single server case, for a function h on $\mathbb{X}$ redefine the mapping T,

$$
\begin{aligned}
Th(0, j) &= \lambda h(1, j) + j\beta_2 h(0, j-1) - (\lambda + j\beta_2)h(0, j) \\
&\quad + \mathbb{1}(j \geq 1)\mu_2[R_2 + h(0, j-1) - h(0, j)] \\
Th(i, 0) &= \lambda h(i+1, 0) + i\beta_1 h(i-1, 0) - (\lambda + i\beta_1)h(i, 0) \\
&\quad + \mathbb{1}(i \geq 1)\mu_1[R_1 + ph(i-1, 1) + qh(i-1, 0) - h(i, 0)].
\end{aligned}
$$

For $i, j \geq 1$

$$
\begin{aligned}
Th(i, j) = \lambda h(i+1, j) &+ i\beta_1 h(i-1, j) + j\beta_2 h(i, j-1) - (\lambda + i\beta_1 + j\beta_2)h(i, j) \\
&+ \max\{\mu_1[R_1 + ph(i-1, j+1) + qh(i-1, j) - h(i, j)], \\
&\quad \mu_2[R_2 + h(i, j-1) - h(i, j)]\}.
\end{aligned}
$$

Theorem 2.4 implies, for example, that in the discounted rewards model, for $i, j \geq 1$, it is optimal to serve a class 1 (2) customer if $\mu_1[R_1 + pv_\alpha(i-1, j+1) + qv_\alpha(i-1, j) - v_\alpha(i, j)] \geq (\leq)\mu_2[R_2 + v_\alpha(i, j-1) - v_\alpha(i, j)]$, with a direct analogue in the average case. We conclude this section with a tacit statement of the monotonicity of the value functions. In the interest of brevity (and because the proof is simple, but lengthy) we omit the proof.

Proposition 3.2 *Under the $\alpha-$discounted reward criterion the following hold for all $i, j \geq 0$,*

1. *$v_\alpha(i+1, j) \geq v_\alpha(i, j)$ and*
2. *$v_\alpha(i, j+1) \geq v_\alpha(i, j)$.*

Similarly, if (g, w) is a solution to the average reward optimality equations, the above statements hold with v_α replaced with w.

3.1.1 Priority Rules

Recall, we seek easily implementable policies. In this section, we provide conditions under which priority rules are optimal.

Theorem 3.3 *Suppose $\beta_1 = 0, \beta_2 > 0$. In this case, the following hold:*

1. *Under the $\alpha-$discounted reward criterion, if $\mu_2 R_2 \geq \mu_1 R_1$, then it is optimal to serve at station 2 whenever station 2 is not empty.*
2. *Under the average reward criterion, if $\lambda\left(\frac{1}{\mu_1} + \frac{1}{\mu_2+\beta_2}\right) < 1$, it is optimal to serve at station 2 whenever station 2 is not empty.*

Proof in the discounted reward case. A little algebra yields that it suffices to show

$$\begin{aligned} \mu_2 R_2 - \mu_1 R_1 &+ \mu_2[v_\alpha(i, j-1) - v_\alpha(i,j)] \\ &+ \mu_1[v_\alpha(i,j) - \big(pv_\alpha(i-1, j+1) + qv_\alpha(i-1,j)\big)] \geq 0 \end{aligned} \tag{3.1}$$

for all $i, j \geq 1$. We follow the sample paths of five processes (on the same probability space) to show (3.1) via a sample path argument. Processes 1-5 begin in states $(i, j-1)$, (i, j), (i, j), $(i-1, j+1)$, and $(i-1, j)$, respectively. Processes 2, 4, and 5 use stationary optimal policies, which we denote by π_2, π_4, and π_5, respectively. In what follows, we show how to construct (potentially sub-optimal) policies for Processes 1 and 3, which we denote by π_1 and π_3, so that

$$\begin{aligned} \mu_2 R_2 - \mu_1 R_1 &+ \mu_2[v_\alpha^{\pi_1}(i, j-1) - v_\alpha^{\pi_2}(i,j)] \\ &+ \mu_1[v_\alpha^{\pi_3}(i,j) - \big(pv_\alpha^{\pi_4}(i-1, j+1) + qv_\alpha^{\pi_5}(i-1,j)\big)] \geq 0. \end{aligned} \tag{3.2}$$

Since π_1 and π_3 are potentially sub-optimal, (3.1) follows from (3.2).

Case 1 *All five processes see an arrival (with probability $\frac{\lambda}{\lambda+\mu_1+\mu_2+(j+1)\beta_2}$) or all five processes see an abandonment at station 2 (with probability $\frac{(j-1)\beta_2}{\lambda+\mu_1+\mu_2+(j+1)\beta_2}$).*

In either scenario, no reward is accrued by any of the processes and the relative position of the new states as measured with respect to the starting states is maintained. We may relabel the states and continue as though we started in these states.

Case 2 *Suppose that the first event is an abandonment in Process 4 only (with probability $\frac{\beta_2}{\lambda+\mu_1+\mu_2+(j+1)\beta_2}$), after which all processes follow an optimal control.*

In this case, it follows that the remaining rewards in the left side of (3.2) are

$$\begin{aligned}\mu_2 R_2 - \mu_1 R_1 + \mu_2[v_\alpha(i, j-1) - v_\alpha(i, j)]\\ + \mu_1[v_\alpha(i, j) - v_\alpha(i-1, j)].\end{aligned} \tag{3.3}$$

Since (3.3) is greater than or equal to the expression on the left side of the inequality in (3.2), it follows that it is enough to show (3.2) whenever the first event in a sample path argument is not an abandonment seen by Process 4 only. That is to say, we may disregard this case provided we show inequality (3.2) for all other remaining cases.

Case 3 *Suppose that the first event is an abandonment in Processes 2-5 (with probability* $\frac{\beta_2}{\lambda+\mu_1+\mu_2+(j+1)\beta_2}$*).*

From this point on assume $\pi_1 = \pi_2$, so that the first two terms in inequality (3.2) are the same. To complete the proof it suffices to show that the remaining rewards are non-negative

$$\mu_2 R_2 - \mu_1 R_1 + \mu_1[v_\alpha(i, j-1) - \big(pv_\alpha(i-1, j) + qv_\alpha(i-1, j-1)\big)]. \tag{3.4}$$

There are several subcases to consider.

Subcase 3.1 *Suppose the next event is an abandonment in Process 4 only (with probability* $\frac{\beta_2}{\lambda+\mu_1+\mu_2+j\beta_2}$*).*

After state transitions assume π_3 uses a stationary optimal policy. Thus, (3.4) becomes

$$\begin{aligned}\mu_2 R_2 - \mu_1 R_1 + \mu_1[v_\alpha(i, j-1) - \big(pv_\alpha(i-1, j-1) + qv_\alpha(i-1, j-1)\big)]\\ = \mu_2 R_2 - \mu_1 R_1 + \mu_1[v_\alpha(i, j-1) - v_\alpha(i-1, j-1)] \geq 0,\end{aligned} \tag{3.5}$$

where the inequality follows since the optimal value function is non-decreasing in each coordinate. The result (in this subcase) follows.

Subcase 3.2 *Suppose policies* π_4 *and* π_5 *assign the server to work at station 1 and 2, respectively. Assume that* π_3 *also works at station 1.*

Suppose that the next event is a service completion at station 1 for Process 4 (with probability $\frac{\mu_1}{\lambda+\mu_1+\mu_2+j\beta_2}$). Suppressing the denominator of the probability, the left hand side of inequality (3.4) becomes

$$\begin{aligned}\mu_1\Big[\mu_2 R_2 - \mu_1 R_1 + p\mu_1[qR_1 + v_\alpha^{\pi_3}(i-1, j) - \big(pv_\alpha^{\pi_4}(i-2, j+1) + qv_\alpha^{\pi_5}(i-1, j-1)\big)]\\ + q\mu_1[qR_1 + v_\alpha^{\pi_3}(i-1, j-1) - \big(pv_\alpha^{\pi_4}(i-2, j) + qv_\alpha^{\pi_5}(i-1, j-1)\big)]\Big]\end{aligned}$$

Assuming that π_3 follows the optimal actions from this point on, a little algebra yields,

$$\mu_1\Big[\mu_2 R_2 - p\mu_1 R_1 + p\mu_1[v_\alpha(i-1,j) - \big(pv_\alpha(i-2,j+1) + qv_\alpha(i-2,j)\big)]\Big]. \tag{3.6}$$

If, however, the next event is a service completion at station 2 in Process 5, then the left hand side of inequality (3.4) becomes

$$\mu_2\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_1[v_\alpha^{\pi_3}(i,j-1) - \big(pv_\alpha^{\pi_4}(i-1,j) + qR_2 + qv_\alpha^{\pi_5}(i-1,j-2)\big)]\Big].$$

Again, assuming π_3 follows an optimal stationary policy from this point on, a little algebra yields

$$\mu_2\Big[\mu_2 R_2 - \mu_1 R_1 - q\mu_1 R_2 + \mu_1[v_\alpha(i,j-1) - \big(pv_\alpha(i-1,j) + qv_\alpha(i-1,j-2)\big)]\Big]. \tag{3.7}$$

Adding (3.6) and (3.7) (with a little algebra) we get

$$\begin{aligned}
&p\mu_1[\mu_2 R_2 - \mu_1 R_1 + \mu_1\big(v_\alpha(i-1,j) - (pv_\alpha(i-2,j+1) + qv_\alpha(i-2,j))\big)] \\
&\quad + \mu_2[\mu_2 R_2 - \mu_1 R_1 + \mu_1\big(v_\alpha(i,j-1) - (pv_\alpha(i-1,j) + qv_\alpha(i-1,j-1)) \\
&\quad + qv_\alpha(i-1,j-1) - qv_\alpha(i-1,j-2)\big)].
\end{aligned}$$

The first expression is (3.4) evaluated at $(i-1,j+1)$. The second is (3.4) repeated. In each case, we can relabel the starting states and repeat the argument. The last expression is non-negative since the value function is non-decreasing in each coordinate.

Subcase 3.3 *Suppose policies π_4 and π_5 assign the server to work at station 2 and 1, respectively. Assume that π_3 also works at station 1.*

If the next event is a service completion at station 2 in Process 4 (with probability $\frac{\mu_2}{\lambda+\mu_1+\mu_2+j\beta_2}$), then the remaining rewards of the processes in (3.4) are (suppressing the denominator)

$$\begin{aligned}
&\mu_2\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_1\big(v_\alpha^{\pi_3}(i,j-1) - pv_\alpha^{\pi_4}(i-1,j-1) - pR_2 - qv_\alpha^{\pi_5}(i-1,j-1)\big)\Big] \\
&\qquad = \mu_2[\mu_2 R_2 - \mu_1 R_1 - p\mu_1 R_2 + \mu_1\big(v_\alpha^{\pi_3}(i,j-1) - v_\alpha(i-1,j-1)\big)]. \qquad (3.8)
\end{aligned}$$

If, however, the next event is a service completion at station 1 in Processes 3 and 5, then from (3.4) we get

$$\begin{aligned}
&\mu_2 R_2 - \mu_1 R_1 + p\mu_1\big[pR_1 + v_\alpha^{\pi_3}(i-1,j) - \big(pv_\alpha(i-1,j) + qv_\alpha(i-2,j)\big)\big] \\
&\quad + q\mu_1\big[pR_1 + v_\alpha^{\pi_3}(i-1,j-1) - \big(pv_\alpha(i-1,j) + qv_\alpha(i-2,j-1)\big)\big].
\end{aligned}$$

Assuming π_3 starting in state $(i-1,j)$ follows the optimal control from this point on, a little algebra yields

$$\mu_1[\mu_2 R_2 - q\mu_1 R_1 + q\mu_1\big(v_\alpha^{\pi_3}(i-1,j-1) - pv_\alpha(i-2,j) - qv_\alpha(i-2,j-1)\big)]. \tag{3.9}$$

Adding expressions (3.8) and (3.9) we get the following lower bound

$$q\mu_1[\mu_2 R_2 - \mu_1 R_1 + \mu_1\big(v_\alpha(i-1,j-1) - pv_\alpha(i-2,j) - qv_\alpha(i-2,j-1)\big)]$$
$$+ \mu_2[\mu_2 R_2 - \mu_1 R_1 + \mu_1\big(v_\alpha(i,j-1) - v_\alpha(i-1,j-1)\big)].$$

The first expression above is (3.4) evaluated at $(i-1,j)$. In this case, we may relabel the states and continue as though we had started in these states. The second expression is non-negative since the value functions are non-decreasing in each coordinate.

There are 8 cases left to consider all with algebra that is directly analogous. We state Case 4 completely below and state the remaining cases with only the associated actions of the other processes for that case (complete details are available in the Online Appendix). In each case, assume that π_1 has the server work at station 1 which leaves the only decision to relay is that of π_3; Processes 2, 4, and 5 use an optimal policy.

Case 4 *Suppose that policies $\pi_i (i = 2, 4, 5)$ all have the server work at station 2. In this case, let π_3 have the server work at station 2 as well.*

If the first event is a service completion at station 1 in Process 1, after which all processes follow an optimal control, then the remaining rewards in (3.2) are

$$\mu_1\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_2\big(pv_\alpha(i-1,j) + qv_\alpha(i-1,j-1) - v_\alpha(i,j) + R_1\big)$$
$$+ \mu_1\big[v_\alpha(i,j) - pv_\alpha(i-1,j+1) - qv_\alpha(i-1,j)\big]\Big]. \tag{3.10}$$

If the first event is a service completion at station 2 in Processes 2-5, again after which optimal controls are used, then the remaining rewards from (3.2) are

$$\mu_2[\mu_2 R_2 - \mu_1 R_1 - \mu_2 R_2 + \mu_1\big(v_\alpha(i,j-1) - pv_\alpha(i-1,j) - qv_\alpha(i-1,j-1)\big)]. \tag{3.11}$$

Adding expressions (3.10) and (3.11) yields

$$\mu_1\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_2\big(v_\alpha(i,j-1) - v_\alpha(i,j)\big)$$
$$+ \mu_1\big[v_\alpha(i,j) - \big(pv_\alpha(i-1,j+1) + qv_\alpha(i-1,j)\big)\big]\Big]. \tag{3.12}$$

Note that expression (3.12) is implied by the left hand side of inequality (3.2). That is, we may restart the argument from here.

Case 5 *Suppose that policies π_5 has the server work at station 1 and policies $\pi_i (i = 2, 4)$ have the server work at station 2. In this case, let π_3 have the server work at station 2 also.*

Case 6 *Suppose policies $\pi_i (i = 4, 5)$ have the server work at station 1 and policy π_2 has the server work at station 2. In this case, let policy π_3 have the server work at station 1.*

Case 7 *Suppose that policies π_4 has the server work at station 1 and that policies $\pi_i (i = 2, 5)$ have the server work at station 2. In this case, let policy π_3 have the server work at station 1.*

Case 8 *Suppose policy π_2 has the server work at station 1 and policies $\pi_i (i = 4, 5)$ have the server work at station 2. In this case, let policy π_3 have the server work at station 2.*

Case 9 *Suppose that policies π_2 and π_5 have the server work at station 1 and that policy π_4 has the server work at station 2. In this case, let policy π_3 have the server work at station 1.*

Case 10 *Suppose that policies π_2 and π_4 have the server work at station 1 and that policy π_5 has the server work at station 2. In this case, let policy π_3 have the server work at station 1.*

Case 11 *Assume that each policy $\{\pi_2, \ldots, \pi_5\}$ serves at station 1.*

In every case save Case 3 we may relabel the states and continue. By doing so we wait until Case 3 occurs. In particular, Subcase 3.1 then yields the result in the discounted reward case.

Proof in the average reward case. For the average case, we make two observations. First, the condition $\lambda\left(\frac{1}{\mu_1} + \frac{1}{\mu_2+\beta_2}\right) < 1$ implies the policy that prioritizes station 2 yields a stable Markov chain (the system acts as an $M/G/1$ system with phase-type service times). Prioritizing station 2 means that compared to any other non-idling policy, the server spends the least amount of time working at station 1. Since station 2 is always stable (because $\beta_2 > 0$) the condition implies the stability of the Markov chain induced by **any** stationary, non-idling policy. Second, if (g, w) is a solution to the average reward optimality equations, the difference $w(x) - w(y)$ represents the difference in the total rewards for an optimal policy starting in states x and y.

Using the same sample path argument provided in the discounted rewards case, we continue to relabel states and restart the argument until Subcase 3.1 occurs. At this time we have (for some integers $i, j \geq 1$).

$$\mu_2 R_2 - \mu_1 R_1 + \mu_1[w(i, j-1) - w(i-1, j-1)]. \tag{3.13}$$

Consider now two processes. Process 1 (2) starts in state $(i, j-1)$ ($(i-1, j-1)$) with Process using an optimal policy and Process 1 using a policy that follows the same actions as Process 2 except to avoid unforced idling. The only way the two processes couple is when $i = j = 1$ and Process 1 sees an extra service completion (and Process 2 is forced to idle. At this point the processes couple and (3.13) is bounded below by $\mu_2 R_2 - \mu_1 R_1 + \mu_1 R_1 = \mu_2 R_2 \geq 0$. The fact that the state $i = j = 1$ will be reached in finite expected time under the optimal policy is a consequence of positive recurrence and the proof is complete. ■

Our next result gives a sufficient condition under which it is optimal to prioritize station 1 (serve at station 1 whenever station 1 is not empty) and is the direct analog of Theorem 3.3. To ease the nomenclature, let

$$A(x,y) = \frac{xe^y}{y^x}\gamma(x,y) = 1 + \sum_{j=1}^{\infty} \frac{y^j}{\Pi_{k=1}^{j}(x+k)}; x > 0; y \geq 0.$$

Theorem 3.4 *Suppose $\beta_2 = 0, \beta_1 > 0$. In this case, the following hold:*

1. *Under the $\alpha-$discounted reward criterion, if $\mu_1 R_1 \geq \mu_2 R_2$, then the policy that prioritizes station 1 is optimal.*
2. *Under the average reward criterion,*
   (a) *Let π_0 be the long-run fraction of time that station 1 is empty under the policy prioritizes station 1. Then*
   $$\pi_0 = \left[A(\frac{\mu_1}{\beta_1}, \frac{\lambda}{\beta_1})\right]^{-1}.$$
   (b) *Suppose under the policy that prioritizes station 1, $P(Ab)$ is the probability a randomly arriving customer abandons the system before completing service. Then*
   $$P(Ab) = \left[A(\frac{\mu_1}{\beta_1}, \frac{\lambda}{\beta_1})\right]^{-1} \cdot \left[\frac{\mu_1}{\lambda} + \left(1 - \frac{\mu_1}{\lambda}\right)A(\frac{\mu_1}{\beta_1}, \frac{\lambda}{\beta_1})\right].$$
   (c) *Suppose π_0 and $P(Ab)$ are as in the previous two parts. If $\frac{p\lambda(1-\Pr(Ab))}{\pi_0\mu_2} < 1$, then it is optimal to prioritize station 1.*

Proof. See Appendices A.2 and A.3. ■

A few remarks regarding Theorems 3.3 and 3.4 are in order. For Theorem 3.3, in the discounted reward case, we require the condition $\mu_2 R_2 \geq \mu_1 R_1$ which is closely related to the classic result of the $c - \mu$ rule. In the average case, to get sufficiency we do not need this inequality. Intuitively, if it is known that there have been several arrivals, in the average case it does not matter when they arrived. In fact, we could collect the reward upon arrival (instead of after service completion at station 1). This is not true for customers moved to station 2 due to the potential abandonment. The way to get the most reward is by serving customers in station 2 as soon as possible; by prioritizing station 2. Similar remarks apply to Theorem 3.4.

In regards to Theorem 3.4, notice that due to the abandonments station 1 is always stable. Under the policy that prioritizes station 1, station 1 acts as a birth-death process (birth rate λ, death rate $i\beta_1 + \mu_1$). The condition provided in Theorem 3.4 is derived by computing the fraction of customers arriving to station 1, that enter station 2 (some abandon, some leave after completing service at station 1) and then computing the fraction of time the server works at station 2 (under the prioritize station 1 policy). That is to say, $\frac{p\lambda(1-\Pr(\text{Ab}))}{\pi_0\mu_2} < 1$ is sufficient for positive recurrence of all states under the prioritize station 1 policy (see the Online Appendix for complete details). Moreover, the policy that prioritizes station 1 spends the least amount of time at station 2 among all non-idling policies. Thus, if station 2 is stable under the policy that prioritizes station 1, it is stable under any non-idling policy.

Note now from the proofs of Theorems 3.3 and 3.4 that in all of the cases not involving abandonments (i.e. cases 1, 4-11 in Theorem 3.3 and cases 2-4 in Theorem 3.4) we may relabel the states and continue until all five processes couple. This implies the following result.

Proposition 3.5 *Suppose $\beta_1 = \beta_2 = 0$. In this case, the following hold:*

1. *If $\mu_1 R_1 \geq \mu_2 R_2$ ($\mu_2 R_2 \geq \mu_1 R_1$), then under the $\alpha-$discounted reward criterion it is optimal to serve at station 1 (2) whenever station 1 (2) is not empty.*
2. *If $\lambda\left(\frac{1}{\mu_1} + \frac{1}{\mu_2}\right) < 1$, then under the average reward criterion, then any non-idling policy is optimal.*

For the rest of the section assume $\beta_1, \beta_2 > 0$. That is to say, we examine the model with abandonments from both stations. Unlike the cases with one of the abandonment rates equal to 0, the decision-maker has to balance abandonments from each station. This presents an interesting challenge since tradeoffs between abandonments and prioritizing services are not at all clear. The main result is in Theorem 3.8. We begin with two lemmas.

Lemma 3.6 *If $\beta_1 \geq \beta_2$, then*

$$v_\alpha(i-1,j) - v_\alpha(i,j-1) \geq -\frac{\mu_1 R_1}{\lambda + \mu_1 + \mu_2 + \beta_1 + \beta_2}. \tag{3.14}$$

Similarly in the average reward case with v_α replaced with w from the AROE.

Proof. We show the result in the discounted case. The average case is similar. Suppose we start two processes in the same probability space. Processes 1-2 begin in states $(i-1,j)$ and $(i,j-1)$, respectively. Process 2 uses a non-idling stationary optimal policy which we denote by π_2. We show how to construct potentially sub-optimal policy π_1 for Process 1 and show that

$$v_\alpha^{\pi_1}(i-1,j) - v_\alpha^{\pi_2}(i,j-1) \geq -\frac{\mu_1 R_1}{\lambda + \mu_1 + \mu_2 + \beta_1 + \beta_2} \tag{3.15}$$

By the optimality of π_2, (3.15) implies (3.14). Without any loss of generality we suppress the discounting.

Let π_1 choose exactly the same action that policy π_2 chooses at every subsequent decision epoch. That is to say, if π_2 chooses to serve at station 1 (2), then π_1 also serves at station 1 (2). Note that Process 2 (1) starts with one more customer in station 1 (2). If π_2 chooses to serve at station 1 (in Process 2) and there are no customers in station 1 in Process 1, then we let π_1 idle the server. It follows from this that the relative position between the two states will be maintained until one of the following two cases occurs.

Case 1 *Customer abandonments that are not seen by both processes.*

Suppose that there is an abandonment in station 1 in Process 2 that is not seen by Process 1 (with probability $\frac{\beta_1-\beta_2}{\lambda+\mu_1+\mu_2+i'\beta_1+j'\beta_2}$ for some positive integers i' and j'). Note we may use this construction since $\beta_1 \geq \beta_2$. In this case, the remaining rewards in the left side of inequality (3.15) are

$$v_\alpha^{\pi_1}(i'-1, j') - v_\alpha^{\pi_2}(i'-1, j'-1). \tag{3.16}$$

Since the value function is non-decreasing in each coordinate, (3.16) is bounded below by 0, so that (3.15) holds.

Suppose next that there is an abandonment at station 2 in Process 1 that is not seen by Process 2 and an abandonment in station 1 in Process 2 that is not seen by Process 1 (with probability $\frac{\beta_2}{\lambda+\mu_1+\mu_2+i'\beta_1+j'\beta_2}$ for i' and j'). In this case, the remaining rewards in the left side of inequality (3.15) are 0 since the processes enter the same state $(i'-1, j'-1)$, so the inequality holds trivially.

Case 2 *Customer services that are not seen by both processes.*

Suppose that there is a service completion in station 1 in Process 2 that is not seen by Process 1 (with probability $\frac{\mu_1}{\lambda+\mu_1+\mu_2+\beta_1+j'\beta_2}$ for some positive integer j'). Note, by the specification of the policies, this can only occur when station 1 is empty for Process 1 and has one customer for Process 2. The remaining rewards on the left side of (3.15) (with the denominator of the probability suppressed) are

$$-\mu_1 R_1 + q\mu_1\Big[v_\alpha(0, j') - v_\alpha(0, j'-1)\Big] \geq -\mu_1 R_1,$$

where the last inequality follows since the value functions are monotone nondecreasing (Proposition 3.2). In this case, the result follows (i.e. inequality (3.15) holds) by noting that $j' \geq 1$, and that the probability $\frac{\mu_1}{\lambda+\mu_1+\mu_2+\beta_1+j'\beta_2}$ (and discounting) times $-R_1$ is greater than or equal to $-R_1$. ∎

Lemma 3.7 *For all $i, j \geq 1$, the following inequality holds:*

$$v_\alpha(i-1, j-1) - v_\alpha(i-1, j) \geq -\frac{\mu_2 R_2}{\lambda + \mu_2 + \beta_2}. \tag{3.17}$$

Similarly in the average reward case with v_α replaced with w from the AROE.

Proof. We show the result in the discounted case. The average case is similar. We show that inequality (3.17) holds using a sample-path argument. Suppose we start two processes in the same probability space. Processes 1-2 begin in states $(i-1, j-1)$ and $(i-1, j)$, respectively. Process 2 uses a non-idling stationary optimal policy which we denote by π_2. Process 1 uses a policy π_1. We show that

$$v_\alpha^{\pi_1}(i-1, j-1) - v_\alpha^{\pi_2}(i-1, j) \geq -\frac{\mu_2 R_2}{\lambda + \mu_2 + \beta_2}. \tag{3.18}$$

By the optimality of π_2, (3.18) implies (3.17). Without any loss of generality we suppress the discounting.

Let π_1 choose exactly the same action that policy π_2 chooses at every subsequent decision epoch. Note that Process 2 starts with one more customer in station 2. If π_2 chooses to serve at station 2 (in Process 2) and there are no customers in station 2 in Process 1, then we let π_1 idle the server. It follows from this that the relative position between the two states is maintained until one of the following two cases occurs.

Case 1 *There is an abandonment at station 2 in Process 2 that is not seen by Process 1.*

In this case, the remaining rewards in the left side of inequality (3.18) are 0 since the processes enter the same state, so that the inequality holds strictly.

Case 2 *There is a service completion in station 2 in Process 2 that is not seen by Process 1 (with probability no greater than $\frac{\mu_2}{\lambda+\mu_2+\beta_2}$).*

The remaining rewards on the left side of (3.18) are

$$\frac{\mu_2}{\lambda + \mu_2 + \beta_2}\Big(- R_2 + v_\alpha(i-1, j-1) - v_\alpha(i-1, j-1)\Big) = -\frac{\mu_2 R_2}{\lambda + \mu_2 + \beta_2},$$

and the result follows (i.e. inequality (3.18) holds) . ■

This leads to the following result on the tandem system with abandonments from both stations.

Theorem 3.8 *The following hold:*

1. *Under the $\alpha-$discounted and the average reward criteria, if $\beta_1 \geq \beta_2$ and $\mu_1 R_1 \geq 2\mu_2 R_2$, then it is optimal to serve at station 1 whenever station 1 is not empty.*

2. *Under the $\alpha-$discounted and the average reward criteria, if $\beta_2 \geq \beta_1$ and $(1-\frac{\mu_2}{\lambda+\mu_2+\beta_2})\mu_2 R_2 \geq \mu_1 R_1$, then it is optimal to serve at station 2 whenever station 2 is not empty.*

Proof of Statement 1. The result is shown in the discounted reward case. The average case is similar. The discounted reward optimality equations imply that it is optimal to assign the worker to station 1 in state (i, j) if

$$\begin{aligned} &\mu_1 R_1 - \mu_2 R_2 + \mu_1\big[\big(pv_\alpha(i-1, j+1) + qv_\alpha(i-1, j)\big) - v_\alpha(i, j)\big] \\ &\quad + \mu_2[v_\alpha(i, j) - v_\alpha(i, j-1)] \geq 0. \end{aligned} \tag{3.19}$$

We thus prove (3.19) via a sample path argument. Suppose we start five processes in the same probability space. Processes 1-5 begin in states $(i-1, j+1)$, $(i-1, j)$, (i, j), (i, j), and $(i, j-1)$, respectively. Processes 3 and 5 use stationary optimal policies which we denote by π_3 and π_5, respectively. In what follows we show how to construct potentially sub-optimal policies π_1, π_2, and π_4 for processes 1, 2, and 4, respectively, so that

$$\begin{aligned} &\mu_1 R_1 - \mu_2 R_2 + \mu_1\big[\big(pv_\alpha^{\pi_1}(i-1, j+1) + qv_\alpha^{\pi_2}(i-1, j)\big) - v_\alpha^{\pi_3}(i, j)\big] \\ &\quad + \mu_2[v_\alpha^{\pi_4}(i, j) - v_\alpha^{\pi_5}(i, j-1)] \geq 0. \end{aligned} \tag{3.20}$$

Since policies π_1, π_2, and π_4 may be sub-optimal, (3.19) follows from (3.20). We note that as long as the relative positions of the processes remains the same, we may relabel the initial states and continue from the beginning of the argument. This occurs (for example) when any of the uncontrolled events occur that are seen by all 5 processes (e.g. arrivals and abandonments seen by all 5 processes). There are several cases to consider, how many of them are precisely the same as those discussed in the proof of Theorem 3.4.

Case 1 *Abandonments that are not seen by all five processes.*

Suppose the next event is an abandonment at station 1 in Processes 3-5 only (with probability $\frac{\beta_1-\beta_2}{\lambda+\mu_1+\mu_2+i\beta_1+(j+1)\beta_2}$), after which all processes follow an optimal control. Note we may use this construction since $\beta_1 \geq \beta_2$. In this case, the remaining rewards on the left side of the inequality in (3.20) (with the denominator of the probability suppressed) are

$$\begin{aligned} \beta_1\Big[&\mu_1 R_1 - \mu_2 R_2 + \mu_1\big[\big(pv_\alpha(i-1, j+1) + qv_\alpha(i-1, j)\big) - v_\alpha(i-1, j)\big] \\ &+ \mu_2[v_\alpha(i-1, j) - v_\alpha(i-1, j-1)]\Big] \geq 0, \end{aligned}$$

where the inequality follows since the value functions are monotone nondecreasing in each coordinate. Next, suppose the next event is an abandonment at station 2 in Process 1 only (with probability $\frac{\beta_2}{\lambda+\mu_1+\mu_2+i\beta_1+(j+1)\beta_2}$), or, an abandonment at station 2 in Processes 1-4 only (with probability $\frac{\beta_2}{\lambda+\mu_1+\mu_2+i\beta_1+(j+1)\beta_2}$), or, an abandonment at station 1 in Processes 3-5 only (with probability $\frac{\beta_2}{\lambda+\mu_1+\mu_2+i\beta_1+(j+1)\beta_2}$). In each case, assume that all processes follow an optimal control immediately after. In the first case, the remaining rewards on the left side of the inequality in (3.20) (with the denominator of the probability suppressed) are

$$\beta_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1[v_\alpha(i-1,j) - v_\alpha(i,j)] + \mu_2[v_\alpha(i,j) - v_\alpha(i,j-1)]\Big]. \tag{3.21}$$

In the second case,

$$\beta_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1[pv_\alpha(i-1,j) + qv_\alpha(i-1,j-1) - v_\alpha(i,j-1)]\Big], \tag{3.22}$$

and in the third case,

$$\begin{aligned}\beta_2\Big[&\mu_1 R_1 - \mu_2 R_2 + \mu_1[\big(pv_\alpha(i-1,j+1) + qv_\alpha(i-1,j)\big) - v_\alpha(i-1,j-1)]\\ &+ \mu_2[v_\alpha(i-1,j) - v_\alpha(i-1,j-1)]\Big]\end{aligned} \tag{3.23}$$

Since each of the three scenarios occurs with the same probability, we can add expressions (3.21), (3.22), and (3.23) to obtain that the remaining rewards in the expression on the left side of the inequality in (3.20) in this case (with the denominator of the probability suppressed) are

$$\begin{aligned}\beta_2\Big[&\mu_1 R_1 - \mu_2 R_2 + \mu_1[pv_\alpha(i-1,j+1) + qv_\alpha(i-1,j) - v_\alpha(i,j)]\\ &+ \mu_2[v_\alpha(i,j) - v_\alpha(i,j-1)]\Big]\\ &+ \Big[2\mu_1 R_1 - 2\mu_2 R_2 + \mu_1[v_\alpha(i-1,j) - v_\alpha(i,j-1)]\Big]\\ &+ \Big[p\mu_1[v_\alpha(i-1,j) - v_\alpha(i-1,j-1)]\Big]\\ &+ \Big[\mu_2[v_\alpha(i-1,j) - v_\alpha(i-1,j-1)]\Big].\end{aligned} \tag{3.24}$$

The expression inside the first pair of brackets above is the left side of inequality (3.20). For this expression, we may restart the argument and continue as though we had started in these states. Consider the expression inside the second pair of brackets above (i.e. $2\mu_1 R_1 - 2\mu_2 R_2 + \mu_1[v_\alpha(i-1,j) - v_\alpha(i,j-1)]$). Lemma 3.6 implies that this expression is bounded below by $\mu_1 R_1 - 2\mu_2 R_2$, which by assumption, is nonnegative. Since the value functions are monotone nondecreasing in each coordinate (Proposition 3.2), it follows that the expressions inside the third and fourth pair of brackets in (3.24) are also nonnegative.

The remaining cases are the same as Cases 2-4 in in the proof of Theorem 3.4, where we show that the relative positions of the processes remains the same, so we may relabel the initial states and continue from the beginning of the argument. In other words, Statement 1 follows by (possibly) relabeling the states and continuing until an abandonment in station 1 in Processes 3-5 only occurs. In the latter case, the remaining rewards in the left side of (3.20) are nonnegative.

Proof of Statement 2. Suppose that $\beta_2 \geq \beta_1$. Recall that it is optimal to prioritize station 2 in state (i, j) whenever it is not empty provided

$$\begin{aligned} &\mu_2 R_2 - \mu_1 R_1 + \mu_2[v_\alpha(i, j-1) - v_\alpha(i, j)] \\ &\qquad + \mu_1[v_\alpha(i, j) - p v_\alpha(i-1, j+1) - q v_\alpha(i-1, j)] \geq 0. \end{aligned} \tag{3.25}$$

Start five processes in the same probability space. Processes 1-5 start in states $(i, j-1)$, (i, j), (i, j), $(i-1, j+1)$, and $(i-1, j)$, respectively, and Process 2, 4 and 5 use stationary optimal policies, which we denote by π_2, π_4 and π_5, respectively. In what follows, we show how to construct policies for Processes 1 and 3, which we denote by π_1, and π_3, respectively, so that

$$\begin{aligned} &\mu_2 R_2 - \mu_1 R_1 + \mu_2[v_\alpha^{\pi_1}(i, j-1) - v_\alpha^{\pi_2}(i, j)] \\ &\qquad + \mu_1[v_\alpha^{\pi_3}(i, j) - p v_\alpha^{\pi_4}(i-1, j+1) - q v_\alpha^{\pi_5}(i-1, j)] \geq 0. \end{aligned} \tag{3.26}$$

Since policies π_1 and π_3 are potentially sub-optimal, (3.25) follows from (3.26). Again we consider only the events that are not seen by all 5 processes.

Case 1 *Suppose that the first event is an abandonment in Process 4 only (with probability* $\frac{\beta_2-\beta_1}{\lambda+\mu_1+\mu_2+i\beta_1+(j+1)\beta_2}$*).*

After this event assume all processes follow an optimal control. In this case, it follows that the remaining rewards in the left side of inequality (3.26) are

$$\begin{aligned} \mu_2 R_2 - \mu_1 R_1 + \mu_2[v_\alpha(i, j-1) - v_\alpha(i, j)] \\ + \mu_1[v_\alpha(i, j) - v_\alpha(i-1, j)]. \end{aligned}$$

Since this last expression is bounded below by the expression on the left side of the inequality in (3.25), it follows that it is enough to show (3.26) whenever the first event in a sample path argument is not an abandonment in station 2 seen by Process 4 only. That is to say, we may disregard this case provided we show inequality (3.26) for all other remaining cases.

Case 2 *Suppose next that the first event is an abandonment in station 2 in Processes 2-5 (with probability* $\frac{\beta_2-\beta_1}{\lambda+\mu_1+\mu_2+i\beta_1+(j+1)\beta}$*) that is not seen by Process 1.*

From this point on assume Process 1 follows an optimal control so that $\pi_1 = \pi_2$. That is, the first two terms in inequality (3.26) are the same. To complete the proof in this case it suffices to show that the remaining rewards

$$\mu_2 R_2 - \mu_1 R_1 + \mu_1[v_\alpha^{\pi_3}(i, j-1) - \big(pv_\alpha^{\pi_4}(i-1, j) + qv_\alpha^{\pi_5}(i-1, j-1)\big)] \tag{3.27}$$

are non-negative. There are several subcases to consider.

Subcase 2.1 *Suppose the next event is an abandonment in station 2 in Process 4 only (with probability $\frac{\beta_2-\beta_1}{\lambda+\mu_1+\mu_2+i\beta_1+j\beta_2}$).*

After state transitions assume π_3 uses a stationary optimal policy. Thus, (3.27) becomes

$$\mu_2 R_2 - \mu_1 R_1 + \mu_1[v_\alpha(i, j-1) - v_\alpha(i-1, j-1)] \geq 0, \tag{3.28}$$

where the inequality follows since the optimal value function is non-decreasing in each coordinate (Proposition 3.2. The result (in this subcase) follows.

Subcase 2.2 *Suppose the next event is an abandonment in station 1 in Process 3 only (with probability $\frac{\beta_1}{\lambda+\mu_1+\mu_2+i\beta_1+j\beta_2}$) or an abandonment in station 2 in Process 4 only (with probability $\frac{\beta_1}{\lambda+\mu_1+\mu_2+i\beta_1+j\beta_2}$).*

After state transitions assume π_3 uses a stationary optimal policy. Thus, (3.27) becomes

$$\mu_2 R_2 - \mu_1 R_1 \geq 0, \tag{3.29}$$

where the inequality follows by assumption. The result (in this subcase) follows.

Subcase 2.3 *Suppose policies π_4 and π_5 assign the server to work at station 1 and 2, respectively. Assume that π_3 also works at station 1.*

Suppose the next event is a service completion at station 1 for Processes 3 and 4 (with probability $\frac{\mu_1}{\lambda+\mu_1+\mu_2+i\beta_1+j\beta_2}$). Suppressing the denominator of the probability, the left hand side of inequality (3.27) becomes

$$\begin{aligned}\mu_1\Big[&\mu_2 R_2 - \mu_1 R_1 + p\mu_1[qR_1 + v_\alpha^{\pi_3}(i-1, j) - \big(pv_\alpha^{\pi_4}(i-2, j+1) + qv_\alpha^{\pi_5}(i-1, j-1)\big)] \\ &+ q\mu_1[qR_1 + v_\alpha^{\pi_3}(i-1, j-1) - \big(pv_\alpha^{\pi_4}(i-2, j) + qv_\alpha^{\pi_5}(i-1, j-1)\big)]\Big]\end{aligned}$$

Assuming that π_3 follows the optimal actions from this point on, a little algebra yields,

$$\mu_1\Big[\mu_2 R_2 - p\mu_1 R_1 + p\mu_1[v_\alpha(i-1, j) - \big(pv_\alpha(i-2, j+1) + qv_\alpha(i-2, j)\big)]\Big]. \tag{3.30}$$

If, however, the next event is a service completion at station 2 in Process 5, then the left hand side of inequality (3.27) becomes

$$\mu_2\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_1[v_\alpha^{\pi_3}(i,j-1) - \big(pv_\alpha^{\pi_4}(i-1,j) + qR_2 + qv_\alpha^{\pi_5}(i-1,j-2)\big)]\Big].$$

Again, assuming π_3 follows an optimal stationary policy from this point on, a little algebra yields

$$\mu_2\Big[\mu_2 R_2 - \mu_1 R_1 - q\mu_1 R_2 + \mu_1[v_\alpha(i,j-1) - \big(pv_\alpha(i-1,j) + qv_\alpha(i-1,j-2)\big)]\Big]. \tag{3.31}$$

Adding (3.30) and (3.31) (with a little algebra) we get

$$\begin{aligned}
&p\mu_1[\mu_2 R_2 - \mu_1 R_1 + \mu_1\big(v_\alpha(i-1,j) - (pv_\alpha(i-2,j+1) + qv_\alpha(i-2,j))\big)] \\
&\quad + \mu_2[\mu_2 R_2 - \mu_1 R_1 + \mu_1\big(v_\alpha(i,j-1) - (pv_\alpha(i-1,j) + qv_\alpha(i-1,j-1)) \\
&\quad + qv_\alpha(i-1,j-1) - qv_\alpha(i-1,j-2)\big)].
\end{aligned}$$

The first expression (with coefficient $p\mu_1$) is (3.27) evaluated at $(i-1,j+1)$. The second is (3.27) repeated. In each case, we can relabel the starting states and repeat the argument. The last expression is non-negative since the value function is non-decreasing in each coordinate.

Subcase 2.4 *Suppose policies π_4 and π_5 assign the server to work at station 2 and 1, respectively. Assume that π_3 also works at station 1.*

If the next event is a service completion at station 2 in Process 4 (with probability $\frac{\mu_2}{\lambda+\mu_1+\mu_2+i\beta_1+j\beta_2}$), then the remaining rewards of the processes in (3.27) are (suppressing the denominator)

$$\begin{aligned}
&\mu_2\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_1\big(v_\alpha^{\pi_3}(i,j-1) - pv_\alpha^{\pi_4}(i-1,j-1) - pR_2 - qv_\alpha^{\pi_5}(i-1,j-1)\big)\Big] \\
&\quad = \mu_2[\mu_2 R_2 - \mu_1 R_1 - p\mu_1 R_2 + \mu_1\big(v_\alpha^{\pi_3}(i,j-1) - v_\alpha(i-1,j-1)\big)]
\end{aligned} \tag{3.32}$$

If, however, the next event is a service completion at station 1 in Processes 3 and 5, then from (3.27) we get

$$\begin{aligned}
&\mu_1\Big[\mu_2 R_2 - \mu_1 R_1 + p\mu_1\big[pR_1 + v_\alpha^{\pi_3}(i-1,j) - \big(pv_\alpha(i-1,j) + qv_\alpha(i-2,j)\big)\big] \\
&\qquad + q\mu_1\big[pR_1 + v_\alpha^{\pi_3}(i-1,j-1) - \big(pv_\alpha(i-1,j) + qv_\alpha(i-2,j-1)\big)\big]\Big]
\end{aligned}$$

Assuming π_3 starting in state $(i-1,j)$ follows the optimal control from this point on, a little algebra yields

$$\mu_1[\mu_2 R_2 - q\mu_1 R_1 + q\mu_1\big(v_\alpha^{\pi_3}(i-1,j-1) - pv_\alpha(i-2,j) - qv_\alpha(i-2,j-1)\big)] \tag{3.33}$$

Adding expressions (3.32) and (3.33) we get the following lower bound

$$q\mu_1[\mu_2 R_2 - \mu_1 R_1 + \mu_1\big(v_\alpha(i-1,j-1) - pv_\alpha(i-2,j) - qv_\alpha(i-2,j-1)\big)] \\ + \mu_2[\mu_2 R_2 - \mu_1 R_1 + \mu_1\big(v_\alpha(i,j-1) - v_\alpha(i-1,j-1)\big)].$$

The first expression above is (3.27) evaluated at $(i-1,j)$. In this case, we may relabel the states and continue as though we had started in these states. The second expression is non-negative since the value functions are non-decreasing in each coordinate. This completes the proof under Case 2.

Case 3 *Suppose the next event is an abandonment at station 1 in Processes 1-3 only, or, an abandonment at station 2 in Process 4 only, or, an abandonment at station 2 in Processes 2-5 only (each with probability* $\frac{\beta_1}{\lambda+\mu_1+\mu_2+i\beta_1+(j+1)\beta_2}$*).*

In all cases, assume all processes follow optimal controls immediately after. In the first case, the remaining rewards in the left side of the inequality in (3.26) are (with the denominator of the probability suppressed)

$$\beta_1\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_2[v_\alpha(i-1,j-1) - v_\alpha(i-1,j)] \\ + p\mu_1[v_\alpha(i-1,j) - v_\alpha(i-1,j+1)]\Big]. \tag{3.34}$$

In the second case, the remaining rewards are

$$\beta_1\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_2[v_\alpha(i,j-1) - v_\alpha(i,j)] \\ + \mu_1[v_\alpha(i,j) - v_\alpha(i-1,j)]\Big]. \tag{3.35}$$

In the third case, the remaining rewards are

$$\beta_1\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_1[v_\alpha(i,j-1) - pv_\alpha(i-1,j) - qv_\alpha(i-1,j-1)]\Big]. \tag{3.36}$$

Adding expressions (3.34), (3.35), and (3.36) yields

$$\beta_1\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_2[v_\alpha(i,j-1) - v_\alpha(i,j)] \\ + \mu_1[v_\alpha(i,j) - pv_\alpha(i-1,j+1) - qv_\alpha(i-1,j)]\Big] \\ + \beta_1\Big[2\mu_2 R_2 - 2\mu_1 R_1 + \mu_2[v_\alpha(i-1,j-1) - v_\alpha(i-1,j)] \\ + \mu_1[v_\alpha(i,j-1) - pv_\alpha(i-1,j) - qv_\alpha(i-1,j-1)]\Big]$$

The expression inside the first pair of brackets above is the left side of inequality (3.26). For this expression, we may restart the argument and continue as though we had started in these states. Consider the expression inside the second pair of brackets above and note that Lemma 3.7 implies that $\mu_2 R_2 - \mu_1 R_1 + \mu_2[v_\alpha(i-1,j-1) - v_\alpha(i-1,j)] \geq \mu_2 R_2 - \mu_1 R_2 - \mu_2 \frac{\mu_2 R_2}{\lambda+\mu_2+\beta_2} \geq 0$, where the last inequality follows by assumption. Lastly, note that the terms $\mu_2 R_2 - \mu_1 R_1 + \mu_1[v_\alpha(i,j-1) - p v_\alpha(i-1,j) - q v_\alpha(i-1,j-1)] \geq 0$ equal the terms in Case 3 of Theorem 3.3 (see expression (3.4)), where we showed that they are nonnegative whenever $\mu_2 R_2 - \mu_1 R_1 \geq 0$.

The remaining cases are the same as Cases 4–11 in the proof of Theorem 3.3, where we show that the relative positions of the processes remains the same, so we may relabel the initial states and continue from the beginning of the argument. In other words, the result now follows since either we arrive at a case in which we continue, or the expression is non-negative. ∎

4 Conclusions

In this paper we add customer impatience to the classic scheduling model in a two phase stochastic service system under a common reward structure; maximizing rewards per service. Our work is motivated by patient care in health care service systems like the triage and treatment process in the ED, and in particular, by the TTR Program at the LMC in Brooklyn, NY, where patients can abandon after triage so that servers must carefully weigh prioritizing triage versus treatment.

Conditions for simple priority rules are provided using a continuous-time Markov decision process formulation. To do this, we must contend with several technical challenges as a result of abandonments: they lead to unbounded transitions rates, so that the model cannot be analyzed by first uniformizing the process. Instead, we use the continuous-time dynamic programming optimality equations and sample path arguments to determine when priority rules are optimal under discounted expected reward and long-run average reward criterion.

There are several avenues for further research. Characterizing the optimal policy in general when priority rules do not hold is of clear interest. We have attempted to prove structural results in this case (in particular the optimality of monotone threshold policies), but have been unable to do so. There are several model extensions worthy of consideration. One might consider time-dependent arrival processes and/or generally distributed service times; both of which occur commonly in practice, such as ED applications. One may also consider multiple servers for both the collaborative and noncollaborative case. As observed in this paper, it is abandonments that make the classic $c\mu$-rule fail, and more broadly, makes the analysis and computation of optimal controls more complicated. While abandonments are well-observed in practice in a broad range of problems, unbounded rates provide significant technical challenges for optimal control and are a bright research direction.

References

[1] H. S. Ahn, I. Duenyas, and M. E. Lewis. Optimal control of a two-stage tandem queuing system with flexible servers. *Probability in the Engineering and Informational Sciences*, 16:453–469, 2002. 1, 2

[2] S. Andradóttir and H. Ayhan. Throughput maximization for tandem lines with two stations and flexible servers. *Operations Research*, 53(3):516–531, May–June 2005. 2

[3] C. Buyukkoc, P. Varaiya, and J. Walrand. The $c\mu$-rule revisited. *Advances in Applied Probability*, 17(1):237–238, 1985. 2

[4] G. Dobson, T. Tezcan, and V. Tilson. Optimal workflow decisions for investigators in systems with interruptions. *Management Science*, 59(5):1125–1141, 2013. 2

[5] I. Duenyas, D. Gupta, and T. L. Olsen. Control of a single-server tandem queueing system with setups. *Operations Research*, 46(2):218–230, March-April 1998. 1, 2

[6] X. Guo and O. Hernández-Lerma. *Continuous-time Markov decision processes*. Springer, 2009. 30, 32

[7] X. Guo and W. Zhu. Denumerable state continuous time Markov decision processes with unbounded cost and transition rates under average criterion. *The ANZIAM Journal*, 43(4):541–557, April 2002. 28, 29

[8] O. Hernández-Lerma, X. Guo, and T. Prieto-Rumeau. A survey of recent results on continuous-time Markov decision processes. *TOP*, 14:177–261, 2006. 27, 28, 29

[9] J. Huang, B. Carmeli, and A. Mandelbaum. Control of patient flow in emergency departments, or multiclass queues with deadlines and feedback. *Operations Research*, 63(4):892–908, 2015. 2

[10] S. Iravani, M. J. M. Posner, and J. A. Buzacott. A two-stage tandem queue attended by a moving server with holding and switching costs. *Queueing Systems*, 26(3-4):203–228, 1997. 1, 2

[11] P. Johri and M. N. Katehakis. Scheduling service in tandem queues attended by a single-server. *Stochastic Analysis and Applications*, 6(3):279–288, 1988. 2

[12] T. Katayama. Analysis of an exhaustive service type tandem queue attended by a moving server with walking time. *Transactions of IECE*, J63-B:10551062, 1980. 2

[13] T. Katayama. Analysis of a tandem queueing system with gate attended by a moving server with walking time. *Transactions of IECE*, J64-B:931–938, 1981. 2

[14] T. Katayama. Analysis of a finite limiting service tandem queue attended by a moving server with walking time. *Review of the Electrical Communication Laboratory*, 31:439–446, 1983. 2

[15] T. Katayama and K. Kobayashi. Sojourn time analysis for a cyclic-service tandem queueing model with general decrementing service. *Mathematical And Computer Modelling*, 22(10-12):131–139, November-December 1995. 2

[16] S. Lippman. Applying a new device in the optimization of exponential queuing systems. *Operations Research*, 23(4):687–710, 1975. 1

[17] S. Nair. A single server tandem queue. *Journal of Applied Probability*, 8(1):95–109, March 1971. 2

[18] R. Nelson. Labor assignment as a dynamic control problem. *Operations Research*, 14(3):369–376, May-June 1966. 2

[19] S. Saghafian, W. J. Hopp, M. P. Van Oyen, J. S. Desmond, and S. L. Kronick. Complexity-based triage: A tool for improving patient safety and operational efficiency. *Manufacturing and Service Operations Management*, 2014. to appear. 2

[20] M. Taube-Netto. Two queues in tandem attended by a single server. *Operations Research*, 25(1):140–147, January-February 1977. 2

A Appendix

A.1 Optimality Equations

In this section we show that the optimality equations have a solution. That is, that Theorem 2.4 holds. This is important since throughout the paper we use the relative values at starting states to show that structural properties hold for the optimal control. For states $y \neq x$, let $q(y|x, a)$ be the rate at which a process leaves x and goes to y given that action a is chosen. Moreover, let $-q(x|x, a)$ be the rate at which a Markov process leaves state x under action a. Note for the current study the transition rate kernal is *conservative*. That is

$$\sum_{y \in \mathbb{X}} q(y|x, a) = 0.$$

The following notation will be used throughout this section.

- $q(x) := \sup\{-q(x|x, a) : a \in A(x)\}$.
- A **deterministic stationary policy** is a function $f : \mathbb{X} \to \mathbb{A}$ such that $f(x) \in A(x)$ for all $x \in \mathbb{X}$. A deterministic stationary policy is simply referred to as a **stationary policy**. Let F denote the set of all stationary policies.
- Let $Q(f) := [q(y|x, f)]$ be the associated matrix of transition rates with the (x, y)th element $q(y|x, f(x))$. For any fixed $f \in F$, the **transition function** associated with the *Markov process* generated by the matrix $Q(f)$ is denoted by

  $$P(u, f) = \{P_{x,y}(u, f)\},$$

  where $P_{x,y}(u, f) := P(X(u) = y|X(0) = x, f(x))$.

Since the abandonments imply that the transition rates are unbounded, we verify that each Markov process generated by a stationary policy yields a transition kernel that (in the one-dimensional case) has row sums equal to 1 for all time. In short, we do not have an infinite number of transitions in finite time. To do so, we verify the following assumption from [8] (see Assumption A in [8]).

Assumption I. There exists a sequence of subsets of $\mathbb{X}_m \in \mathbb{X}$, a non-decreasing function $h \geq 1$ on $\mathbb{X}$ and constants $b \geq 0$ and $c \neq 0$ such that

1. $\mathbb{X}_m \uparrow \mathbb{X}$ and for each $m \geq 1$, $\sup\{q(x)|x \in \mathbb{X}_m\} < \infty$.
2. $\min\{h(x)|x \notin \mathbb{X}_m\} \to \infty$ as $m \to \infty$.

3. $\sum_{y \in \mathbb{X}} h(y)q(y|x,a) \leq ch(x) + b$ for all (x,a).

Lemma A.1 *Assumption I holds with* $\mathbb{X}_m = \{(i,j)|i+j \leq m\}$*;* $h(i,j) = \frac{i}{\mu_1} + \frac{i+j}{\mu_2} + 1$ *(the workload function +1);* $c = \lambda\Big[\frac{1}{\mu_1} + \frac{1}{\mu_2}\Big]$*;* $b = 0$.

Proof. Only the third statement of Assumption I is nontrivial. For $(i,j) \neq (0,0)$ and $a = (n_1, n_2)$ (with $n_1 + n_2 \leq N$) a little algebra yields (the $(0,0)$ case is similar),

$$\begin{aligned}\sum_{(k,\ell)\in\mathbb{X}} q((k,\ell)|(i,j),a)h(k,l) &= \lambda\Big[\frac{1}{\mu_1} + \frac{1}{\mu_2}\Big] - i\beta_1\Big[\frac{1}{\mu_1} + \frac{1}{\mu_2}\Big] - j\beta_2\Big[\frac{1}{\mu_2}\Big] \\ &\quad - \min\{i, n_1\}\mu_1\Big[\frac{1}{\mu_1} + q\frac{1}{\mu_2}\Big] \\ &\quad - \min\{j, n_2\}\mu_2\Big[\frac{1}{\mu_2}\Big] \\ &\leq c,\end{aligned} \tag{A.1}$$

as desired. ∎

The previous result implies that the process generated by each Markov policy yields a (Markov) process that is *regular*. In particular, the row sums of the transition matrix equal to 1, no matter the starting and ending time (cf. [8])

The first set of assumptions stated below, **Assumption A**, are a slight modification of Assumption **A** that appears in Guo and Zhu [7]. If **Assumption A** below holds, then Lemmas 3 and 4 in the paper by Guo and Zhu [7] hold. In particular, v_α is the unique bounded solution to the DROE; for any $\alpha > 0$ there must exist an α-discounted reward optimal policy f_α^*; and a stationary policy $f \in F$ is optimal if and only if it realizes the maximum in the right-hand side of the DROE. In short, the first result in Theorem 2.4 holds. There the authors assume that the countable state space is the non-negative integers, whereas we require them for the non-negative integer quadrant. The assumptions are restated here for completeness.

Assumption A. The following hold:

1. There exists k non-negative functions $w_n, n = 1, \ldots, k$, such that
   (a) for all $x \in \mathbb{X}$ and $a \in A(x), n = 1, \cdots, k-1, \sum_{y\in\mathbb{X}} q(y|x,a)w_n(y) \leq w_{n+1}(x)$;
   (b) for all $x \in \mathbb{X}$ and $a \in A(x), \sum_{y\in\mathbb{X}} q(y|x,a)w_k(y) \leq 0$.
2. $W := (w_1 + \cdots + w_k) \geq 1$, and for all $x \in \mathbb{X}$, $t > 0$, $f \in F$,

$$-\int_0^t \sum_{y\in\mathbb{X}} P_{x,y}(u,f)q(y|y,f(y))W(y)\,du < \infty,$$

where each w_n comes from the previous statement.

3. Moreover, for some $M > 0$,
$$|r(x,a)| \leq MW(x)$$
for each $x \in \mathbb{X}, a \in A(x)$.

Consider now the function on $\mathbb{X}$, $\delta(i,j) = (i+1)\beta_1 + (j+1)\beta_2$. Notice that for $a = (n_1, n_2) \in A(x) = A(i,j)$ (with $n_1 + n_2 \leq N$)

$$\begin{aligned}\sum_{(k,\ell)\in\mathbb{X}} q((k,\ell)|(i,j),a)\delta(k,\ell) &= \lambda(\beta_1) - i\beta_1(\beta_1) - j\beta_2(\beta_2) - \min\{i,n_1\}\mu_1(\beta_1 - p\beta_2) - \min\{j,n_2\}\mu_2(\beta_2)\\ &\leq \lambda(\beta_1) + \min\{i,n_1\}\mu_1(p\beta_2)\\ &\leq \lambda(\beta_1) + N\mu_1 p\beta_2\\ &\leq \big(\lambda + N\mu_1 p\big)\delta(i,j)\end{aligned}$$

Applying Lemma 3.1 from [8] yields for $0 \leq s \leq t$ and a fixed policy π

$$\mathbb{E}^{\pi}_{s,(i,j)}\,\delta(X(t)) \leq e^{-(\lambda+N\mu_1 p)(t-s)} \tag{A.2}$$

Lemma A.2 *Assumption **A** above holds with $w_n = 1$, and $M = R_1 + R_2$.*

Proof. With $w_n = 1$, Assumptions **A**(1a) and (1b) hold for any $x \in \mathbb{X}$ since the row sums of the generator matrix are 0. For Assumption **A**(2), note that for any given stationary policy $f \in F$ and $x = (i,j)$,

$$\begin{aligned}-\sum_{(r,s)\in\mathbb{X}} P_{(i,j),(r,s)}(u,f)q((r,s)|(r,s),f((r,s))W((r,s)) &= k\sum_{(r,s)\in\mathbb{X}} P_{(i,j),(r,s)}(u,f)\\ \Big(\lambda + \min\{r, f((r,s))\}\mu_1 + \min\{s, N - f((r,s))\}\mu_2\Big) &+ kE^{f}_{0,(i,j)}[\delta(X(u)) - \beta_1 - \beta_2]\\ &\leq k(\lambda + N\mu_1 + N\mu_2 + e^{-(\lambda+N\mu_1 p)u})\end{aligned}$$

where the inequality holds by replacing $\min\{r, f((r,s))\}$ and $\min\{s, N - f((r,s))\}$ by the upper bound N and applying (A.2). It follows that

$$-\int_0^t \sum_{y\in\mathbb{X}} P_{(i,j),y}(u,f)q(y|y,f(y))W(y)\,du \leq k[t(\lambda + N\mu_1 + N\mu_2 + \frac{1}{-(\lambda+N\mu_1 p)}e^{-(\lambda+N\mu_1 p)t})] < \infty,$$

so that Assumption **A**(2) holds. Assumption **A**(3) holds since the reward function is bounded by assumption. ∎

Remark A.3 *If **Assumption A** holds, then Lemmas 3 and 4 in the paper by Guo and Zhu [7] hold. In particular, v_α is the unique bounded solution to the DROE.*

The next two sets of assumptions stated below, **Assumption 4** and **Assumption 5.4**, are slight modifications of Assumption **4** and Assumption *5.4* that appear in Guo and Zhu [6]. If **Assumption A** and **Assumption 5.4** below hold, then Theorem 5.9 in Guo and Hernandez-Lerma [6] hold. In particular, there exist a constant $g^* \geq 0$, a stationary policy $f^* \in F\ \alpha$, and a real valued function u^* on $\mathbb{X}$ satisfying the AROE; any $f \in F$ satisfying the maximum in the AROE is average-reward optimal, and thus f^* is an average-reward optimal stationary policy, and the optimal average-reward function, $\rho^{f^*}(i,j)$, equals the constant g^*. There the authors also assume that the countable state space is the non-negative integers, whereas we require them for the non-negative integer quadrant. The assumptions are restated here for completeness.

Assumption 4.

1. $A(x)$ is compact for each $x = (i,j) \in \mathbb{X}$.
2. For all $x = (i,j), y = (k,\ell) \in \mathbb{X}$ the functions $r((i,j),a)$ and $q((k,\ell)|(i,j),a)$ are continuous on $A(x)$.

Lemma A.4 *Assumption **4** above holds*

Proof. Since the set $A(x)$ is finite for all $x = (i,j) \in \mathbb{X}$, **4.**1 and **4.**2 hold trivially. ∎

Assumption 5.4 For some sequence $\{\alpha_n, n \geq 0\}$ of discount factors such that $\alpha_n \downarrow 0$ (as $n \to \infty$) and some $x_0 = (i_0, j_0) \in \mathbb{X}$, there exist a non-negative real valued function H on $\mathbb{X}$ and a constant L_1, such that

1. $\alpha_n v_{\alpha_n}(x_0)$ is bounded in n (this implies that $|v_{\alpha_n}(x_0)| < \infty$, and so we may define the function $u_{\alpha_n}(x) := v_{\alpha_n}(x) - v_{\alpha_n}(x_0) = v_{\alpha_n}(i,j) - v_{\alpha_n}(i_0,j_0)$ on $\mathbb{X}$ for each $n \geq 1$).
2. $L_1 \leq u_{\alpha_n}(x) := v_{\alpha_n}(x) - v_{\alpha_n}(x_0) = v_{\alpha_n}(i,j) - v_{\alpha_n}(i_0,j_0) \leq H(i,j) = H(x)$;
3. $\sum_{(k,\ell)\in\mathbb{X}} H((k,\ell))q((k,\ell)|(i,j),a)$ is continuous in $a \in A((i,j))$ for any $x = (i,j) \in \mathbb{X}$.

To verify **5.4.**1, we will use the following fact (see Proposition 5.6 in [6]).

Proposition A.5 *The following condition implies Assumption **5.4.**1:*

1. *The exists $\tilde{f} \in F$, a function $\tilde{w} > 0$ on $\mathbb{X}$, constants $\tilde{\rho} > 0, \tilde{b} \geq 0$, and $\tilde{M} > 0$ such that, for all $x = (i,j) \in \mathbb{X}$ and $a \in A(x)$, we have*

$$r((i,j),a) \leq \tilde{M}\tilde{w} \tag{A.3}$$

and

$$\sum_{(k,\ell)\in\mathbb{X}} \tilde{w}((k,\ell))q((k,\ell)|(i,j),a) \leq -\tilde{\rho}\tilde{w}((i,j)) + \tilde{b} \tag{A.4}$$

Lemma A.6 *Assumption **5.4** above holds with $x_0 = (0,0)$; $H(i,j) = i(R_1 + R_2) + jR_2$; $L = 0$; $\tilde{f}$ arbitrary; $\tilde{w} \equiv 1$ (i.e. the constant function 1); $\tilde{\rho}$ arbitrary; $\tilde{b} = \tilde{\rho}$; and $M = \mu_1 R_2 + \mu_2 R_2$.*

Proof. The (optimal) α-discounted values are monotone non-decreasing in each coordinate for any given discount factor $\alpha \in (0,1)$. It follows that for any $x = (i,j) \in \mathbb{X}$

$$0 \leq u_{\alpha_n}(x) = v_{\alpha_n}(x) - v_{\alpha_n}(x_0) = v_{\alpha_n}(i,j) - v_{\alpha_n}(0,0),$$

giving us the lower bound in **5.4**(2). We show that the upper bound in **5.4**(2) holds by the following sample path argument. Suppose you start two processes in the same probability space. Processes 1-2 begin in states (i,j) and $(0,0)$, respectively, and Process 1 uses a stationary optimal policy, which we will denote by π_1. To establish the bound, we show how to construct a potentially sub-optimal policy for Process 2, say π_2, so that

$$v_{\alpha_n}^{\pi_1}(i,j) \leq v_{\alpha_n}^{\pi_2}(0,0) + i(R_1 + R_2) + jR_2. \tag{A.5}$$

Since policy π_2 may not be optimal, the upper bound in **5.4**(2) follows from (A.5).

To that end, at each decision epoch, let policy π_2 choose exactly the same action that policy π_1 chooses whenever possible. That is to say, if at any decision epoch, Process 2 has customers waiting for service in station k and policy π_1 chooses to have a server work at station k (for $k = 1,2$) Process 2 does the same. Otherwise, let π_2 idle the server. From this it follows that the only extra rewards seen by Process 1 are those when Process 2 idles and Process 1 sees a service completion. There are at most i extra service completions at station 1 for Process 1 amd $i + j$ service completions at station 2, leading to the upper bound in (A.5).

For **5.4**.1 note that

$$\begin{aligned} r(x,a) = r((i,j),(n_1,n_2)) &= \frac{\min\{i,n_1\}\mu_1 R_1}{\lambda + \min\{i,n_1\}\mu_1 + i\beta_1 + j\beta_2} + \frac{\min\{j,n_2\}\mu_2 R_2}{\lambda + \min\{j,n_2\}\mu_2 + i\beta_1 + j\beta_2}. \\ &\leq \mu_1 R_1 + \mu_2 R_2 := \tilde{M} \\ &= \tilde{M}1 \\ &= \tilde{M}\tilde{w}((i,j)). \end{aligned}$$

so that the inequality in (A.3) holds. Also note that

$$\begin{aligned} \sum_{(k,\ell)\in\mathbb{X}} \tilde{w}((k,\ell))q((k,\ell)|(i,j),a) &= \sum_{(k,\ell)\in\mathbb{X}} q((k,\ell)|(i,j),a) \\ &= 0 \\ &= -\tilde{\rho}1 + \tilde{\rho} \\ &= -\tilde{\rho}\tilde{w}((i,j)) + \tilde{b} \end{aligned}$$

so that inequality (A.4) holds. As a result, Proposition A.5 holds, and hence, Assumption 5.4.(1) also holds.

Lastly, since the set $A(x)$ is finite for all $x = (i, j) \in \mathbb{X}$, **5.4.**3 holds trivially. ■

Theorem A.7 *Suppose that Assumptions **4** and **5.4** above hold. Then the following hold:*

1. *There exists a constant $g^* \geq 0$, a stationary policy f^*, and a real-valued function u^* on $\mathbb{X}$ sastisfying the AROE;*

$$g^* = r((i,j), f^*(i,j)) + \sum_{(k,\ell)\in\mathbb{X}} q((k,\ell)|(i,j), f^*(i,j))u^*(k,\ell)$$
$$= \max_{a\in A(i,j)} \{r((i,j),a) + \sum_{(k,\ell)\in\mathbb{X}} q((k,\ell)|(i,j),a)u^*(k,\ell)\}; \tag{A.6}$$

2. *Any f realizing the maximum on the right-hand side of (A.6) is average reward optimal. Therefore, f^* in (A.6) is an average reward optimal stationary policy, and the optimal average reward satisfies $\rho^{f^*}(i,j) = g^*$, for all $(i,j) \in \mathbb{X}$*

Proof. The proof is the same as the proof of Theorem 5.9 that appears in Guo and Hernandez-Lerma [6]. ■

A.2 Proof of Theorem 3.4 under the discounted reward criterion

An argument similar to the one given in Theorem 3.3 can be used to show that in state (i, j) it is optimal to assign the worker to station 1 in state (i, j) if

$$\mu_1 R_1 - \mu_2 R_2 + \mu_1[\big(pv_\alpha(i-1,j+1) + qv_\alpha(i-1,j)\big) - v_\alpha(i,j)]$$
$$+ \mu_2[v_\alpha(i,j) - v_\alpha(i,j-1)] \geq 0. \tag{A.7}$$

We thus prove (A.7) via a sample path argument. Suppose we start five processes in the same probability space. Processes 1-5 begin in states $(i-1, j+1)$, $(i-1, j)$, (i, j), (i, j), and $(i, j-1)$, respectively. Processes 3 and 5 use stationary optimal policies which we denote by π_3 and π_5, respectively. In what follows we show how to construct potentially sub-optimal policies π_1, π_2, and π_4 for processes 1, 2, and 4, respectively, so that

$$\mu_1 R_1 - \mu_2 R_2 + \mu_1[\big(pv_\alpha^{\pi_1}(i-1,j+1) + qv_\alpha^{\pi_2}(i-1,j)\big) - v_\alpha^{\pi_3}(i,j)]$$
$$+ \mu_2[v_\alpha^{\pi_4}(i,j) - v_\alpha^{\pi_5}(i,j-1)] \geq 0. \tag{A.8}$$

Since policies π_1, π_2, and π_4 are potentially sub-optimal, (A.7) follows from (A.8). In what follows, discounting is suppressed without any loss of generality. Notice that arrivals, services or abandonments seen by all 5 processes result in the same change of state for each process. We may relabel the starting states and continue the argument below. We focus on the events not seen by all 5 processes.

Case 1 *Suppose that Processes 3-5 see an abandonment at station 1 not seen by Processes 1-2, after which all processes follow an optimal control.*

In this case, the remaining rewards in (A.8) are

$$\begin{aligned}\mu_1 R_1 - \mu_2 R_2 + \mu_1[\big(pv_\alpha(i-1, j+1) + qv_\alpha(i-1, j)\big) - v_\alpha(i-1, j)] \\ + \mu_2[v_\alpha(i-1, j) - v_\alpha(i-1, j-1)].\end{aligned} \tag{A.9}$$

Since the value function is non-decreasing in each coordinate, it follows that (A.9) is non-negative.

Case 2 *Suppose that policies π_1 and π_2 have the server work at station 2 and policies π_3 and π_5 all have the server work at station 1. In this case, let π_4 have the server work at station 1 also.*

If the first event is a service completion at station 1 in Processes 3-5, after which all processes follow an optimal control, then the remaining rewards in (A.8) are

$$\begin{aligned}&\mu_1\Big[\mu_1 R_1 - \mu_2 R_2 - \mu_1 R_1 \\ &\quad + \mu_2\big(pv_\alpha(i-1, j+1) + qv_\alpha(i-1, j) - pv_\alpha(i-1, j) - qv_\alpha(i-1, j-1)\big)\Big].\end{aligned} \tag{A.10}$$

If the first event is a service completion at station 2 in Processes 1-2, again after which optimal controls are used, then the remaining rewards from (A.8) are

$$\begin{aligned}\mu_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1 R_2 + \mu_1\big((pv_\alpha(i-1, j) + qv_\alpha(i-1, j-1)) - v_\alpha(i, j)\big) \\ + \mu_2\big(v_\alpha(i, j) - v_\alpha(i, j-1)\big)\Big].\end{aligned} \tag{A.11}$$

Adding expressions (A.10) and (A.11) yields

$$\begin{aligned}\mu_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1\big(pv_\alpha(i-1, j+1) + qv_\alpha(i-1, j) - v_\alpha(i, j)\big) \\ + \mu_2\big(v_\alpha(i, j) - v_\alpha(i, j-1)\big)\Big].\end{aligned} \tag{A.12}$$

Note that expression (A.12) is implied by the left hand side of inequality (A.8). That is, we may restart the argument from here.

Case 3 *Suppose that policies π_1 and π_2 have the server work at station 2 and policies π_3 and π_5 have the server work at station 1 and 2, respectively. In this case, let π_4 have the server work at station 1.*

If the first event is a service completion at station 1 in Processes 3 and 4, after which all processes follow an optimal control, then the remaining rewards in (A.8) are

$$\mu_1\Big[\mu_1 R_1 - \mu_2 R_2 - \mu_1 R_1 + \mu_2 R_1 \\ + \mu_2\big(pv_\alpha(i-1,j+1) + qv_\alpha(i-1,j) - v_\alpha(i,j-1)\big)\Big]. \tag{A.13}$$

If the first event is a service completion at station 2 in Processes 1, 2, and 5, again after which optimal controls are used, then the remaining rewards from (A.8) are

$$\mu_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1 R_2 - \mu_2 R_2 \\ + \mu_1\big((pv_\alpha(i-1,j) + qv_\alpha(i-1,j-1)) - v_\alpha(i,j)\big) \\ + \mu_2\big(v_\alpha(i,j) - v_\alpha(i,j-2)\big)\Big]. \tag{A.14}$$

Adding expressions (A.13) and (A.14) yields

$$\mu_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1\big(pv_\alpha(i-1,j+1) + qv_\alpha(i-1,j) - v_\alpha(i,j)\big) \\ + \mu_2\big(v_\alpha(i,j) - v_\alpha(i,j-1)\big)\Big] \\ + \mu_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1\big(pv_\alpha(i-1,j) + qv_\alpha(i-1,j-1) - v_\alpha(i,j-1)\big) \\ + \mu_2\big(v_\alpha(i,j-1) - v_\alpha(i,j-2)\big)\Big]$$

Note that the expression inside the first pair of brackets in the expression above is implied by the left hand side of inequality (A.8). That is, we may restart the argument from here. The expression inside the second pair of brackets in the expression above is the same as the one on the left hand side of (A.8) but evaluated at $(i, j-1)$. In the latter case, we may relabel the states and continue as though we had started in these states.

Case 4 *Suppose that policies π_3 and π_5 have the server work at station 2 and 1, respectively. In this case, let policies π_1, π_2, and π_4 have the server work at station 2, 2, and 1, respectively.*

If the first event is a service completion at station 1 in Processes 4 and 5, after which all processes follow an optimal control, then the remaining rewards in (A.8) are

$$\mu_1\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1\big(pv_\alpha(i-1,j+1) + qv_\alpha(i-1,j) - v_\alpha(i,j)\big) \\ + \mu_2\big(pv_\alpha(i-1,j+1) + qv_\alpha(i-1,j) - pv_\alpha(i-1,j) - qv_\alpha(i-1,j-1)\big)\Big]. \tag{A.15}$$

If the first event is a service completion at station 2 in Processes 1-3, again after which optimal controls are used, then the remaining rewards from (A.8) are

$$\mu_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1\big((pv_\alpha(i-1,j) + qv_\alpha(i-1,j-1)) - v_\alpha(i,j-1)\big) \\ + \mu_2\big(v_\alpha(i,j) - v_\alpha(i,j-1)\big)\Big]. \tag{A.16}$$

Adding expressions (A.15) and (A.16) yields

$$\Big(\mu_1 + \mu_2\Big)\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1\big(pv_\alpha(i-1,j+1) + qv_\alpha(i-1,j) - v_\alpha(i,j)\big) \\ + \mu_2\big(v_\alpha(i,j) - v_\alpha(i,j-1)\big)\Big]$$

The expression inside the pair of brackets in the expression above is the left side of inequality (A.8). In this case, we may relabel the states and continue as though we had started in these states.

In every case save one we may relabel the states and continue. By doing so we wait until Case 1 occurs. In particular, Case 1 yields the result.

A.3 Proof of Theorem 3.4 under the average reward criterion

Suppose (g, w) satisfy the average reward optimality equations and use the same sample path argument provided in the discounted rewards case. We continue to relabel the starting states and restarting the argument until such time that Case 1 holds. At this point, let π_1, π_2 and π_3 follow optimal controls. Also, let π_4 be a non-idling policy that chooses exactly the same action that policy π_5 chooses at every subsequent decision epoch when possible. In this case, the analogous expression to (A.8) is

$$\mu_1 R_1 - \mu_2 R_2 + \mu_1[\big(pw(i-1,j+1) + qw(i-1,j)\big) - w(i-1,j)] \\ + \mu_2[w^{\pi_4}(i-1,j) - w^{\pi_5}(i-1,j-1)]. \tag{A.17}$$

Since the relative value function is non-decreasing in each coordinate, the previous expression is bounded below by

$$\mu_1 R_1 - \mu_2 R_2 + \mu_2[w^{\pi_4}(i-1,j) - w^{\pi_5}(i-1,j-1)].$$

Since Process 4 starts with one more customer in station 2 than Process, the relative position between Processes 4 and 5 is maintained until $i = j = 1$ (this occurs in finite time due to positive recurrence). When Process 4 sees a service completion in station 2 that is not seen by Process 5 the two processes couple and the remaining rewards in (A.17) are $\mu_1 R_1 \geq 0$. It follows that in the average reward case, it is optimal to serve at station 1 whenever station 1 is not empty. ∎