DYNAMIC ALLOCATION OF HEALTHCARE RESOURCES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Gabriel Zayas-Cabán

January 2015

DYNAMIC ALLOCATION OF HEALTHCARE RESOURCES

Gabriel Zayas-Cabán, Ph.D.
Cornell University 2015

This thesis collects two papers proposing systematic frameworks for dynamic (i.e. state dependent) resource allocation in healthcare delivery. The first focuses on disaster response, and the second, on care delivery in an Emergency Department (ED) triage and treatment process.

The first paper considers how should Emergency Medical Services (EMS) vehicles be allocated in the aftermath of a catastrophic event. The control of EMS vehicles is modeled as a multi-server parallel queueing system and classical results from the queueing theory literature are used to determine how many vehicles can be made available to the affected region. In addition, we propose that a centralized decision-maker coordinate the transfer of resources from the unaffected region to the affected region. A knapsack model is developed to determine how many vehicles to allocate in the different areas within the affected region and a Markov decision process formulation is developed to dynamically reallocate these added resources within each area until a return to normalcy is achieved. We compare our approach to several other reasonable approaches in a numerical study.

The second paper considers how should a medical provider be allocated in a two-phase stochastic service system. To do this, we first propose a single-server tandem queueing model where patients can abandon before receiving treatment during the second phase of service. We then use a Markov decision process formulation and sample path arguments to determine optimal dynamic policies for the medical service provider. In particular, we provide sufficient conditions under which it is optimal to prioritize phase-one service (triage) and sufficient conditions under which it is optimal to prioritize phase-two service (treatment). In addition, we introduce a new class of threshold policies as reasonable alternatives to priority rules. Using data from an actual hospital, we compare the aforementioned policies and several other potential service policies in a simulation study.

BIOGRAPHICAL SKETCH

Gabriel was born and raised in Mayagüez, Puerto Rico until his mid teens, when he moved to Columbia, Missouri. He completed most of his high school education in Columbia, before moving to Florida, where he completed his undergraduate degree in Mathematics at the University of South Florida. After that, he moved to Ithaca, NY to pursue his doctoral studies at Cornell University. In 2015, he will be a visiting researcher at the Center for Healthcare Engineering and Patient Safety at the University of Michigan in Ann Arbor.

This thesis is dedicated to my grandparents Antonia, Natividad, Eduardo, and José; my parents María and José; and my siblings Teresa and José A.

ACKNOWLEDGEMENTS

I am extremely grateful to my committee members Mark E. Lewis, Shane G. Henderson, and James Renegar for all of their time, help, and advice. I thank Mark and Shane for their longstanding collaboration and ideas, and I particularly want to thank my main advisor Mark for not only introducing me to Markov decision processes and queueing systems, but also for his constant support, encouragement, guidance, and detailed feedback throughout my time at Cornell. I am truly lucky to have had him as an adviser and this particular journey would not have been possible without him.

I would also like to thank Matthew Olson and Samuel Schmitz from the MITRE Corporation for their expertise and guidance in formulating the model presented in Chapter 2, and thanks also to Professor Linda V. Green from Columbia University and Professor Jingui Xie from the University of Science and Technology of China for their expertise and feedback in the work presented in Chapter 3. I thank the National Defense Science & Engineering Graduate Fellowship and the Cornell/Sloan Fellowship for their support of this research.

I want to thank my fellow Operations Research and Center for Applied Mathematics graduate students for their friendship throughout our years in Ithaca (Rolf, Brad, Brandon, Zach, Kenneth, Ravi, Danielle, Franz, Joyjit) and one of my oldest and closest friends (and now best "man") Kemi. I also want to thank my wife Amy for filling my life with joy and aspirations and for her infinite patience. Finally, I especially want to thank my parents María and José, my sister Teresa, and my brother José for all of their love and support without which this and all of my previous journeys would have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Healthcare spending in the United States is astronomical and is expected to increase. As a result, improvements in the quality and efficiency of healthcare delivery need consideration. This thesis considers the management of different healthcare operations and proposes systematic frameworks for improving them. Although different methods can be used, the focus of this thesis is on resource allocation.

Considerations for resource allocation in healthcare systems are similar to those in other settings: demand can be both predictable and random; multiple resources need to be allocated at the same time; and services are often provided in multiple steps, through a network. Hence, similar concepts and tools for developing resource allocation frameworks can be used. Yet, there are certain features of healthcare systems that demand special considerations, and hence, different modeling concepts and tools. For example, patients waiting for treatment may require continuing care. Additionally, the outcome of service depends on the patient's health state which can deteriorate while waiting resulting in a need for more or different care.

The considerations above are especially critical when resources for care delivery are expensive or strained due to increased demand. Often, both criteria are satisfied. One example is EMS vehicles, such as ambulances. The day-to-day management of EMS vehicles can be disrupted when catastrophic events, such as hurricanes, occur. These events can severely strain available EMS vehicle capacity, rendering it unable to provide timely access to care. Similar to management of EMS vehicles, resource allocation is also challenging in emergency room (ER) departments. Recent years have seen an increase in annual ER visits at the same time that hospital managers are under increasing pressure to cut costs causing hospitals to merge, downsize, or close. As a result, a growing number of hospitals report overcrowding at the ER, and hence, longer waiting times, boarding times, and periods of ambulance diversion, along with higher rates of patients leaving without treatment and a higher incidence of adverse events. How care is delivered in these two settings, inside the hospital with the ER or outside the hospital with EMS vehicles, differs in significant ways. Yet, in both settings, it is equally important that quality and timely care be delivered as efficiently as possible, and how resources are allocated plays an important role. Dynamic (i.e. state-dependent) resource allocation in these two scenarios can significantly improve performance (e.g., reduced costs, congestion, time spent waiting to be served, etc.), as compared to static (state-independent) resource allocation, but they have not been adequately addressed in the literature.

The work collected in this dissertation aims to fill the existing gap in the literature related to dynamic resource allocation for care delivery in two specific contexts: disaster response where the relevant resources are EMS vehicles and patient flow in the ER where the relevant resources are medical service providers (physicians, physician assistants, or nurses).

In the first part of the thesis, we consider how to allocate EMS vehicles in response to catastrophic or large scale events such as hurricanes, major floods, or terrorist attacks. In such cases, demand for EMS vehicles will almost certainly overwhelm the available supply, and it is necessary for cities in the affected region to request aid (in the form of added capacity) from neighboring municipalities in order to bring the affected region back to its day-to-day levels of operation. We consider a region consisting of several cities, where each city is in charge of managing its own EMS vehicles. We propose that a centralized or statewide decision-maker coordinate the temporary transfer of resources (EMS vehicles) from cities in the unaffected region into cities in the affected region. The control of each city's EMS vehicles is modeled as a multi-server queueing system and classical results are used to estimate the number of vehicles available at each city. We then develop a knapsack model to guide the allocation of available vehicles from the donor area into the affected one and a clearing system model to dynamically control the added resources. As the dimension of the problem is large, a heuristic we call the buddy system is proposed where cities are paired to form city groups. This reduces the size of the problem enough to solve the knapsack problem. Within the city groups, the clearing system model is solved by Markov decision processes. The performance of our heuristic is compared to several other reasonable heuristics via a detailed numerical study, which show that the buddy system exhibits significant cost and time savings and is generally robust to varying parameters.

To our knowledge, this work is the first to consider the assignment and dynamic allocation of EMS resources after a large scale event and the first to provide a general framework from which to begin this allocation.

While the first part of the thesis focuses on the management of health care resources outside the hospital, the second part focuses on the facilitation of care in the ER. Specifically, patient care in many healthcare systems, such as the ER, consists of two phases of service: assessment, or triage, and treatment. It is often the case that these phases are carried out by a single medical provider. We consider how to prioritize the work by the medical provider to balance initial delays for care with the need to discharge patients in a timely fashion. We present a single-server two-stage tandem queueing model for the aforementioned hospital ER triage and treatment process. After completing service in the first station, patients may leave the system. Otherwise, they join the second station where they may abandon before receiving treatment. We use a Markov decision process formulation and sample path arguments to determine the optimal dynamic schedule for the medical service provider.

To our knowledge, the work presented in the second part of the thesis is the first to consider the dynamic allocation of health care delivery in the ER by medical service providers using continuous-time Markov decision processes (CTMDP). In addition, our work is one of few relatively recent papers to address a CTMDP model with unbounded transition rates. Finally, it is the first to propose and analyze a new class of threshold service disciplines, which act as a reasonable compromise between two priority service disciplines.

1.0.1 Dissertation Outline

Chapter 2 covers the allocation of EMS vehicles in response to large-scale events problem. Chapter 3 covers the optimal control of an ER triage and treatment process problem. Corresponding chapter introductions can be found in Sections 2.1 and 3.1. Also given in these chapter introductions are chapter outlines in Sections 2.1.1 and 3.1.1. The literature related to each problem may be found after respective chapter introductions in Sections 2.2 and 3.2. A detailed modeling approach for the EMS problem is given in Section 2.3. Model description and the optimal control results for the triage and treatment work are given in

Section 3.3. The main results in Chapter 3 are given in Section 3.4. Numerical results and analysis for corresponding chapters are given in Sections 2.4 and 3.5. Corresponding chapter summaries are given in Sections 2.5 and 3.6. Directions for future research are discussed in the concluding chapter, Chapter 4.

CHAPTER 2

EMERGENCY MEDICAL SERVICE ALLOCATION IN RESPONSE TO LARGE SCALE EVENTS

2.1 Introduction

After a large scale or catastrophic event it may be the case that no amount of savvy on the part of local emergency medical service (EMS) dispatchers can direct enough resources to alleviate the problem in a timely fashion. In this case other cities in the region may be asked to lend some of their resources until the original city has recovered from the event. The problem can be viewed from several lenses.

- From the perspective of the lending municipality,
  - How many vehicles are available to lend?
  - How long will it be before they are returned?
- From the perspective of the affected region, how should the extra capacity be allocated?

In this chapter, we provide a systematic approach for resource allocation when the demand for EMS vehicles exceeds what is typically experienced. Instead of suggesting that the two (or more) city managers/dispatchers make decisions without coordination, we recommend that a centralized or statewide administrator advise several municipalities to send some of their EMS vehicles to the affected region. These vehicles would then be allocated within the affected region until a return to normalcy is achieved. We note that except under special circumstances, for example in situations where the federal government takes over governance of a particular area, this is not the way large scale emergencies are handled. Thus, our research examines the potential of a centrally controlled approach and how this approach should be implemented.

To complete the analysis, prior to the event, we assume that the control of each city's EMS vehicles can be modeled by a multi-server queueing system. By relaxing the measure of quality of service slightly, we estimate the number of vehicle available for lending to other cities. We then provide a heuristic for static allocation from lending cities to the affected region and another dynamic allocation heuristic for allocation within the affected region. Taken together these questions constitute a macroscopic view of resource allocation from outside the affected region and a microscopic dynamic control problem from within.

After the modeling is complete, it becomes clear that the dimension of the state and action spaces are too large to solve the problem in reasonable time. We propose a heuristic we call the "buddy system" where cities or municipalities are paired to form city groups. It is left to the decision-maker to decide how these groups are formed (for example by location or need). This reduces the size of the problem and allows us to solve a deterministic resource allocation problem for initially allocating vehicles to city groups. Within the city groups a clearing system queueing model is solved via Markov decision processes. Our heuristic is tested by varying several parameters in an extensive numerical study.

2.1.1 Chapter Outline

The rest of the chapter is organized as follows: Section 2.2 contains a brief survey of the literature. Section 2.3 contains a detailed description of how a decision-maker might decide how many vehicles are available from the donor region and how they might be allocated to the affected region. We present the numerical analysis in Section 2.4 and conclude the paper in Section 2.5. A dictionary of symbols and variables can be found in Appendix. A.1.

2.2 Literature Review

Early work on emergency medical response systems is too plentiful to provide a complete review here. Instead we review only the most recent work that closely relates to the current study. First, we discuss some of the work that considers day-to-day allocation of EMS since some of this research can either serve as a baseline or provide insights into our problem. Secondly, we consider some of the closely related models considering response to large scale events. In each case, the work considers particular parts of the response (locally), while we are considering a larger response as might be undertaken by a statewide controller. The reader interested in the early work should consult the excellent survey papers by Swersey [**?**], Brotcorne et al. [**?**], and Goldberg [**?**].

There are several ambulance location models to deal with day-to-day operations. Most often the goal is to design emergency response systems to cover as many calls as possible while maintaining a target average waiting time (see for example [**?**] and the references therein). Ingolfsson [**?**] et al. propose a location model that minimizes the number of ambulances needed to guarantee a certain fraction of the calls is reached within a given time. The authors allow response time to be composed of a random delay (prior to travel to the scene) plus a random travel time. Gerolimisnis et al. [**?**] develop a spatial queueing model for locating emergency vehicles on urban networks. In addition, they incorporate varying service rates depending on incident characteristics. The model is applied using real data for locating freeway service patrol vehicles. In our work, we capture costs for relocating from one region to the next (as opposed to modeling spatial concerns explicitly), but assume that within the affected region the travel costs are negligible. Each case above is significantly different than the current work since we assume that the goal after a large scale event is to provide *best effort* service to return the system to normalcy. We also include a deterministic optimization component that answers the question of regional (pre-)allocations.

One of the major issues with modeling stochastic behavior in EMS vehicle routing is scalability. Maxwell et al. [**?**] use approximate dynamic programming techniques to approximate optimal ambulance redeployment. That is, once ambulances have been deployed, they may be reassigned to a new home base consistent with current network characteristics. This is somewhat similar to the current work, since we allow EMS vehicles to be reassigned within the affected region. Of course our objective is different since they are interested in day-to-day operations whereas we are interested in what happens after a major disaster. Schmid and Doerner [**?**] develop a multi-period version of the ambulance location problem that takes into account time-varying coverage areas and allows vehicles to be repositioned in order to maintain a certain coverage standard throughout the planning horizon. It is formulated as a mixed integer program that tries to optimize coverage at various points in time simultaneously. Each of these studies consider the problem on a day-to-day basis and differ significantly from the way we handle the allocation component of the assignment problem. Our approach is to use a hybrid deterministic optimization component on the macro-scale and a heuristic plus dynamic programming component on the micro-scale.

While all of the work mentioned above is moving toward implementation, a few others stand out as handling particular instances of day-to-day operations (whereas we seek a general framework after a large scale event). Doerner et al. [**?**] propose and implement solution heuristics for an ambulance location model, namely the double coverage ambulance location problem, in Austria. Moroshi [**?**] compares several location models using actual patient call data from Tokyo metropolitan area in order to improve ambulance service. Iannoni et al. [**?**] propose methods to optimize the configuration and operation of emergency medical systems on highways and is specifically applied to highway systems in Brazil. Iannoni et al. [**?**] use a hypercube queueing model (see the surveys cited above for a description of this model) to address location and districting.

Propelled by the events of hurricane Katrina, the attacks on September 11, and the looming possibility of a pandemic, work to respond to large scale events is more recent than some of the models described above. For example, Jain and Mclean [**?**] propose a framework with which to integrate modeling and simulation for emergency response while Aaby et al. [**?**, **?**] used simulation (among several other operations research techniques) to address vaccine dispensing. Banomyong and Sopadang [**?**] developed a Monte-Carlo simulation model to improve emergency logistics in current response models while Bjarnason et al.

[?] used simulation-based optimization for resource placement and emergency response. In each case above simulation is used along with either analysis (to see how well a proposed policy works) or deterministic optimization to try to improve on a policy online. Argon et al. [?] and Jacobson et al. [?] also consider a problem related to disaster response. It is motivated by the patient triage problem that arises after a mass-casualty event where medical resources are severely strained. In the latter scenario, emergency responders perform triage on patients and then have to determine the order in which these patients will be admitted to other resources. They formulate and analyze a dynamic priority assignment problem in a clearing system model with multiple classes of impatient jobs. Jobs are classified based on their lifetime (i.e., their tolerance for waiting until service), service time, and in [?], by a reward distribution as well. In [?], the objective is to minimize the total number of abandonment while in [?], it is to maximize the expected total reward. Similar to their work, we assume that the decision-maker has access to service rate information and can develop relative weights to assess the importance of assigning vehicles to different locations. However, to our knowledge none of the aforementioned work considers the assignment and dynamic allocation of multiple resources after a large scale event which spans different areas within a region and certainly does not attempt to provide a general framework from which to begin this allocation.

2.3 Modeling Approach

Before explaining our approach we make several remarks. First, we have generically defined a large scale event as one where the affected region might call for (or simply require) significant aid in the form of extra capacity. While we do not formally define these events, we mention that clearly catastrophic events like hurricanes, major floods or terrorist attacks would fall in this category. Other events that occur more frequently may also be categorized as large scale events (depending perhaps on the resources of the affected region) like large fires, power outages, or some large automobile accidents. Second, we will not formally define what is meant by a "return to normalcy". Instead, in our modeling approach we assume that there is a zero state. This depends heavily on the type of event and the municipality or region under consideration. What is crucial is that the decision-maker has knowledge of (or a good estimate of) the number of jobs or outstanding calls that need to be serviced at each location before the system is back to the nominal state. Finally, there is the assumption that the central decision-maker has final authority over who gets sent to each municipality. Currently, in practice decisions are made by local authorities in a decentralized manner. Given that it is quite often the case that some municipalities are too generous (cf. [**?**]["The FDNY requested and received mutual aid from Nassau and Westchester counties on September 11"]), we feel that the above assumption (i.e. that the central decision-maker has final authority) is a reasonable one. Indeed, one of the recommendations we would make is that a centralized control can be beneficial.

2.3.1 Preprocessing

After a catastrophic event, a decision-maker needs to assess which cities should act as donors and which should be candidates for added capacity. For the donor cities, the nominal quality of service measures are slightly relaxed in preparation for lost capacity. For the affected region, the same measures are relaxed so that capacity can be allocated to the most needy portion of the region. The lost (or actually lent) capacity from the donor region becomes added temporary capacity to the cities in the affected region. In this section we propose a method for estimating the amount of available capacity. Indeed, this is simply one method. Others could be explored and used as inputs to the allocation steps in Section 2.3.2.

Suppose there is a set of N nodes (in total), where each node represents the operation of an (independent) EMS system for a particular city or municipality. Assume that node k has a fixed and known capacity n_k for $k = 1, 2, \ldots, N$. After the large scale event, we will assume the time to complete a service at a particular node k is exponential, but for now we allow the service times to follow a general distribution with rate μ_k. We make the assumption that under nominal conditions the time between EMS calls at a particular municipality are exponential with rate λ_k. Note that by making these assumptions we are ignoring spatial considerations (like those that might guide which server is assigned to each job) at a particular node. Since demands that are met leave the system, the above description is that of N parallel queues each of which acts as an $M/G/n_k/\infty$ queue.

2.3.1.1 General Service Times

When the time to complete calls is also exponential we provide an explicit estimate of the number of available vehicles based on the average waiting time (see below). Unfortunately, there are no closed form expressions for the average waiting time for the $M/G/n_k/\infty$ system. We suggest a systematic method to estimate the number of available vehicles that follows closely the analysis suggested by [**?**].

Fix a node k. To guaranty stability assume that $\frac{\lambda_k}{n_k \mu_k} < 1$. Instead of the average waiting time as a quality of service measure, we make the tacit assumption that a vehicle that is immediately assigned to a call can complete the call in a timely fashion. If this holds, then only those calls that find all vehicles currently busy are made to wait longer than usual. As an approximation for the probability with which calls made to dispatch take longer than usual we use the probability that a call is blocked in a system without buffer space. In this case, calls that do not find a vehicle immediately available are lost. Computing the blocking probability is done by the well-known *Erlang* blocking formula for an $M/G/n_k/n_k$ queueing system,

$$\pi_{n_k} = \frac{\frac{\left(\frac{\lambda_k}{\mu_k}\right)^{n_k}}{n_k!}}{\sum_{i=0}^{n_k} \frac{\left(\frac{\lambda_k}{\mu_k}\right)^{i}}{i!}}.$$

During normal day-to-day operations the local dispatcher sets the targeted quality of service (QoS) as the proportion of "blocked" calls, say $\tilde{\pi}_{n_k}$. After a large scale event at a nearby municipality, this measure is relaxed to $\alpha > \tilde{\pi}_{n_k}$ (for QoS level α). Since the blocking probability is non-increasing in n_k, exhaustive search can then be used to obtain the maximum $n'_k < n_k$ such that $\pi_{n'_k} > \alpha$. The number of vehicles available for reallocation is $n_k - n'_k$. Repeating the argument for each city in the donor region yields the total number of vehicles available for lending.

Since the actual system calls are allowed to wait for a vehicle to become available, this is an underestimate of the actual blocking probability; the estimates are aggressive. On the other hand, since the same algorithm is used to estimate the blocking probability before and after the large scale event, the estimates are compatible.

2.3.1.2 Exponential Service Times

Continue to consider a single node in the network with a Poisson arrival process. Suppose now, however, that the time to complete each call to dispatch can be modeled as an exponential random variable with rate μ_k. Under this assumption, the system at a fixed node can be modeled as an $M/M/n_k/\infty$ queueing system.

Let $\rho = \frac{\lambda_k}{n_k \mu_k}$ and $r = \frac{\lambda_k}{\mu_k}$. It is well-known that for this system the long-run average number of jobs waiting for a vehicle to be assigned and the average number of total jobs in the system can be computed, respectively by

$$L_q^{n_k} = C\frac{r^{n_k}(\rho)}{n_k!(1-\rho)^2} \qquad\qquad L^{n_k} = r + C\frac{r^{n_k}(\rho)}{n_k!(1-\rho)^2},$$

where C is the normalizing constant such that $C^{-1} = \sum_{n=0}^{n_k} \frac{\lambda_k^n}{n!\mu_k^n} + \sum_{n=n_k+1}^{\infty} \frac{\lambda_k^n}{n_k! n_k^{n-n_k} \mu_k^n}$. Using *Little's Law*, $L = \lambda W$ in each case yields the average wait before being assigned a vehicle is $W_q^{n_k} = \frac{L_q^{n_k}}{\lambda_k}$ and the average wait until service is complete is $W^{n_k} = \frac{L^{n_k}}{\lambda_k}$. In an analogous manner to the previous section, suppose the quality of service measure is the average waiting time. In the normal day-to-day operations, the average wait is W^{n_k}, but after a large scale event, the decision-maker decides to relax this to $\beta > W^{n_k}$. Since the average wait is non-increasing in n_k, exhaustive search yields the maximum n'_k such that $W^{n'_k} > \beta$. The difference $n_k - n'_k$ represents the capacity available to be reallocated.

Remark. *After this analysis is complete, the decision-maker has a vector of "extra" capacity at each city. For the donor cities, this is the capacity that can be reallocated to the affected region, and for cities in the affected region this is capacity that can be used* *to help return the city to its nominal state.*

2.3.2 Allocation from Donor Cities to the Affected Region

As has been alluded to, before beginning the allocation a decision-maker needs to decide which city(ies) will be designated as *donor* cities and which cities are within the *affected* region. The decision-maker will then take the following steps.

1. For each of the cities use one of the results from Section2.3.1.1 (if the nominal case assumes general services) or Section 2.3.1.2 (if the nominal case assumes exponential services)

   (a) Fix a specific region k (i.e. a city or group of cities) and for the cities in this affected region (the number of which is denoted L_k), the available vehicles are put into the vector $\vec{\ell_k}$. That is, the available vehicles at city i are ℓ_i, $i = 1, 2, \ldots, L$.

   (b) The total number of vehicles available from all donor cities is obtained and denoted N_D.

2. Suppose $v_{N_k}(\vec{m_k}, \vec{v_k})$ $(w_{N_k}(\vec{m_k}, \vec{v_k}))$ denotes the total cost (time) of returning region k to normalcy starting with the initial vector of $\vec{m_k} = \{m_1, m_2, \ldots, m_{L_k}\}$ calls above normalcy in the affected region, the current allocation captured in the vector $\vec{v_k} = \{v_1, v_2, \ldots, v_{L_k}\}$, and N_k vehicles from the donor cities.

   (a) Solve the following optimization problem (MP) if objective is to minimize the total cost of returning the system to normalcy:

$$\begin{aligned}
\text{minimize} \quad & \sum_{k=1}^{L} v_{N_k}(\vec{m}_k, \vec{v}_k) \\
\text{subject to} \quad & \sum_{k=1}^{L} N_k = N_D, \\
& N_k \geq 0.
\end{aligned} \tag{2.1}$$

(b) Solve the following optimization problem (MP′) if objective is to minimize the time of returning the system to normalcy:

$$\begin{aligned}
\text{minimize} \quad & \max_{k=1,\ldots,L} w_{N_k}(\vec{m}_k, \vec{v}_k) \\
\text{subject to} \quad & \sum_{k=1}^{L} N_k = N_D, \\
& N_k \geq 0.
\end{aligned} \tag{2.2}$$

Note that constraints can be easily added to (MP) ((MP′)). For example, if the decision-maker decides that city k must have M_k vehicles from the excess capacity, adding the constraint $N_k = M_k$ achieves the goal. Suppose that there is a budget, b and that sending vehicles to the location associated with city k is c_k. The constraint $\sum_{k=1}^{\ell} c_k N_k \leq b$ can be added.

The difficulty in implementing this algorithm is in the assumption that $v_{N_k}(\vec{m}, \vec{v})$ ($w_{N_k}(\vec{m}, \vec{v})$) can be obtained quickly and easily. In the next section, we provide a general case that quickly becomes intractable due to the curse of dimensionality. We then show how the computational issue can be alleviated, but at the cost of an oversimplification. Our heuristic is a compromise between these two extremes.

2.3.3 Allocation within the Affected Region

Consider the reallocation of available vehicles between the cities in the affected region. Mathematically, our decision-making scenario is modeled as a clearing system. That is, each arrival rate is set to zero ($\lambda_k = 0$ for all $k = 1, 2, \ldots, N$), the nominal state is referred to as the zero state and we seek the control policy that empties or "clears" the system at minimal cost. To model the decision-maker's need to prioritize certain areas more than others, the system is charged holding costs h_k per unit time for each job waiting at station k. From the standpoint of the decision-maker, this is not a value judgement on the importance of meeting one service call or another, but it might (for example) capture the severity of various service calls. We assume that the time to move vehicles from one locale to the next is negligible. This is done for tractability and may not directly reflect reality. On the other hand, we model the cost of assigning vehicles from node k to node m as $c_{k,m}$, or more generally, define the function $c(\cdot)$ to model the action of assigning vehicles from one location to another.

Another concern of the decision-makers may be that within the affected region, vehicles are moved around too often. Thus far, the cost function is a function of the allocation to be used in the coming decision epoch, and not of the current decision. Recall, $\vec{v} = \{v_1, v_2, \ldots, v_L\}$ is the current allocation of vehicles to cities and let $d(\vec{v}, a)$ be the cost of changing the allocation from $\vec{v}$ to a. As an example of a possible function d, one might consider the total number of vehicles moved by the allocation $\vec{n}$,

$$d(\vec{v}, \vec{n}) = \sum_{k=1}^{L} |v_k - n_k|.$$

Thus, once the donor city vehicles have been assigned to the affected region the decision-making scenario is as follows: The decision-maker views the state of the system in the state space $\mathbb{X} := \{(\vec{m}, \vec{v}) | \vec{m} \in (\mathbb{Z}^+)^L, \sum_{k=1}^{L} v_k = M\}$, where M is the total number of vehicles available for assignment. The vector $\vec{m} = \{m_1, m_2, \ldots, m_L\}$ is the current number of jobs (patients for example) at each municipality. After each service completion, the decision maker chooses from the set of available (re-)assignments; a subset of $A(M) := \{(n_1, n_2, \ldots, n_L) | \sum_{k=1}^{L} n_k = M\}$. In other words, the decision maker may reallocate vehicles after each service completion until the nominal or "zero" state is reached. In (MP) we seek a control policy that minimizes the cost to empty the system while in (MP′) we find the policy that returns the system to normalcy in the minimum time.

We note that although we allow for the possibility of constantly reallocating vehicles until the "zero" state is reached, the model is robust enough to restrict the set of available assignments to reflect the decision-maker's desire to model reality. For example, we may only allow reassignment of vehicles from one subset of the cities to another.

2.3.3.1 Independent Allocation Heuristic

To search for an optimal control we begin by assuming that cities in the affected region operate independently. That is, once the allocation from the donor region to each city in the affected region in completed, vehicles stay at their assigned location until the affected city has returned to normal after which they return to their respective donors. In this case there is no need to keep track of how many vehicles are currently assigned to each city in the state space. The total cost for city k is computed

$$v_k(i_k, N_k) = c_k(N_k) + \frac{i_k h_k}{\min\{N_k + \ell_k, i_k\}\mu_k} + v_{N_k}(i_k - 1) \tag{2.3}$$

where i_k and N_k, respectively, are the number of calls above normalcy that need to served and the number of vehicles allocated to city k that may be used to bring it back to the "zero" state. Analogously, the total time until city k returns to normal is computed

$$w_{N_k}(i_k, N_k) = \frac{1}{\min\{N_k + \ell_k, i_k\}\mu_k} + w_{N_k}(i_k - 1). \tag{2.4}$$

The recursions above are easy to solve making the allocation problems in (2.1) and (2.2) also simple to solve. This simplified version of the problem has the advantage that the decision-maker does not need to be in constant communication with each municipality (and continually update the state space). The down side is that this is quite the rudimentary heuristic. After a large scale event, it is possible that within the affected region EMS vehicles may be reallocated dynamically to alleviate difficulties in more highly affected areas. The next section discusses the adjustment to the model above to capture these decisions.

2.3.3.2 The L-station Clearing System

In this section we present the other extreme when compared to the independent allocation of Section 2.3.3.1. Suppose that we allow the decision-maker to reallocate vehicles within the affected region. For ease of notation and (computational) tractability, we will assume from here on out that the cost of assigning vehicles depends on the city the vehicles are going and not on the initial donor city. Recall the total number of vehicles available for assignment is M and that the set of available assignments is then $A(M) := \{(n_1, n_2, \ldots, n_L) | \sum_{i=1}^{L} n_k = M\}$. Let e_i be the vector of all zeros except for a 1 in the i^{th} element; $(0, 0, \ldots, 0, 1, 0, \ldots, 0)$. Including the cost of the current allocation and the (smoothing) cost to change allocations the discrete-time dynamic programming cost optimality equations are now

$$v(\vec{m},\vec{v}) = \min_{\vec{n}\in A(M)} \Big\{ c(\vec{n}) + d(\vec{v},\vec{n}) + \sum_{i=1}^{L} \frac{h_i m_i}{\sum_{k=1}^{L} \min\{n_k + \ell_k, m_k\}\mu_k} + \sum_{i=1}^{L} \frac{\min\{n_i + \ell_i, m_i\}\mu_i v(\vec{m} - e_i, \vec{n})}{\sum_{k=1}^{L} \min\{n_k + \ell_k, m_k\}\mu_k} \Big\}. \tag{2.5}$$

where $\vec{m}$ and $\vec{v}$, respectively, are the vectors containing the number of calls above normalcy at each city that need to served and the number of vehicles presently at city k that may be reallocated to bring it back to the "zero" state. The analogous time optimality equations are

$$w(\vec{m},\vec{v}) = \min_{\vec{n}\in A(M)} \Big\{ \frac{1}{\sum_{k=1}^{L} \min\{n_k + \ell_k, m_k\}\mu_k} + \sum_{i=1}^{L} \frac{\min\{n_i + \ell_i, m_i\}\mu_i w(\vec{m} - e_i, \vec{n})}{\sum_{k=1}^{L} \min\{n_k + \ell_k, m_k\}\mu_k} \Big\}. \tag{2.6}$$

The value function $v(\vec{m},\vec{v})$ ($w(\vec{m},\vec{v})$) must be computed for each element of the state space. The difficulty (of course) in computing v (w) is the fact that $(\vec{m},\vec{v})$ could be high dimensional. In each state, the set of actions is the number of ways to allocate M vehicles to L cities. Using the (multi-choose) identity,

$$\left(\!\!\binom{L}{M}\!\!\right) = \binom{L+M-1}{M},$$

shows that the number of actions becomes intractable as the number of vehicles and/or cities gets large. Moreover, even when the recursion can be computed, implementation requires that a centralized decision-maker monitor each city and make decisions dynamically.

2.3.3.3 The Buddy System

Unfortunately, the recursion in (2.5) ((2.6)) is computationally infeasible as the number of affected cities grows. It may also suffer from a lack of implementability. Suppose we can solve the recursion in (2.5) ((2.6)) for city **groups** of size B.

1. Divide the cities into $r := \lceil \frac{L}{B} \rceil$ groups.
2. Parameterize the optimal cost (time) when a city cluster has N_k movable vehicles added as a single node with zero vehicles as $v_{N_k}(\vec{m}, \vec{v})$ ($w_{N_k}(\vec{m}, \vec{v})$).
3. (a) If the objective is to minimize the total cost of returning the system to normalcy, then solve the following optimization problem (MP$_r$):

$$\begin{aligned} \text{minimize} \quad & \sum_{k=1}^{r} v_{N_k}(\vec{m}, \vec{v}) \\ \text{subject to} \quad & \sum_{k=1}^{r} N_k = N_D, \qquad\qquad N_k \geq 0. \end{aligned}$$

   (b) If the objective is to minimize the time of returning the system to normalcy, then solve the following optimization problem (MP$'_r$):

$$\begin{aligned} \text{minimize} \quad & \max_{k=1,\ldots,r} w_{N_k}(\vec{m}, \vec{v}) \\ \text{subject to} \quad & \sum_{k=1}^{r} N_k = N_D, \qquad\qquad N_k \geq 0. \end{aligned}$$

4. Allocate the optimal N_k to city cluster k and use $v_{N_k}(\vec{m}, \vec{v})$ ($w_{N_k}(\vec{m}, \vec{v})$) to allocate the vehicles in that cluster dynamically.

It should clear that the models of Sections 2.3.3.1 and 2.3.3.2 are the cases of this heuristic with $B = 1$ and $B = L$, respectively. As an alternative, suppose instead that the decision-maker has paired cities in the affected region making groups of size 2. That is to say, that each city has a "buddy" city which it will borrow from or lend to depending on the current number of jobs to be completed at the city. The above description can still be used to discern how much excess capacity is added to each 2-city sub group in the affected region. Another benefit in using the buddy system is that the central decision-maker really only needs to know the state of the system immediately after the large scale event. Once the allocation has been made, (s)he can pass control to the city group for local communication.

2.4 Numerical Analysis

In this section we discuss how well the buddy system described in Section 2.3.3.3 performs when compared to several heuristics. Accordingly, we calculated the percentage savings in expected cost and time using the optimization formulation (MP_r) in Section 2.3.3.3 coupled with the buddy system with $B = 2$ when compared to four alternative heuristics. In addition, we calculated the percentage savings in expected time using the optimization formulation (MP'_r) in Section 2.3.3.3 coupled with the buddy system with $B = 2$ when compared to the same four alternative heuristics. The heuristics are described below:

Heuristic 1: The centralized dispatcher allocates available vehicles evenly between the cities. Once vehicles are allocated, they remain in the city until system has returned to the nominal state.

- This heuristic reflects the centralized dispatcher's preference for equity. Each city is treated the same without regard to the number of incidences, processing rates or an assessment of variable severity.

Heuristic 2: Cities are ordered such that the vector of initial jobs above normalcy is such that $m_1 \geq m_2 \geq \cdots \geq m_L$. The centralized dispatcher allocates $a_1 := \min\{N_D, m_1\}$ to city 1, $a_2 := \min\{N_D - a_1, m_2\}$ to city 2, etc., until all available vehicles have been exhausted. Vehicles remain in each city until all jobs are completed.

- The cities are ordered from the one with the most jobs to the one with the least. The decision-maker allocates enough vehicles so that as few as possible jobs in city 1 are made to wait. If there is a surplus, the same is done for city 2, etc.

- In this heuristic, the centralized dispatcher prioritizes cities with the most emergencies.

Heuristic 3: The centralized dispatcher allocates available vehicles evenly amongst city groups of size 2. Once available vehicles have been evenly distributed, vehicles within each city group are allocated dynamically according to the 2-station clearing system.

- This heuristic is similar to the first with the exception that once the macroscopic allocation is made, the microscopic allocation is dynamic.
- We also remark that Heuristic 1 is **not** always a sub-case of Heuristic 3. Indeed, if there 4 total cities and 10 available vehicles, Heuristic 1 divides the first 8 vehicles evenly $(2, 2, 2, 2)$ and then starts from city 1, then city 2, etc. until all vehicles are exhausted (ending with $(3, 3, 2, 2)$ in this case). Heuristic 3 divides the vehicles into two city groups of size 2, each having 5 vehicles. Assuming that cities 1 and 2 form a group and 3 and 4 form a group, the assignment $(3, 3, 2, 2)$ is not achievable by Heuristic 3.

Heuristic 4: The centralized dispatcher determines the city (or cities) with the highest number of jobs. If there is (only) one, allocate a vehicle to that city. If not, among the cities having the highest number of jobs, determine the city with the lowest processing rate and allocate a vehicle there arbitrarily choosing one if there are several. Decrease the number of jobs by one for the recipient city and repeat this process (with the updated "number of jobs" vector) until the available vehicles have been exhausted.

- This heuristic reflects the centralized dispatcher's desire to provide

Parameter Symbol	Definition
r	number of city groups
N_D	number of vehicles available for (re-)allocation
$i_{n,m}$	number of jobs in city n of city group m
$h_{n,m}$	per unit holding cost in city n of city group m
$\mu_{n,m}$	processing rate for city n of city group m

Table 2.1: Parameters and their description

aid to cities that are in most need; those cities with either the largest number of accidents or slowest processing rates.

There are several scenarios in which a policy may not allocate any vehicles to some cities. We have assumed that the vector of excess capacity within the affected region is $\ell = \{1, 1, \ldots, 1\}$. Thus no matter what the assignment of donor vehicles, work is still being completed at a minimum rate. Throughout the numerical analysis we assume the cost of moving vehicles between cities as well as the smoothing cost is zero. We list the parameters tested in Table 2.1.

Two cases are considered. In Case I, there are 4 cities in the affected region with a total of 144 instances. In Case II there 8 cities and a total of 300 different instances. The parameter values for each case are displayed in Tables 3.2 and B.1. We also remark that throughout the study in each case the policies are implemented in a system where cities are paired as in the buddy system. In this system, the buddy system uses the optimal control while for example Heuristic 1 splits the extra capacity evenly amongst the paired sub-groups. This is done for consistency and alleviates the subtle difference between the maximum of expectations and the expectation of maximums. The main takeaways from the numerical study are as follows:

Parameters	Values/Levels
N_D	20, 50
$i_{n,m}(n = 1; m = 1, 2)$	8
$i_{n,m}(n = 2; m = 1, 2)$	8, 20, 50
$h_{1,1}, h_{2,1}, h_{1,2}$	2
$h_{2,2}$	2, 8
$\mu_{n,m}(n = 1; m = 1, 2)$	1
$\mu_{n,m}(n = 2; m = 1, 2)$	1, 5

Parameters	Values/Levels
N_D	15
$i_{n,m}(n = 1, 2; m = 1, 2)$	8
$i_{1,3}, i_{2,3}, i_{1,4}, i_{2,4}$	16, 16, 24, 24
$h_{1,1}, h_{2,1}$	2
$h_{n,m}(n = 1, 2; m = 2, 3, 4)$	1, 8
$\mu_{1,1}, \mu_{2,1}, \mu_{1,3}, \mu_{2,3}$	1
$\mu_{n,m}(n = 1, 2; m = 2, 4)$	1, 5

Table 2.2: Parameter values for two different cases: 4 cities with $r = 2$ (Case I; left panel) and 8 cities with $r = 4$ (Case II; right panel).

- The buddy system performs well in each case and does so consistently. However, since implementation of the system is not without cost one may consider one of the heuristics as an alternative.
- If the donor cities provide enough capacity to cover the number of jobs to be completed, the benefit of the buddy system is significantly reduced.
- There are two components of the decision process; the macroscopic allocation from donors to affected cities and the microscopic (potentially dynamic) allocation of vehicles within the region. Both components can have a significant effect on cost savings.
- Each of these observations are displayed in the 4 city example and confirmed in the 8 city example.
- Finally, we note that the buddy system also performs well under the minimal completion time criterion, even when the control is obtained by minimizing the total cost. So a decision-maker can prioritize more severely injured patients while still guaranteeing a rapid return to normalcy.

2.4.1 Case I: 4 cities ($r = 2$)

2.4.1.1 Expected Cost

Results for the comparison of the total expected cost between the buddy system and the 4 heuristics described above for Case I are summarized in Table 2.3. The percentage savings in expected cost from implementing the buddy system versus Heuristics 1-4 is nonnegative for all instances. In fact, the buddy system is significantly better than Heuristics 1, 2, and 4 in a large proportion of the cases. In each case, the percentage difference was more than 30% more than 35% of the time. Heuristic 3, which was on average more than 7% more costly than the buddy system was more than 10% more costly in 30% of the cases tested.

Compared to the buddy system, Heuristic 1 fared well in the instances where the number of available vehicles was at least the total number of jobs. Indeed when the initial number of jobs vector is $(8, 20, 8, 8)$ the average percent savings when N_D is 50 is only 3.4%. On the other hand, when the initial jobs vector is $(8, 20, 8, 20)$ or $(8, 20, 8, 50)$, with $N_D = 20$, the average cost savings is 32.1%. Heuristic 1 also fared well in those instances where all of the cities had the same number of jobs. Out of the remaining instances, Heuristic 1 fared well when there were high service rates and low holding costs corresponding to cities with a high job count. This stands to reason since the concern of the high job count is mitigated by the high service rates and lower costs.

Similarly, Heuristic 2 fared well in the instances when enough vehicles were available to cover all of the jobs. However, its performance worsened dramatically as the difference between the total number jobs and the number of available vehicles increased. In particular, when the initial vectors were $(8, 20, 8, 20)$

and $(8, 20, 8, 50)$ and $N_D = 20$ the average cost savings were 77.73% and 53.86%, respectively. In general, Heuristic 2 fared the poorest and we obtained significant savings from implementing our heuristic.

Heuristic 3 fared well in all of the same instances as Heuristic 1. It turns out that in these cases the macroscopic allocation decision is the same for the two heuristics. Heuristic 3 then allocates according to what is optimal while Heuristic 1 does no further allocation. Heuristic 3 also performed well when exactly two cities had either 20 or 50 jobs to complete, regardless of the number of vehicles available for reallocation. Upon examining the macroscopic allocation for Heuristic 3 we note that it did not differ significantly from that of our heuristic while the microscopic allocation for the two is in essence the same. When restricted to these cases, the average difference in the costs is 1.3%. The instances in which it performed poorly correspond to those in which there was only one city where the number of jobs was either 20 or 50 and the total number of jobs exceeded the number of available vehicles. This is because the macroscopic allocation decision for Heuristic 3 differs significantly from that of our heuristic; at the macroscopic level, Heuristic 3 does not take into consideration the severity of one city with high incidences while the buddy system does. When restricted to these cases the average cost difference is 18.37%.

We also note that Heuristic 3 fared better than Heuristic 1 in all instances. This suggests that significant benefits can be obtained by incorporating system dynamics at the microscopic decision-making level even when the macroscopic decision is egalitarian.

Lastly, Heuristic 4 fared well in the instances where enough vehicles were available to cover all of the jobs. In particular, when the initial number of jobs vector is either $(8, 8, 8, 8)$ or $(8, 20, 8, 8)$ the average percent savings when $N_D = 50$ is negligible. Furthermore, when the initial number of jobs vector is $(8, 8, 8, 8)$ the average percent savings when $N_D = 20$ is 3.1% and when the initial number of jobs vector is $(8, 20, 8, 20)$ the average percent savings when $N_D = 50$ is only 0.51%. If the processing rates and costs corresponding to cities with 20 or 50 jobs took on high and low values, respectively, then Heuristic 4 also performed well. It performed dramatically worse otherwise. This is due to the fact that in the remaining instances, when the initial number of jobs vector are $(8, 8, 8, 50)$, $(8, 20, 8, 50)$, and $(8, 50, 8, 50)$, or when it is $(8, 20, 8, 8)$ with $N_D = 20$, Heuristic 4 typically ignores the processing rates and always ignores the costs. In these cases the average cost savings using the buddy system is 34.2%.

Percentage Savings	**Heuristic 1**	**Heuristic 2**	**Heuristic 3**	**Heuristic 4**
<= 1%	10	24	53	32
Between 1% and 5%	21	2	22	10
Between 5% and 10%	16	4	26	5
Between 10% and 20%	15	6	30	20
Between 20% and 30%	25	9	10	26
Between 30% and 40%	31	7	3	17
Between 40% and 50%	18	8	0	9
Between 50% and 60%	7	23	0	13
> 60%	1	61	0	12
Average	23.66%	47.99%	7.13%	24.71%
Standard Deviation	17.12%	30.68%	8.40%	22.13%
Max	61.78%	95.17%	35.50%	84.28%
Min	0%	0%	0%	0%

Table 2.3: Percentage savings in expected cost for Case I: The Buddy System v. Heuristics 1-4. Note the values in columns 2-5 represent the number of instances in the range in column 1.

2.4.1.2 Expected Completion Time for Expected Cost Optimal Policies

In this section, we use the policies derived using the buddy system under the minimal cost criterion (solving (MP)) and compare the results to Heuristics 1-4 under the expected completion time criterion. The results are summarized in Table 2.4. The expected completion time for Heuristic 1 is lower than or equal to the buddy system in those instances where the number of available vehicles is at least the total number of jobs. If we consider only the cases where the initial jobs vector is $(8, 20, 8, 8)$ and $N_D = 50$ the average times savings is 0.76%. When the total number of jobs exceeds the number of vehicles available, Heuristic 1 outperforms the buddy system when high processing rates and/or high cost corresponded to at least one city with a high number of jobs. For example, when the initial jobs vector is $(8, 20, 8, 50)$, $\vec{h} = (2, 2, 2, 8)$, and either $\vec{\mu} = (1, 1, 1, 5)$ or $\vec{\mu} = (1, 5, 1, 5)$ the average times savings is -6.24%. It also fared well in those instances where all the cities had the same number of jobs. This makes sense since the buddy system is taking into account different holding costs while the criterion does not call for such consideration. For instance, when the initial jobs vector is $(8, 8, 8, 8)$ and $N_D = 20$ the average times savings is 0.82%. For all other cases, the expected completion time for the buddy heuristic heuristic is lower, with an average times savings of 22.98%.

The expected completion time for Heuristic 2 is better than that of the buddy system when the number of vehicles available exceeds the total number of jobs. It also fared well in those instances where all the cities had the same number of jobs. Note that in the latter case, the initial allocation for Heuristic 2 is the same as that for Heuristic 1. Lastly, Heuristic 2 fared well in the instances where $N_D = 50$, the initial jobs vector is $(8, 20, 8, 20)$, and $\vec{\mu} = (1, 1, 1, 5)$ with average times savings is 0.05%. Otherwise, the buddy system had a lower expected time of completion by an average of 57.08%.

As in the expected cost section, the instances where the expected completion time for Heuristic 3 is comparable or better than the buddy system correspond to the same instances where the expected completion time for Heuristic 1 also performs well. It also performed comparably to the buddy instances when the number of jobs exceeded the total number of vehicles available for reallocation, the initial jobs vector is one of $(8, 20, 8, 20)$, $(8, 20, 8, 50)$, or $(8, 50, 8, 50)$ and $\vec{\mu} = (1, 1, 1, 1)$ with average times savings is 1.52%. In the remaining cases, where the buddy system outperforms Heuristic 3, the average time savings is 8.1%.

Lastly, the expected completion time for Heuristic 4 is comparable to that of the buddy system in those instances where the number of available vehicles is at least the total number of jobs, when there is at least one high processing rates to offset the effects of high job count, and when the number of jobs in each city are the same. To be specific, these instances include the cases when the initial jobs vectors is $(8, 20, 8, 8)$ with $N_D = 50$ and when the initial jobs vectors are $(8, 50, 8, 8)$ and $(8, 20, 8, 20)$ with $\vec{\mu} = (1, 5, 1, 1)$ or $\vec{\mu} = (1, 5, 1, 5)$, and $N_D = 50$. In the former case, the average times savings is 0% and in the latter cases, it is 6.32%. In the remaining cases, where the buddy system outperforms Heuristic 4, the average time savings is 35.8%.

2.4.1.3 Expected Completion Time for ($\mathbf{MP}'_r$)

While the last section used the policies tuned to the cost criterion and showed that the buddy system did well when compared to the other heuristics, in this

Percentage Savings	Heuristic 1	Heuristic 2	Heuristic 3	Heuristic 4
Between -30% and -20%	1	0	0	0
Between -20% and -10%	6	0	3	0
Between -10% and -5%	10	0	5	0
Between -5% and -1%	14	0	6	0
Between -1% and 0%	20	28	38	24
Between 0% and 1%	10	0	17	10
Between 1% and 5%	23	0	25	22
Between 5% and 10%	3	6	33	1
Between 10% and 20%	24	0	13	10
Between 20% and 30%	14	3	5	17
Between 30% and 40%	7	9	0	14
> 40%	12	98	0	54
Average	9.25%	46.02%	3.39%	27.70%
Standard Deviation	16.00%	28.63%	6.93%	25.05%
Max	44.73%	90.52%	22.07%	78.76%
Min	-27.12%	-0.41%	-19.79%	-0.00%

Table 2.4: Percentage savings in expected completion time for Case I: The Buddy System v. Heuristics 1-4 (dynamic control computed using expected cost). Note the values in columns 2-5 represent the number of instances in the range in column 1.

section we tune the buddy system to minimize the total expected completion time. Results for the expected completion time are summarized in Table 2.5. The percentage savings in expected time from implementing the minimax buddy system versus Heuristics 1-4 is nonnegative for all instances.

The expected completion time for Heuristic 1 is comparable to that of the minimax buddy system in those instances where the number of available vehicles is at least the total number of jobs. If we consider only the cases where the initial jobs vector is $(8, 20, 8, 8)$ and $N_D = 50$ the average times savings is 0.01%. Out of the remaining instances, the expected completion time for Heuristic 1 is comparable to the minimax buddy system when high processing rates correspond to cities with a high number of jobs. For example, when the initial jobs vector is $(8, 20, 8, 50)$ and $\vec{\mu} = (1, 5, 1, 5)$ the average times savings is 3.49%. That is, when the rates offset the negative effects of a high jobs count. For all other cases, the average times savings is 26.15%.

The expected completion time for Heuristic 2 is comparable to that of the minimax buddy system only when the number of vehicles available exceeds the total number of jobs, when high processing rates correspond to cities with 20 or 50 jobs, and when the number of jobs at each city is the same. Otherwise, our heuristic had a lower expected time of completion by an average of 62.92%.

The expected completion time for Heuristic 3 is no greater than that of the minimax buddy system. This stands to reason since allocating vehicles according to what Heuristic 3 prescribes serves as a feasible allocation/solution to the minimax buddy system heuristic. Consequently, the expected completion time obtained from Heuristic 3 for any given instance is an upper bound to the expected completion time of the minimax buddy system heuristic.

Lastly, the expected completion time for Heuristic 4 is comparable to that of the buddy system only in those instances where the number of available vehicles was at least the total number of jobs or if at least one high processing rate corresponds to a city with a high job count. In these instances, the average times savings is 0.09%. Otherwise, our heuristic had a lower expected time of completion by an average of 43.47%.

Percentage Savings	Heuristic 1	Heuristic 2	Heuristic 3	Heuristic 4
Between 0% and 1%	26	20	44	22
Between 1% and 5%	26	0	18	10
Between 5% and 10%	10	4	12	4
Between 10% and 20%	18	6	18	6
Between 20% and 30%	18	2	22	16
Between 30% and 40%	10	8	14	20
Between 40% and 50%	20	8	8	12
Between 50% and 60%	8	12	2	8
Between 60% and 70%	2	46	6	36
> 70%	6	38	0	10
Average	21.46%	52.66%	15.21%	36.84%
Standard Deviation	21.37%	28.61%	17.34%	26.65%
Max	72.21%	94.55%	60.81%	86.35%
Min	0.00%	0.00%	0.00%	0.00%

Table 2.5: Percentage savings in expected completion time for Case I: The Minimax Buddy System v. Heuristics 1-4 (dynamic control computed using expected cost). Note the values in columns 2-5 represent the number of instances in the range in column 1.

2.4.2 Case II: 8 cities ($r = 2$)

2.4.2.1 Expected Cost

The results from Case I support the hypothesis that larger savings in expected cost can be obtained when implementing the buddy system if the large scale event leaves the total number of jobs significantly higher than the total number of vehicles available for reallocation. In particular, the buddy system is useful when resources are scarce relative to the need. To check the scalability of this observation the parameters in Case II are chosen so that resources are scarce in a slightly larger model. The remaining parameters are varied. The results for Case II are summarized in Table 2.6.

As in the previous case, Heuristics 2 and 4 performed poorly compared to the buddy system and were worse than Heuristics 1 and 3. In general, Heuristic 2 performs worse than Heuristic 4. We obtain savings of at least 4% from implementing the buddy system versus Heuristic 1 and 3. In line with our hypothesis, higher processing rates improved the performance of Heuristics 1 and 3 relative to the buddy system. This stands to reason since higher processing rates help reduce the need for scarce resources. For example, in the case where the processing rates in city groups 1-4 are all 1, the average cost savings is 36.27%. On the other hand, when the processing rates for city groups 1,2, and 3 are all 1 but the processing rate for city group 4 is 5, the average cost savings is 30.67%. Similarly, an increase in cost increases the importance of allocating resources effectively. For example, in the case where the cost vector is $\vec{h} = (2, 2, 1, 1, 1, 1, 1, 1)$, the average cost savings is 21% but when the cost vector is $\vec{h} = (2, 2, 8, 8, 8, 8, 8, 8)$ the average cost savings is 30.81%.

Furthermore, similar to earlier results, Heuristic 3 fared better than Heuristic 1 in all of the instances. This is further support for considering dynamics when making the reallocation decisions.

Percentage Savings	**Heuristic 1**	**Heuristic 2**	**Heuristic 3**	**Heuristic 4**
<= 1%	0	0	0	0
Between 1% and 5%	0	0	3	0
Between 5% and 10%	1	0	9	6
Between 10% and 20%	13	0	57	4
Between 20% and 30%	38	6	97	6
> 30%	248	294	134	284
Average	39.70%	68.51%	27.97%	57.26%
Standard Deviation	10.79%	10.23%	10.42%	15.87%
Max	66.51%	82.82%	55.25%	79.76%
Min	9.08%	28.02%	4.03%	7.55%

Table 2.6: Percentage savings in expected cost for Case II: The Buddy System v. Heuristics 1-4. Note the values in columns 2-5 represent the number of instances in the range in column 1.

2.4.2.2 Expected Completion Time for Expected Cost Optimal Policies

Percentage Savings	Heuristic 1	Heuristic 2	Heuristic 3	Heuristic 4
Between −50% and −40%	0	0	19	0
Between −40% and −30%	21	0	24	0
Between −30% and −20%	0	0	39	0
Between −20% and −10%	3	0	7	0
Between −10% and −5%	65	0	12	0
Between −5% and −1%	0	0	2	0
Between −1% and 0%	0	0	15	0
Between 0% and 1%	0	0	0	0
Between 1% and 5%	9	0	6	0
Between 5% and 10%	6	0	40	0
Between 10% and 20%	66	0	74	0
Between 20% and 30%	94	0	45	0
Between 30% and 40%	24	0	17	4
Between 40% and 50%	10	82	0	141
Between 50% and 60%	0	33	0	46
Between 60% and 70%	0	168	0	102
> 70%	0	17	0	7
Average	11.41%	59.01%	1.00%	55.44%
Standard Deviation	20.02%	7.99%	22.04%	8.64%
Max	46.99%	72.11%	38.18%	72.11%
Min	-34.89%	47.45%	-41.10%	36.51%

Table 2.7: Percentage savings in expected completion time for Case II: The Buddy System v. Heuristics 1-4. Note the values in columns 2-5 represent the number of instances in the range in column 1.

The results for Case II are summarized in Table 2.7. As can been seen from Table 2.7, the expected completion time for the buddy system is lower than that of Heuristics 2 and 4 in all of the instances. It is lower than that of Heuristic 1 in approximately 70% of the cases, and it is lower than that of Heuristic 3 in approximately 66% of the cases.

Heuristic 1 outperforms the buddy system heuristic when the processing rate for at least one city in city group 4 is high (5), and either the holding costs for at least one city in city group 4 is high (8) and the holding costs for at least two additional cities is also high (8), or, the holding costs for at least 4 cities are high (8). For instance, when $\vec{\mu} = (1, 1, 1, 1, 1, 1, 1, 5)$ and either $\vec{h} = (2, 2, 8, 8, 8, 8, 1, 1)$ or $\vec{h} = (2, 2, 1, 1, 1, 8, 8, 1)$, the average time savings is -8.43%.

Heuristic 3 outperforms the buddy system heuristic when the processing rate for at least one city in city group 4 is 5 and either the holding cost for at least one city in city group 4 is 8 or the holding cost for at least 2 other cities is also 8. For example, when $\vec{\mu} = (1, 1, 1, 1, 1, 1, 1, 5)$ and $\vec{h} = (2, 2, 1, 1, 1, 1, 8, 1)$ or $\vec{h} = (2, 2, 1, 1, 8, 8, 1, 1)$ the average time savings is -4.15%.

2.4.2.3 Expected Completion Time for (MP'_r)

The results for Case II are summarized in Table 2.8. As can been seen from Table 2.8, the expected completion time for the buddy system is lower than all of the heuristics in all of the instances, with savings of at least 50.16%.

Percentage Savings	Heuristic 1	Heuristic 2	Heuristic 3	Heuristic 4
<= 1%	0	0	0	0
Between 1% and 5%	0	0	0	0
Between 5% and 10%	0	0	0	0
Between 10% and 20%	0	0	0	0
Between 20% and 30%	0	0	0	0
Between 30% and 40%	0	0	0	0
Between 40% and 50%	0	0	0	0
Between 50% and 60%	90	0	210	0
Between 60% and 70%	120	0	90	0
> 70%	90	300	0	300
Average	62.41%	82.80%	56.67%	81.46%
Standard Deviation	8.17%	2.04%	8.10%	0.00%
Max	73.37%	85.91%	68.94%	81.46%
Min	52.35%	81.46%	50.16%	81.46%

Table 2.8: Percentage savings in expected completion time for Case II: The Minimax Buddy System v. Heuristics 1-4. Note the values in columns 2-5 represent the number of instances in the range in column 1.

2.5 Chapter Summary

In this chapter, we consider how to respond to an emergency scenario in which a centralized dispatcher decides how to allocate EMS resources to affected cities from an unaffected region nearby. Our goal is to aid the decision-maker so that (s)he can bring the affected region back to its day-to-day levels of operations as efficiently as possible. We thus proposed a model in which EMS for individual cities are represented by Markovian queues with finite capacity. This model guides the pre-allocation stage. Additionally, the model is used to make the location decision via a queueing clearing system. It is well known that dynamic programming becomes computationally intractable when the size of the state space increases. We propose an implementable heuristic which at the macroscopic level is analogous to the classic resource allocation problem. At the microscopic level it allocates vehicles dynamically. Our results show that significant advantages in both cost and time savings are obtained when using our proposed heuristic as opposed to some common sense heuristics. The savings are robust over various parameter settings when the total number of jobs exceeds the total number of available vehicles.

CHAPTER 3

OPTIMAL CONTROL OF AN EMERGENCY ROOM TRIAGE AND TREATMENT PROCESS

3.1 Introduction

The treatment of individual patients in health care systems often consists of two phases of service provided by a single medical service provider. For example, in health clinics, that operate in lower income communities or at universities, a patient may arrive without an appointment. After some time, a physician will first interview the patient to obtain a brief medical history and description of the symptoms, and the same physician then conducts the physical examination. Another example is Emergency Room (ER) departments, where some patients can either be treated and discharged from the ER while others have to be admitted into the hospital. For patients in the former category, it is sometimes the case that the person who does the triage, i.e. determines their severity status, also diagnoses, treats, and discharges them. Both illustrations are of patient care that is comprised of two phases of service undertaken by a single person: triage, which may include taking medical history, and treatment, which may include writing prescriptions, providing care information, administering shots, or splinting broken fingers.

Motivated in part by the Lutheran Medical Center (LMC) Triage-Treat-and-Release (TTR) program, we model the ER triage and treatment process as a two-phase stochastic service system with a single service provider (i.e., physician or physician assistant). The LMC is a full-service, 462- bed academic teaching hospital. It offers a wide range of clinical programs to diverse communities in Brooklyn, New York. Patients arriving to the ER are automatically registered and then proceed to triage (phase one service) which is conducted on a first-come-first-served (FCFS) basis. After triage, high severity patients who are likely to be admitted to the hospital are assigned to another part of the ER for testing and/or treatment while those who are low severity and low complexity await treatment, called phase-two service, in the triage area. Patients awaiting triage rarely leave before being seen but patients that are awaiting phase-two service have limited patience and may abandon the system before receiving final treatment. The service in both phases is carried out by the same medical provider, usually a physician assistant (PA). After receiving phase-two service, patients are discharged and leave the system. We assume that rewards are received by the system for completing phase i service, $i = 1, 2$; R_1 for phase 1 and R_2 for phase 2. No rewards are received in phase 2 for patients who abandon the system.

The LMC TTR program is a new and creative way to help reduce long waiting times in the ER. Placing a physician in triage, like the TTR program does, has been one suggested intervention to alleviate ER overcrowding. This is because, in theory, it allows for earlier patient contact with a physician and resultant earlier advanced decision-making. In addition, there have been several studies that suggest that physicians or physician assistants are more reliable in their assessment of patients during triage ([**?**], [**?**]). If a hospital is going to use a more trained professional to do triage, it may also be desirable that the same person would also be involved in treatment. We also note that the fast-track system another suggested intervention for alleviating long waiting times. This system deals with patients who are identified at triage as those who can be quickly discharged after treatment in the ED. The TTR program is different in that the person who does the triage is the one doing the treatment rather than the triage and treatment being decoupled as in a normal fast-track. Finally, many hospitals, including the LMC, have multiple people doing triage. The single-server model we consider here can act as a proxy for the multi-server model when the traffic intensity is high. This can be done by replacing the (single-server) service rate with the total service rate in a multi-server system. Since not much time is spent with the system having only a few customers, it is not difficult to decide what to do on the boundary (just avoid idling if possible).

The goal of this research is to develop service disciplines for the medical service provider that maximize the total discounted expected reward or long-run average reward of the system and to compare these with alternative service policies. Accordingly, we formulate the medical provider allocation problem as a continuous-time Markov decision process (CTMDP). In contrast to many existing models, the transition rates of our CTMDP model are unbounded due to potential abandonments by patients waiting for or in service in phase-two. This complicates our analysis, since, for example, a technique known as uniformization, used to transform the CTMDP into an equivalent discrete-time MDP, can no longer be used (see for example [**?**]). It follows that an alternate approach for the solution and analysis of our model is needed.

In this chapter, we use results from [**?**] and provide a mild extension to the results in [**?**] to address some of these technical difficulties. In particular, we use the results from [**?**] to show the existence of discounted expected and average reward optimal stationary policies and a solution of the discounted expected and average optimality equations. We then use the optimality equations and sample path arguments to analyze the structure of the optimal policy. In particular, we give sufficient conditions for when it is optimal to prioritize phase 2 service (i.e. treatment). This policy benefits both the patient and provider since it means that once a patient receives phase-one treatment, (s)he immediately proceeds to phase-two service. In the discounted case, the sufficient condition is akin to the classic c-μ rule. For the long-run average reward case, prioritizing phase-two service is optimal regardless of the reward structure provided the system operating under this policy is stable. We also give a sufficient condition for when to prioritize phase-one service.

We present a new class of policies, that we call K-level Threshold policies as a reasonable compromise between the two priority service disciplines. We then compare several service disciplines, including the two priority rules and the K-level Threshold service disciplines, in a simulation study with respect to long run average reward, and other important practical considerations such as average total number of patients in each queue, average total number of patients in system, and average sojourn times in each queue.

To our knowledge, our work is the first to consider the dynamic allocation of health care delivery in the ER by a single medical service provider using CTMDP. In addition, our work is one of the few of the relatively recent work to address a CTMDP model with unbounded transition rates. Finally, it is the first to propose and analyze a new class of threshold service disciplines, which act as a reasonable compromise between the two priority service disciplines.

3.1.1 Chapter Outline

The rest of this chapter is organized as follows. Section 3.2 contains a summary of the literature relevant to our work. Section 3.3 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 which do not idle the provider whenever there are patients waiting to be served. Section 3.4 contains our main theoretical results: we give sufficient conditions for when to prioritize phase-two service under each reward criteria. In Section 3.5, we introduce a class of threshold policies as a compromise between the two priority service disciplines. We then present a simulation study comparing several service policies with respect to long run average reward, average total number of patients in each phase, average total number of patients in the system, and average sojourn time in phase 1. We conclude with a brief discussion in Section 3.6.

3.2 Literature Review

In this section we highlight those papers most relevant to our work; namely, two-phase stochastic service systems, or tandem queues, with a single service provider, and divide these into two categories. The first are those that analyze the performance of tandem queues under an already specified policy and the second are those that consider optimal service disciplines.

Among those that analyze the performance of single server tandem queues,

[?] 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 is one in which the provider 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. [?] considers the same model but with a zero switching rule at each phase. [?] analyzes the system under a zero switching rule at each phase but with non-zero switchover times.

A K-limited policy is one in which whenever the provider visits a phase, (s)he continues serving that phase until either this phase becomes empty or K patients are served, whichever occurs first. In [?] and [?], Katayama extends his previous work to consider gated service and K-limited service disciplines, respectively. In both papers, intermediate finite waiting room is allowed. A gated service policy is one in which, once the provider switches phases, (s)he serves only patients who are in that phase at the time immediately following the switch. [?] analyze the sojourn time under general K-decrementing service policies; that is policies in which once the server visits phase one, (s)he continues serving that phase until either this phase becomes empty or K patients, are served, whichever occurs first, and then serves at phase-two until it is empty. We mention two additional service disciplines, which we consider in our simulation study. The first are exhaustive service disciplines, where the provider visits phase 1 (or 2) and serves patients in that phase until it becomes empty, and then switches to the other phase. The second are priority policies, in which provider serves patients in phase 1 (or 2) according to a preemptive or non-preemptive priority rule.

To the best of our knowledge, [**?**] is the first to consider optimal service policies for a tandem queueing system over a finite horizon. [**?**] also consider the optimal service disciplines for queues in series. Customers arrive according to a Poisson process and the service times at each queue are independent and exponentially distributed. They show that the discipline that always assigns the provider to the non-empty queue closest to the exit stochastically minimizes the total number of customers in the system. [**?**] consider optimal service disciplines in a forest network of N queues without arrivals and where the service times at each queue are independent and generally distributed. The goal is to minimize the total expected discounted cost, which consists of linear holding costs, lump sum switching costs or set-up delays. They consider the two node model as a special case, which is similar to the holding cost version of our model without abandonments, and in which two nodes are connected probabilistically in tandem. That is to say, a job served at node 1 is routed to 2 with probability p and exits the system with probability $1 - p$. They define an optimal policy for the latter model and show that an optimal policy must be exhaustive if the holding cost in node 1 is larger than node 2. [**?**] consider optimal service disciplines in a two-stage tandem queue with Poisson arrivals and general service times. The goal is to minimize total long run average holding and switching costs. They show that the optimal policy in the second stage is greedy and that if the holding cost rate in the second stage is greater or equal to the rate in the first stage, then it is also exhaustive. They also give a condition under which a sequential service policy is optimal. [**?**] consider optimal service disciplines in a system where there is a setup time for the provider to switch between different phases. The arrival process is a Poisson process and the service time at each phase is an independent and generally distributed random variable. Neither setups nor services can be preempted. They assume that the holding cost at each operation is non-decreasing in operation number and seek to minimize discounted and average cost. The main result is that the optimal policy when the provider is set up for phase one service is completely described by a monotone switching curve. More recently, [**?**] consider optimal service disciplines for two flexible service providers in a two-stage tandem queueing system assuming Poisson arrivals and exponential service times. They consider a collaborative scenario and the non-collaborative scenario. In the former, providers can collaborate to work on the same job, which is similar to the single provider case. They provide sufficient conditions under which it is optimal to allocate both providers to phase 1 or 2 in the collaborative case. They also show that the optimal rule for prioritizing phase-one service in the collaborative case also holds in the non-collaborative case. [**?**] consider queues in tandem with a single service provider, in which the provider may allocate a fraction α of the service capacity to station 1 and $1-\alpha$ to station 2 when both are busy. Using work conservation and FIFO, they show that the wait time in system increases with α on every sample path. For Poisson arrivals and an arbitrary joint distribution of service times of the same customer at each station, the average waiting time at each station is given for the special cases $\alpha = 0$ and $\alpha = 1$.

To our knowledge, our work is the first to consider optimal service disciplines for a single service provider in a two-phase stochastic service system using Markov decision processes in which patients can abandon before receiving medical care. In addition, for our simulation study, we introduce and analyze a new class of policies which we have called K-level threshold policies. In a threshold policy with level K, the provider continues to prioritize station 2, but moves to station 1 when the number of customers at station 1 reaches K. That is to say, the threshold policy has the medical service provider work at station 2 until either station 2 is empty or the number of patients at station 1 reaches K. In the first case, (s)he continues to follow the prioritize station 2 policy. If the threshold causes the station change, then the provider works at 1 until it is empty and then returns to prioritizing station 2. Note that the policy that prioritizes station 2 spends the highest proportion of effort at station 2, while that which prioritizes station 1 spends the least. The prioritize station 1 (2) policy is equivalent to setting K-level Threshold Policy with $K = 1$ (∞). In between these two extremes are threshold policies with higher thresholds spending more time at station 2.

3.3 Model Description and Preliminaries

We approximate the triage-treatment system with the following model. Customers arrive to the service system (ER) according to a Poisson process of rate λ and immediately join the first queue. After receiving service at station 1 a reward of R_1 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 (random and hidden) patience time is generated. Assume that the service requirements at station i are exponential with rate $\mu_i > 0$, $i = 1, 2$ and the abandonment time is exponential with rate β. If the customer does not complete service before the abandonment time ends, the customer leaves the system without receiving service at the second station. After each event (arrival, service completion or abandonment) the provider views 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 $T := \{t_n, n \geq 1\}$ 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). A policy prescribes where the provider should be allocated (0 = idle, 1 = station 1 and 2 = station 2) for all states for all time. The available actions in state $x = (i, j)$ are

$$A(x) = \begin{cases} \{0, 1, 2\} & \text{if } i, j \geq 1; \\ \{0, 1\} & \text{if } i \geq 1, j = 0; \\ \{0, 2\} & \text{if } j \geq 1, i = 0; \\ \{0\} & \text{if } i = j = 0. \end{cases}$$

The function $r(x, a)$ is the expected reward function (since each state change does not represent the accrual of a reward),

$$r((k, \ell), a) = \begin{cases} \frac{\mu_1 R_1}{\lambda + \mu_1 + \ell\beta} & \text{if } k > 0,\ a = 1, \\ \frac{\mu_2 R_2}{\lambda + \mu_2 + \ell\beta} & \text{if } \ell > 0,\ a = 2, \\ 0 & \text{if } a = 0. \end{cases}$$

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

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

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 patients 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 $v_\alpha(i,j) := \sup_{\pi\in\Pi} v^\pi_\alpha(i,j)$, where Π is the set of all non-anticipating policies. Similarly, for the average reward case, the average reward of the policy f is $\rho^f(i,j) := \liminf_{t\to\infty} \frac{v^f_{0,t}(i,j)}{t}$.

3.3.1 Restrict attention to non-idling policies

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

Proposition 1. *Under the $\alpha-$discounted reward (finite or infinite horizon) or the average reward criterion, there exists a non-idling policy that is optimal.*

3.3.1.0.1 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 an abandonment time 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, both are unbeknownst to the provider at the time of the arrival. Fix time $t \geq 0$ and assume $\alpha > 0$. Process 1 uses a policy π_1 that idles the provider unnecessarily (say instead of assigning her or him to station 1) until the first event and uses a stationary optimal policy thereafter. 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 provider to station 1 until the first event and satisfies

$$v^{\pi_2}_{\alpha,t}(i,j) \geq v^{\pi_1}_{\alpha,t}(i,j). \tag{3.1}$$

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

Suppose the first event is an arrival, a service completion seen by the two processes, 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 (3.1) 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 (3.1) 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 provider to station 1 (since it now has one less patient at station 1 than Process 1). In this case, it idles the provider. 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 (3.1) 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 (3.1) holds strictly. Setting $\alpha = 0$ and repeating the proof yields the result for the average case. ∎

3.3.2 Optimality equations

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 tell us what structure of the value functions implies that simple optimal policies exist and then prove that structure via sample path arguments. For a function h on $\mathbb{X}$ define the following mapping

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

For $i, j \geq 1$

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

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 section 3.3.2.1.

Theorem 1. *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),*

     $$\alpha v_\alpha = T v_\alpha. \tag{3.2}$$

   - *(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 cost criterion if $\lambda\left(\frac{1}{\mu_1} + \frac{1}{\mu_2+\beta}\right) < 1$:*
   - *(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.*

3.3.2.0.2 Proof. Since the rewards are bounded and there are only a finite states that can be reached during transitions, Assumptions A of [**?**] hold with $w_n = 1$. The results for the discounted reward case now follow directly from Lemma 3 parts (ii)-(iv) of [**?**]. ∎

3.3.2.1 Existence of a solution to the optimality equations

In this section we show that the optimality equations have a solution. This is important since throughout this chapter 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 kernel is *conservative*. That is

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

The following notation will be used throughout this section.

- $q(x) := \sup\{-q(x|x,a) : a \in A(x)\}$.
- Let F denote the set of all **stationary policies**. That is,

$$F = \{f | f : x \in \mathbb{X} \to f(x) \in A(x) \in \mathbb{X}\}.$$

- For any fixed $f \in F$, the **transition function** associated with the *Markov process* generated by the *Rate 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 [**?**] (see Assumption A in [**?**]).

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 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);* $c = \lambda\Big[\frac{1}{\mu_1} + \frac{1}{\mu_2}\Big]$; $b = 0$.

3.3.2.1.1 Proof. Only the third statement of Assumption I is nontrivial. For $(i,j) \neq (0,0)$ a little algebra yields (the case when $(i,j) = (0,0)$ is similar),

$$\begin{aligned}\sum_{(k,\ell)\in\mathbb{X}} q((k,\ell)|(i,j),a)h(k,l) = \lambda\Big[\frac{1}{\mu_1}+\frac{1}{\mu_2}\Big] - j\beta\Big[\frac{1}{\mu_2}\Big] \\ - \mu_1\mathbb{1}(a=1)\Big[p(\frac{1}{\mu_1}+\frac{1}{\mu_2}) + (1-p)(\frac{1}{\mu_1})\Big] \\ - \mu_2\mathbb{1}(a=2)\Big[\frac{1}{\mu_2}\Big] \qquad (3.3) \\ \leq c,\end{aligned}$$

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 are 1, no matter the starting and ending time (cf. [**?**])

The following set of assumptions are slight modifications of Assumptions **A**,**B**, and **C** that appear in [**?**]. 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 F is the set of all stationary policies, $P(u, f)$ is the transition matrix with respect to the Q−matrix $Q(f)$, and each w_n comes from the previous statement.

3.

$$|r(x, a)| \le MW(x), x \in \mathbb{X}, a \in A(x),$$

for some $M > 0$.

Consider now the function on $\mathbb{X}$, $\delta(i, j) = j\beta$. Notice that

$$\begin{aligned}\sum_{(k,\ell)\in\mathbb{X}} q((k, \ell)|(i, j), a)\delta(k, \ell) &= \lambda \cdot 0 - j\beta(\beta) + \mathbb{1}(a = 1)\mu_1 p\beta - \mathbb{1}(a = 2)\mu_2\beta \\ &\le \beta(\mu_1 - \delta(i, j)).\end{aligned}$$

Applying Lemma 3.1 from [**?**] yields for $0 \le s \le t$ and a fixed policy π

$$\mathbb{E}^{\pi}_{s,(i,j)}\, \delta(X(t)) \le e^{-\beta(t-s)} j\beta + (1 - e^{-\beta(t-s)})(\mu_1/\beta). \tag{3.4}$$

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

3.3.2.1.2 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 $x = (i, j)$,

$$\begin{aligned}-\sum_{y\in\mathbb{X}} P_{(i,j),y}(u, f)q(y|y, f(y))W(y) &= k \sum_{y\in\mathbb{X}} P_{(i,j),y}(u, f)\Big(\lambda + \mathbb{1}(f(y) = 1)\mu_1 + \mathbb{1}(f(y) = 2)\mu_2\Big) \\ &\quad + kE^{f}_{0,(i,j)} m(X(u)) \\ &\le k(\lambda + \mu_1 + \mu_2 + e^{-\beta u} j\beta) + k(1 - e^{-\beta u})(\mu_1/\beta),\end{aligned}$$

where the inequality holds by replacing the indicators by 1 and applying (3.4). It follows that

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

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

Remark 1. *If **Assumption A** holds, then Lemmas 3 and 4 in the paper by [**?**] hold. In particular, v_α is the unique bounded solution to the DROE* (3.2).

Assumption B. For some sequence $\{\alpha_n, n \geq 0\}$ of discount factors such that $\alpha_n \downarrow 0$ and some $x_0 = (i_0, j_0) \in \mathbb{X}$, there exist a non-negative function h and a constant N^*, such that for all $n \geq 1$, $a \in A(x)$, $x = (i, j) \in \mathbb{X}$,

1. $N^* \leq u_{\alpha_n}(x) := v_{\alpha_n}(x) - v_{\alpha_n}(x_0) = v_{\alpha_n}(i, j) - v_{\alpha_n}(i_0, j_0) \leq h(i, j) = h(x)$;
2. $\sum_{(k,\ell)\in\mathbb{X}} q((k, \ell)|(i, j), a)h(k, \ell) < \infty$;
3. there exists a non-negative sequence $\{\phi_{i,j}, i, j \geq 0\}$ that is monotone non-increasing (i.e. $\phi_{i,j} \leq \phi_{k,\ell}$ whenever $k \leq i$ and $\ell \leq j$) such that
$$\sum_{i=1}^{\infty} \phi_{i,j} = \infty$$
for every j, or,
$$\sum_{j=1}^{\infty} \phi_{i,j} = \infty$$
for every i. Moreover, there are positive integers $\hat{u}, \hat{v}$ such that, for any $a \in A(k, \ell)$,
$$r((k, \ell), a) \geq q(k, \ell)h(k, \ell)\phi_{k,\ell},$$
whenever $k \geq \hat{u}, \ell \geq \hat{v}$;
4. there exists positive integers n, m, n', and m', constants $C > 0, \chi \in (0, 1)$ such that, for any $(i, j) \in \mathbb{X}$, $a \in A(i, j)$,
$$q((k, \ell)|(i, j), a)h(k, \ell) \leq \chi^{k+\ell}C,$$
whenever $i \geq m$, $j \geq n$, and $k \geq n' + i$, or, $\ell \geq m' + j$.

Lemma 3. *Assumption **B** above holds with* $x_0 = (0,0)$; $h(i,j) = i(R_1 + R_2) + jR_2$; $N^* = 0$; $\phi_{i,j} = \frac{\min\{\mu_1 R_1, \mu_2 R_2\}}{q(i,j)^2 h(i,j)}$; $\hat{u} = \hat{v} = 1$; $n = m = 0, n' = m' = 2$; *and* $C > 0$ *and* $\chi \in (0,1)$ *arbitrary.*

3.3.2.1.3 Proof. The α-discounted optimality equations 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 **B**(1). We show that the upper bound in **B**(1) 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

$$v_{\alpha_n}^{\pi_1}(i,j) \leq v_{\alpha_n}^{\pi_2}(0,0) + i(R_1 + R_2) + jR_2. \tag{3.5}$$

Since policy π_2 may not be optimal, the upper bound in **B**(1) follows from (3.5).

To that end, at each decision epoch, let policy π_2 choose exactly the same action that policy π_1 chooses. That is to say, if at any decision epoch, Process 2 has patients waiting for service in station 1 and policy π_1 chooses to have the provider work at station 1, or, Process 2 has patients waiting for service in station 2 and policy π_1 chooses to have the provider work at station 2, then let policy π_2 have the provider work at station 1, or, have the provider work at station 2, respectively. Otherwise, let π_2 idle the provider. 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 (3.5).

For **B**(2) note that

$$\begin{aligned}\sum_{(k,\ell)\in\mathbb{X}} q((k,\ell)|(i,j),a)h(k,\ell) = \lambda(R_1 + R_2) + \mathbb{1}(a = 1)\mu_1\Big(p(-R_1) + q(-R_1 - R_2)\Big) \\ + \mathbb{1}(a = 2)\mu_2(-R_2) + j\beta(-R_2) \\ < \infty.\end{aligned}$$

Note that $\phi_{i,j} = \frac{\min\{\mu_1 R_1, \mu_2 R_2\}}{q(i,j)^2 h(i,j)}$ is a non-negative sequence that is monotone decreasing. Moreover,

$$q(k,\ell)^2 h(k,\ell) = (\lambda + \max\{\mu_1,\mu_2\} + \ell\beta)^2 (k(R_1 + R_2) + \ell R_2),$$

so that $q(k,\ell)^2 h(k,\ell)$ is a linear function of k. It follows that

$$\sum_{k=1}^{\infty} \phi_{k,\ell} = \sum_{k=1}^{\infty} \frac{\min\{\mu_1 R_1, \mu_2 R_2\}}{q(k,\ell)^2 h(k,\ell)} = \min\{\mu_1 R_1, \mu_2 R_2\} \sum_{k=1}^{\infty} \frac{1}{q(k,\ell)^2 h(k,\ell)} = \infty,$$

for any fixed ℓ. Also note that whenever $k \geq 1, \ell \geq 1$

$$\begin{aligned}r((k,\ell),a) = \mathbb{1}(a = 1)\frac{\mu_1 R_1}{\lambda + \mu_1 + \ell\beta} &+ \mathbb{1}(a = 2)\frac{\mu_2 R_2}{\lambda + \mu_2 + \ell\beta} \\ &\geq \frac{\min\{\mu_1 R_1, \mu_2 R_2\}}{\max\{\lambda + \mu_1 + \ell\beta, \lambda + \mu_2 + \ell\beta\}} \\ &= \frac{\min\{\mu_1 R_1, \mu_2 R_2\}}{q(k,\ell)} \\ &= q(k,\ell)h(k,\ell)\phi_{k,\ell}.\end{aligned}$$

Thus, **B**(3) holds.

Finally, observe that if the current state of the system is (i, j), then the process can transition to only one of the following states at the next decision epoch:

$(i+1, j), (i-1, j), (i-1, j+1)$, or, $(i, j-1)$. From this it follows that each coordinate in the next state can differ from the corresponding coordinate of the current state by no more than ±1. This implies that $q((k, \ell)|(i, j), a) = 0$ for any $k \geq i + 2$ or for any $\ell \geq j + 2$, and therefore,

$$q((k, \ell)|(i, j), a)h(k, \ell) = 0 \leq \chi^{k+\ell} C,$$

for any constants $C > 0, \chi \in (0, 1)$, whenever $i \geq m = 0, j \geq n = 0$, and $k \geq n' + i = 2 + i$, or, $\ell \geq m' + j = 2 + j$, so that $\mathbf{B}(4)$ holds. ∎

Assumption C. Fix a stationary policy $f \in F$ and define $(\eta_{(i,j),(k,\ell)}(f)) := \lim_{t\to\infty} P(t, f)$. For any initial state $(i, j) \in \mathbb{X}$ we have $\sum_{(k,\ell)\in\mathbb{X}} \eta_{(i,j),(k,\ell)}(f) = 1$.

Proposition 2. *Under each fixed stationary policy, the Markov process generated by said policy has the following properties:*

1. *There is at most one recurrent class.*
2. *Under the assumption that* $\lambda\left(\frac{1}{\mu_1} + \frac{1}{\mu_2+\beta}\right) < 1$ *all states that communicate with* $(0, 0)$ *are positive recurrent. Moreover, transient states reach* $(0, 0)$ *in finite expected time.*
3. *If* $\lambda\left(\frac{1}{\mu_1} + \frac{1}{\mu_2+\beta}\right) < 1$ *then Assumption **C** holds.*

3.3.2.1.4 Proof. Under any stationary (non-idling) policy $(0, 0)$ is reachable from all states; Statement 1 holds. In fact, if π is a stationary policy and $\mathcal{G}(\pi)$ is the set of states that communicate with $(0, 0)$ for the Markov process generated by π, then $\{(i, j)|i \geq 0, j \in \{0, 1\}\} \subseteq \mathcal{G}(\pi)$. Using similar logic as that in (3.3) with

$s(i,j) = \frac{i}{\mu_1} + \frac{i+j}{\mu_2+\beta} + 1$ we have for any non-idling policy and $(i,j) \neq (0,0)$

$$\begin{aligned}
\sum_{(k,\ell)\in\mathbb{X}} q((k,\ell)|(i,j),a)s(i,j) &= \lambda\Big[\frac{1}{\mu_1} + \frac{1}{\mu_2+\beta}\Big] - j\beta\Big[\frac{1}{\mu_2+\beta}\Big] \\
&\quad - \mu_1 \mathbb{1}(a=1)\Big[p(\frac{1}{\mu_1} + \frac{1}{\mu_2+\beta}) + (1-p)(\frac{1}{\mu_1})\Big] \\
&\quad - \mu_2 \mathbb{1}(a=2)\Big[\frac{1}{\mu_2+\beta}\Big] \\
&= \lambda\Big[\frac{1}{\mu_1} + \frac{1}{\mu_2+\beta}\Big] - j\beta\Big[\frac{1}{\mu_2+\beta}\Big] \\
&\quad - \mu_1[\mathbb{1}(a=1) - 1 + 1]\Big[p(\frac{1}{\mu_2+\beta}) + (\frac{1}{\mu_1})\Big] \\
&\quad - \mu_2 \mathbb{1}(a=2)\Big[\frac{1}{\mu_2+\beta}\Big] \\
&= \lambda\Big[\frac{1}{\mu_1} + \frac{1}{\mu_2+\beta}\Big] - 1 \\
&\quad + [\mathbb{1}(a=2)]\Big[(\frac{p\mu_1 - \mu_2}{\mu_2+\beta}) + 1\Big] \\
&\quad - \Big[(\frac{p\mu_1}{\mu_2+\beta})\Big] - j\beta\Big[\frac{1}{\mu_2+\beta}\Big].
\end{aligned}$$

Since for the provider to work at station 2, $j \geq 1$, an upper bound of the quantity on the right hand side is $\lambda\Big[\frac{1}{\mu_1} + \frac{1}{\mu_2+\beta}\Big] - 1 < 0$, where the inequality holds by assumption. Applying an analogue to *Foster's criterion* for continuous-time processes (see [**?**]; Theorems 4.2 with $f = 1$) yields that the Markov process associated with any stationary policy has $(0,0)$ and all states that communicate with it as positive recurrent. Theorem 4.3(i) of [**?**] (again with $f = 1$) implies that under any stationary policy, the Markov process generated starting in any initial transient state reaches $(0,0)$ in finite expected time; Statement 2 of the proposition holds. Statement 3 is a direct consequence of the first two. ∎

Theorem 2. *If Assumptions **A**, **B**, and **C** above hold, then the following hold:*

1. *there exists a stationary policy f^*, a constant g^*, a function u^* on $\mathbb{X}$ and a decreas-*

ing sequence $\{\alpha_k\}$ *tending to zero, such that for* $(i, j) \in \mathbb{X}$,

(a) $g^* = \lim_{k\to\infty} \alpha_k v^*_{\alpha_k}(i, j)$, $u^*(i, j) = \lim_{k\to\infty} u_{\alpha_k}(i, j)$;

(b)

$$\begin{aligned} g^* &= r((i, j), f^*(i, j)) + \sum_{(k,\ell)\in\mathbb{X}} q((k, \ell)|(i, j), f^*(i, j))u^*(k, \ell) \\ &= \max_{a\in A(i,j)} \{r((i, j), a) + \sum_{(k,\ell)\in\mathbb{X}} q((k, \ell)|(i, j), a)u^*(k, \ell)\}; \end{aligned} \tag{3.6}$$

2. f^* *is an average reward optimal and satisfies* $\rho^{f^*}(i, j) = g^*$, $(i, j) \in \mathbb{X}$, *where ;*

3. *any* f *realizing the maximum on the right-hand side of* (3.6) *is average reward optimal.*

3.3.2.1.5 Proof. The proof of 1(a) and 1(b) is the same as the proof of (i) in Theorem 1 that appears in [**?**]. Next, we want to show that f^* is an average reward optimal policy and satisfies

$$\rho^{f^*}(i, j) = g^*,$$

for all $(i, j) \in \mathbb{X}$. Recall we have assumed that

$$\sum_{(k,\ell)\in\mathbb{X}} \eta_{(i,j),(k,\ell)}(f) = 1,$$

for all $(i, j) \in \mathbb{X}, f \in F$. We prove that

$$\sum_{(k,\ell)\in\mathbb{X}} \eta_{(i,j),(k,\ell)}(f^*) r((k, \ell), f^*(k, \ell)) \geq g^*.$$

For now assume $u^* \geq 0$. Pursuant to Assumption **B**(3), assume that for each fixed j we have $\sum_{i=1}^{\infty} \phi_{i,j} = \infty$. Now note that for N, M large,

$$
\begin{aligned}
C_{N,M} &:= \sum_{k=0}^{N} \sum_{\ell=0}^{M} \eta_{(k,\ell)}(f^*)\Big[g^* - r((k,\ell), f^*(k,\ell))\Big] \\
&= \sum_{k=0}^{N} \sum_{\ell=0}^{M} \eta_{(k,\ell)}(f^*)\Big[\sum_{(r,s)\in\mathbb{X}} q((r,s)|(k,\ell), f^*(k,\ell))u^*(r,s)\Big] \\
&= \sum_{k=0}^{N} \sum_{\ell=0}^{M} \eta_{(k,\ell)}(f^*)\Big[q\,((k,\ell)|(k,\ell), f^*(k,\ell))\,u^*(k,\ell) \\
&\quad + \sum_{\substack{(r,s)\in\mathbb{X} \\ (r,s)\neq(k,\ell)}} q\,((r,s)|(k,\ell), f^*(k,\ell))\,u^*(r,s)\Big].
\end{aligned} \tag{3.7}
$$

We can expand the summation inside the brackets to rewrite the right hand side of (3.7):

$$
\begin{aligned}
C_{N,M} &= \sum_{k=0}^{N} \sum_{\ell=0}^{M} \eta_{(k,\ell)}(f^*)\Big[\sum_{r=0}^{\infty} \sum_{s=M+1}^{\infty} q((r,s)|(k,\ell), f^*(k,\ell))u^*(r,s) \\
&\quad + \sum_{r=N+1}^{\infty} \sum_{s=0}^{M} q((r,s)|(k,\ell), f^*(k,\ell))u^*(r,s)\Big] \\
&\quad + \sum_{k=0}^{N} \sum_{\ell=0}^{M} \eta_{(k,\ell)}(f^*)\Big[q((k,\ell)|(k,\ell), f^*(k,\ell))u^*(k,\ell) \\
&\quad + \sum_{r=0}^{k-1} \sum_{s=0}^{M} q((r,s)|(k,\ell), f^*(k,\ell))u^*(r,s) \\
&\quad + \sum_{\substack{s=0 \\ s\neq\ell}}^{M} q((k,s)|(k,\ell), f^*(k,\ell))u^*(r,s) \\
&\quad + \sum_{r=k+1}^{N} \sum_{s=0}^{M} q((r,s)|(k,\ell), f^*(k,\ell))u^*(r,s)\Big].
\end{aligned} \tag{3.8}
$$

Let $\mathbb{X}_{n,m} := \{(i,j) \in \mathbb{X} | i \leq n, j \leq m\}$ so that we may simplify (3.8)

$$
\begin{aligned}
C_{N,M} &= \sum_{k=0}^{N} \sum_{\ell=0}^{M} \eta_{(k,\ell)}(f^*)\Big[\sum_{r=0}^{\infty} \sum_{s=M+1}^{\infty} q((r,s)|(k,\ell), f^*(k,\ell))u^*(r,s) \\
&\quad + \sum_{r=N+1}^{\infty} \sum_{s=0}^{M} q((r,s)|(k,\ell), f^*(k,\ell))u^*(r,s)\Big] \\
&\quad + \sum_{k=0}^{N} \sum_{\ell=0}^{M} \eta_{(k,\ell)}(f^*)\Big[q((k,\ell)|(k,\ell), f^*(k,\ell))u^*(k,\ell) \\
&\quad + \sum_{(r,s)\in\mathbb{X}_{N,M}\setminus\{(k,\ell)\}} q((r,s)|(k,\ell), f^*(k,\ell))u^*(r,s)\Big].
\end{aligned}
$$

We will demonstrate that (3.8) can be bounded above by

$$\sum_{k=0}^{N}\sum_{\ell=0}^{M}\eta_{(k,\ell)}(f^*)\Big[\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty}q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)+\sum_{r=N+1}^{\infty}\sum_{s=0}^{M}q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big]$$

by showing that

$$\begin{aligned}D_{N,M} := \sum_{k=0}^{N}\sum_{\ell=0}^{M}\eta_{(k,\ell)}(f^*)\Big[&q((k,\ell)|(k,\ell),f^*(k,\ell))u^*(k,\ell)\\ &+\sum_{(r,s)\in\mathbb{X}_{N,M}\setminus\{(k,\ell)\}}q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big]\le 0.\end{aligned}$$

From an interchange of the summations,

$$\begin{aligned}D_{N,M} &= \sum_{r=0}^{N}\sum_{s=0}^{M}u^*(r,s)\Big[\eta_{(r,s)}(f^*)q((r,s)|(r,s),f^*(r,s))\\ &\quad+\sum_{(k,\ell)\in\mathbb{X}_{N,M}\setminus\{(r,s)\}}\eta_{(k,\ell)}(f^*)q((r,s)|(k,\ell),f^*(k,\ell))\Big]\\ &\le \sum_{r=0}^{N}\sum_{s=0}^{M}u^*(r,s)\Big[\eta_{(r,s)}(f^*)q((r,s)|(r,s),f^*(r,s))\\ &\quad+\sum_{(k,\ell)\in\mathbb{X}\setminus\{(r,s)\}}\eta_{(k,\ell)}(f^*)q((r,s)|(k,\ell),f^*(k,\ell))\Big].\end{aligned}$$

Since $\{\eta_{(i,j)}|i\ge 0, j\ge 0\}$ is a limiting distribution, it is also a stationary distribution and the term inside the brackets is zero. Using this in (3.7), we arrive at the inequality:

$$\begin{aligned}C_{N,M} &= \sum_{k=0}^{N}\sum_{\ell=0}^{M}\eta_{(k,\ell)}(f^*)\Big[g^*-r((k,\ell),f^*(k,\ell))\Big]\\ &\le \sum_{k=0}^{N}\sum_{\ell=0}^{M}\eta_{(k,\ell)}(f^*)\Big[\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty}q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\\ &\quad+\sum_{r=N+1}^{\infty}\sum_{s=0}^{M}q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big].\end{aligned}$$

Next, choose $N>n$, $M>\max\{m,\hat{v}\}$ to obtain

$$\begin{aligned}C_{N,M} \le \sum_{k=0}^{n}\sum_{\ell=0}^{M}\eta_{(k,\ell)}(f^*)\Big[&\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty}q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\\ &+\sum_{r=N+1}^{\infty}\sum_{s=0}^{M}q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big]\\ +\sum_{k=n+1}^{N}\sum_{\ell=0}^{M}\eta_{(k,\ell)}(f^*)\Big[&\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty}q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\\ &+\sum_{r=N+1}^{\infty}\sum_{s=0}^{M}q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big].\end{aligned}$$

This can be rewritten as

$$
\begin{aligned}
C_{N,M} \le \sum_{k=0}^{n}\sum_{\ell=0}^{M} \eta_{(k,\ell)}(f^*)\Big[\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty} & q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s) \\
& + \sum_{r=N+1}^{\infty}\sum_{s=0}^{M} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big] \\
& + \sum_{k=n+1}^{N}\sum_{\ell=0}^{M} \eta_{(k,\ell)}(f^*)\Big[\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s) \\
& + \sum_{r=N+1}^{\infty}\sum_{s=0}^{M} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big] \\
= \sum_{k=0}^{n}\sum_{\ell=0}^{M} \eta_{(k,\ell)}(f^*)\Big[\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty} & q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s) \\
& + \sum_{r=N+1}^{\infty}\sum_{s=0}^{M} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big] \\
& + \sum_{k=n+1}^{N}\sum_{\ell=0}^{M} \eta_{(k,\ell)}(f^*)\Big[\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s) \\
& + \sum_{r=N+1}^{N+n'} q((r,M)|(k,\ell),f^*(k,\ell))u^*(r,M) \\
& + \sum_{s=0}^{M-1}\sum_{r=N+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s) \\
& + \sum_{r=N+n'+1}^{\infty} q((r,M)|(k,\ell),f^*(k,\ell))u^*(r,M)\Big]
\end{aligned}
$$

where n' is from $\mathbf{B}(4)$. Using the fact that h bounds u^* (by Assumption $\mathbf{B}(1)$) and applying $\mathbf{B}(4)$ yields the following bounds

$$
\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s) \le \frac{\chi^{M+1}C}{(1-\chi)^2},
$$

$$
\sum_{s=0}^{M-1}\sum_{r=N+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s) \le \sum_{s=0}^{M-1}\frac{\chi^{N+1+s}C}{(1-\chi)} \le \frac{\chi^{N+1}C}{(1-\chi)^2},
$$

and

$$
\sum_{r=N+n'+1}^{\infty} q((r,M)|(k,\ell),f^*(k,\ell))u^*(r,M) \le \frac{\chi^{N+n'+1+M}C}{(1-\chi)}
$$

Therefore,

$$
\begin{aligned}
&\sum_{k=0}^{n}\sum_{\ell=0}^{m} \eta_{(k,\ell)}(f^*)\Big[\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s) + \sum_{r=N+1}^{\infty}\sum_{s=0}^{M} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big] \\
&\quad + \sum_{(k,\ell)\in X_{N,M}\setminus X_{n,m}} \eta_{(k,\ell)}(f^*)\Big[\sum_{(r,s)\in \mathbb{X}_{N+n',M+n'}\setminus \mathbb{X}_{N,M}} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s) \\
&\quad + \sum_{r=0}^{\infty}\sum_{s=M+m'+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s) + \sum_{s=0}^{M+m'+1}\sum_{r=N+n'+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big]
\end{aligned}
$$

is bounded above by

$$
\begin{aligned}
C_{N,M} \leq \sum_{k=0}^{n}\sum_{\ell=0}^{M}\eta_{(k,\ell)}(f^*)\Big[&\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\\
&+\sum_{r=N+1}^{\infty}\sum_{s=0}^{M} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big]\\
&+\sum_{k=n+1}^{N}\sum_{\ell=0}^{M}\eta_{(k,\ell)}(f^*)\Big[\sum_{r=N+1}^{N+n'} q((r,M)|(k,\ell),f^*(k,\ell))u^*(r,M)\Big]\\
&+\frac{\chi^{M+1}C}{(1-\chi)^2}+\frac{\chi^{N+1}C}{(1-\chi)^2}+\frac{\chi^{N+n'+1+M}C}{(1-\chi)}.
\end{aligned} \tag{3.9}
$$

Recalling that $\{\eta_{(i,j)}|i\geq 0, j\geq 0\}$ is a stationary distribution, we have the following (for any fixed (r,M))

$$
\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\eta_{(k,\ell)}(f^*)q((r,M)|(k,\ell),f^*(k,\ell)) = 0.
$$

In particular for $r > N$, since the off-diagonal terms of the rate matrix $Q(f^*)$ are non-negative

$$
\sum_{k=n+1}^{N}\sum_{\ell=0}^{M}\eta_{(k,\ell)}(f^*)q((r,M)|(k,\ell),f^*(k,\ell)) \leq -\eta_{(r,M)}(f^*)q((r,M)|(r,M),f^*(r,M)).
$$

Interchanging the order of summation and using this (3.9) yields

$$
\begin{aligned}
C_{N,M} \leq \sum_{k=0}^{n}\sum_{\ell=0}^{M}\eta_{(k,\ell)}(f^*)\Big[&\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\\
&+\sum_{r=N+1}^{\infty}\sum_{s=0}^{M} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big]\\
&+\sum_{r=N+1}^{N+n'}\eta_{(r,M)}(f^*)\Big[-q((r,M)|(r,M),f^*(r,M))u^*(r,M)\Big]\\
&+\frac{\chi^{M+1}C}{(1-\chi)^2}+\frac{\chi^{N+1}C}{(1-\chi)^2}+\frac{\chi^{N+n'+1+M}C}{(1-\chi)}.
\end{aligned}
$$

From Assumption **B**(2), $\sum_{(r,s)\in\mathbb{X}} q((r,s)|(k,\ell),f^*(k,l))u^*(r,s) < \infty$. So the second term tends to zero as N tend to ∞. Letting N tend to ∞, we obtain

$$
\begin{aligned}
\lim_{N\to\infty} C_{N,M} \leq \sum_{k=0}^{n}\sum_{\ell=0}^{M}\eta_{(k,\ell)}(f^*)\Big[&\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big]\\
&+\liminf_{N}\Big[\sum_{(k,\ell)\in\mathbb{X}_{N+n',M}\setminus\mathbb{X}_{N,M}}\eta_{(k,\ell)}(f^*)\Big(-q((k,\ell)|(k,\ell),f^*(k,\ell))u^*(k,\ell)\Big)\Big]\\
&+\frac{\chi^{M+1}C}{(1-\chi)^2}.
\end{aligned}
$$

Let

$$D_N := \sum_{(k,\ell)\in\mathbb{X}_{N+n',M}\setminus\mathbb{X}_{N,M}} \eta_{(k,\ell)}(f^*)\big[-q((k,\ell)|(k,\ell),f^*(k,\ell))u^*(k,\ell)\big]\Big]$$

We would like to show $\liminf_N D_N = 0$. Suppose, for a contradiction, that this is not the case. This means that there exists $\epsilon_0 > 0$ such that $\liminf_N D_N > \epsilon_0$. This in turn implies that there exist a natural number N_0 such that $D_N > \epsilon_0$ whenever $N \geq N_0$. Without loss of generality, assume that $N_0 \geq \hat{u}$. It follows that there is an (infinite) subsequence $\{N_i; i \geq 0\} \subset \mathbb{N}$ such that $N_i \to \infty$, $N_{i+1} - N_i = n'$, and $D_{N_i} > \epsilon_0$ for all i. Assumption **B**(3) yields

$$\begin{aligned}
&\sum_{(k,\ell)\in\mathbb{X}_{N_i+n',M}\setminus\mathbb{X}_{N_i,M}} \eta_{(k,\ell)}(f^*)r((k,\ell),f^*(k,\ell)) \\
&\qquad \geq \sum_{(k,\ell)\in\mathbb{X}_{N_i+n',M}\setminus\mathbb{X}_{N_i,M}} \eta_{(k,\ell)}(f^*)q(k,\ell)h(k,\ell)\phi_{k,\ell} \\
&\qquad \geq \sum_{(k,\ell)\in\mathbb{X}_{N_i+n',M}\setminus\mathbb{X}_{N_i,M}} \eta_{(k,\ell)}(f^*)[-q((k,\ell)|(k,\ell),f^*(k,\ell))u^*(k,\ell)]\phi_{N_i+n',M} \\
&\qquad = D_{N_i}\phi_{N_i+n',M} \\
&\qquad \geq \epsilon_0\phi_{N_i+n',M},
\end{aligned}$$

where the second inequality follows from the assumptions that h bounds u^* and ϕ is non-decreasing. Summing over i and noting that r and η are non-negative yields,

$$\begin{aligned}
&\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} \eta_{(k,\ell)}(f^*)r((k,\ell),f^*(r,s)) \\
&\qquad \geq \sum_i \sum_{(k,\ell)\in\mathbb{X}_{N_i+n',M}\setminus\mathbb{X}_{N_i,M}} \eta_{(k,\ell)}(f^*)r((k,\ell),f^*(k,\ell)) \geq \epsilon \sum_i \phi_{N_i+n',M}.
\end{aligned}$$

Since $\{\phi_{i,j}\}$ is such that $\sum_{i=1}^{\infty}\phi_{i,j} = \infty$ for all j, the limit comparison test implies that $\sum_i \phi_{N_i+n',M} = \infty$. Hence,

$$\sum_{(k,\ell)\in\mathbb{X}} \eta_{(k,\ell)}(f^*)r((k,\ell),f^*(r,s)) = \infty,$$

which is a contradiction (since r is bounded). Thus, $\liminf_N D_N = 0$. As a result

$$\begin{aligned}\lim_{N\to\infty} C_{N,M} &\leq \sum_{k=0}^{n}\sum_{\ell=0}^{M}\eta_{(k,\ell)}(f^*)\Big[\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big] \\ &\quad + \frac{\chi^{M+1}C}{(1-\chi)^2}. \\ &\leq \Big[\sum_{r=0}^{\infty}\sum_{s=M+1}^{\infty} q((r,s)|(k,\ell),f^*(k,\ell))u^*(r,s)\Big] \\ &\quad + \frac{\chi^{M+1}C}{(1-\chi)^2}.\end{aligned}$$

From Assumption **B**(2), $\sum_{(r,s)\in\mathbb{X}} q((r,s)|(k,\ell),f^*(k,l))u^*(r,s) < \infty$. So the first term tends to zero as M tends to ∞. Since $\chi < 1$, the second term also tends to zero as M tends to ∞. Therefore, letting M tend to ∞, we obtain

$$\lim_{M\to\infty}\Big[\lim_{N\to\infty} C_{N,M}\Big] \leq 0.$$

In other words,

$$g^* \leq \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\eta_{(i,j)}(f^*)r((i,j),f^*(i,j)).$$

The other direction follows from Lemma 6 of [**?**]. The boundedness of the rewards implies that

$$\sum_{(k,\ell)\in\mathbb{X}}\eta_{(i,j),(k,\ell)}(f^*)r((k,\ell),f^*(k,\ell) < \infty.$$

Since

$$\sum_{(k,\ell)\in\mathbb{X}} q((k,\ell)|(i,j),f^*((i,j)) = 0,$$

and u^* is bounded below, the components of u^* may be increased by adding any constant without affecting the proof. Hence, the assumption that $u^* \geq 0$ is without loss of generality. The remaining parts of the proof follow exactly the same as those in [**?**]. ■

3.4 Optimal Control

The non-idling result of Proposition 1 leads to the next result which basically says that ignoring one station completely is sub-optimal.

Proposition 3. *In the case of the infinite horizon discounted reward or average reward, if $\lambda < \mu_1$ and $p > 0$, then there is an optimal policy that moves the provider from one station to the other at least once.*

3.4.0.1.6 Proof. We need to show that there exists a policy that does as well (or better) than either the policy that works solely at station 1 (letting every potential abandonment occur) or the policy that works solely at station 2. Note that $\lambda < \mu_1$ implies that under the policy that works at station 1 only, the first station acts like a (stable) $M/M/1$ queueing system. Since $p > 0$, it is well-known that the exit process of the first station (those patients that enter the second station) is a Poisson process with rate λp. Moreover, the second station acts as an $M/M/\infty$ queue with each abandonment acting as a provider. That is to say, the joint process (taking both stations together) is an ergodic continuous-time Markov chain. In particular, any state of the form $(0, j)$, for $j \geq 1$ is eventually reached. The fact that there is an optimal non-idling policy now yields the result.

Consider now the system that works only at station 2. In this case, station 2 acts as a (linear) pure death process, which eventually empties station 2. At this time, the system is in some state $(i, 0)$. If $i \geq 1$ the previous results on non-idling policies yield the result. If $i = 0$, after the first arrival the results apply. ∎

The next result, the main result of the section, provides conditions under which it is optimal to prioritize station 2 under the discounted and average reward criteria.

Theorem 3. *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 $\lambda\left(\frac{1}{\mu_1} + \frac{1}{\mu_2+\beta}\right) < 1$ and $\beta > 0$, it is optimal to serve at station 2 whenever station 2 is not empty.*

3.4.0.1.7 Proof. See Section 3.4.1 ■

A few remarks regarding Theorem 3 are in order. 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.

It should also be noted that the policy that prioritizes station 2 by no means ignores station 1. On the contrary, the set of recurrent states under this policy is $\{(i, j) | i \geq 0, j \in \{0, 1\}\}$ so that customers receive service at station 1, move to station 2 and leave the system; the provider serves a customer completely before starting on the next customer. This leads to several questions

- Do we need the assumption that $\lambda\left(\frac{1}{\mu_1} + \frac{1}{\mu_2+\beta}\right) < 1$ when the system is clearly stabilizable with $\frac{\lambda}{\mu_1} < 1$ (just prioritize station 1)?

- Is it solely the abandonments that make the c-μ result fail?
- Under what conditions is it optimal to prioritize station 1?

The answer to the first question is "no" due to the fact that the policy that prioritizes station 2 is not stable without this assumption. We will study it further numerically. To answer the second question consider the following result.

Proposition 4. *Suppose $\beta = 0$ and $\mu_1 R_1 > \mu_2 R_2$, then under either the α–discounted reward or the average reward criterion (with $\lambda < \mu_1$), it is optimal to serve at station 1 whenever station 1 is not empty.*

3.4.0.1.8 Proof. The result is shown in the discounted reward case. The average case is similar. In this case the problem is uniformizable with uniformization rate $\Psi := \lambda + \mu_1 + \mu_2$ and discrete-time discount factor $\gamma = \frac{\Psi+\alpha}{\Psi}$. The optimal policies are unchanged between the continuous and discrete time models and the optimal values coincide up to a multiplicative constant. Without loss of generality assume that $\Psi = 1$. The discrete time DROE mapping for any function h on $\mathbb{X}$ is now

$$\begin{aligned} Uh(0,j) &= \gamma(\lambda[h(1,j) - h(0,j)] + h(0,j)) \\ &\qquad + \mathbb{1}(j \geq 1)\mu_2[R_2 + \gamma(h(0,j-1) - h(0,j))]\} \\ Uh(i,0) &= \gamma[\lambda(h(i+1,0) - h(i,0)) + h(i,0)] \\ &\qquad + \mathbb{1}(i \geq 1)\mu_1[R_1 + \gamma(ph(i-1,1) + (1-p)h(i-1,0) - h(i,0))]. \end{aligned}$$

For $i, j \geq 1$,

$$\begin{aligned} Uh(i,j) &= \lambda h(i+1,j) + j\beta h(i,j-1) - (\lambda + j\beta)h(i,j) \\ &\qquad + \max\{\mu_1[R_1 + ph(i-1,j+1) + (1-p)h(i-1,j) - h(i,j)], \\ &\qquad \mu_2[R_2 + h(i,j-1) - h(i,j)]\}. \end{aligned}$$

The discrete-time value function satisfies $\tilde{v}_\gamma = U\tilde{v}_\gamma$ and can be computed via successive approximations $\tilde{v}_{n+1,\gamma} = U\tilde{v}_{n,\gamma}$ with $\tilde{v}_{0,\gamma} = 0$ and $\lim_{n\to\infty} \tilde{v}_{n,\gamma} = \tilde{v}_\gamma$. An argument similar to that in Theorem 3 implies that in state (i, j) it is optimal to assign the worker to station 1 in state (i, j) if

$$\mu_1 R_1 - \mu_2 R_2 + \mu_1[(p\tilde{v}_\gamma(i-1, j+1) + q\tilde{v}_\gamma(i-1, j)) - \tilde{v}_\gamma(i, j)] + \mu_2[\tilde{v}_\gamma(i, j) - \tilde{v}_\gamma(i, j-1)] \geq 0. \quad (3.10)$$

We thus prove (3.10) with $\tilde{v}_\gamma$ replaced with $\tilde{v}_{n,\gamma}$ via a sample path argument and induction. Trivially, (3.10) holds for $\tilde{v}_{0,\gamma}$. Assume that the inequality (3.10) holds for $\tilde{v}_{n,\gamma}$ and consider it for $\tilde{v}_{n+1,\gamma}$.

Suppose we start five processes in the same probability space. Processes 1-5 begin at time $n + 1$ in states $(i-1, j+1), (i-1, j), (i, j), (i, j)$, and $(i, j-1)$, respectively. The inductive hypothesis implies that it is optimal to provide service at station 1 when there is work to do there. Processes 3 and 5 use 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

$$\begin{aligned} \mu_1 R_1 - \mu_2 R_2 + \mu_1[(p\tilde{v}^{\pi_1}_{n+1,\gamma}(i-1, j+1) + q\tilde{v}^{\pi_2}_{n+1,\gamma}(i-1, j)) - \tilde{v}^{\pi_3}_{n+1,\gamma}(i, j)] \\ + \mu_2[\tilde{v}^{\pi_4}_{n+1,\gamma}(i, j) - \tilde{v}^{\pi_5}_{n+1,\gamma}(i, j-1)] \geq 0. \end{aligned} \quad (3.11)$$

Since policies π_1, π_2, and π_4 are potentially sub-optimal, (3.10) follows from (3.11).

In any of the cases that $i \geq 2$ each process serves at station 1. So, each process moves to the new state with the relative position of each intact. The inductive hypothesis now yields the result. Similarly if there is an arrival, 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; the inductive hypothesis yields the result. It remains to consider the cases where one or more of the processes cannot serve at station 1.

Case 1. *Suppose that policies π_1 and π_2 have the provider work at station 2 and policies π_3 and π_5 all have the provider work at station 1. In this case, let π_4 have the provider 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 (3.11) are

$$\mu_1\Big[\mu_1 R_1 - \mu_2 R_2 - \mu_1 R_1 \\ + \mu_2(p\tilde{v}_{n,\gamma}(i-1, j+1) + q\tilde{v}_{n,\gamma}(i-1, j) - p\tilde{v}_{n,\gamma}(i-1, j) - q\tilde{v}_{n,\gamma}(i-1, j-1))\Big]. \tag{3.12}$$

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 (3.11) are

$$\mu_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1 R_2 \\ + \mu_1((p\tilde{v}_{n,\gamma}(i-1, j) + q\tilde{v}_{n,\gamma}(i-1, j-1)) - \tilde{v}_{n,\gamma}(i, j)) \\ + \mu_2(\tilde{v}_{n,\gamma}(i, j) - \tilde{v}_{n,\gamma}(i, j-1))\Big]. \tag{3.13}$$

Adding expressions (3.12) and (3.13) yields

$$\mu_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1(p\tilde{v}_{n,\gamma}(i-1, j+1) + q\tilde{v}_{n,\gamma}(i-1, j) - \tilde{v}_{n,\gamma}(i, j)) + \mu_2(\tilde{v}_{n,\gamma}(i, j) - \tilde{v}_{n,\gamma}(i, j-1))\Big] \tag{3.14}$$

Note that expression (3.14) is implied by the left hand side of inequality (3.11) and the inductive hypothesis. That is, we may restart the argument from here.

Case 2. *Suppose that policies π_1 and π_2 have the provider work at station 2 and policies π_3 and π_5 have the provider work at station 1 and 2, respectively. In this case, let π_4 have the provider 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 (3.11) are

$$\begin{aligned}\mu_1\Big[&\mu_1 R_1 - \mu_2 R_2 - \mu_1 R_1 + \mu_2 R_1 \\ &+ \mu_2(p\tilde{v}_{n,\gamma}(i-1, j+1) + q\tilde{v}_{n,\gamma}(i-1, j) - \tilde{v}_{n,\gamma}(i, j-1))\Big].\end{aligned} \tag{3.15}$$

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 (3.11) are

$$\begin{aligned}\mu_2\Big[&\mu_1 R_1 - \mu_2 R_2 + \mu_1 R_2 - \mu_2 R_2 \\ &+ \mu_1((p\tilde{v}_{n,\gamma}(i-1, j) + q\tilde{v}_{n,\gamma}(i-1, j-1)) - \tilde{v}_{n,\gamma}(i, j)) \\ &+ \mu_2(\tilde{v}_{n,\gamma}(i, j) - \tilde{v}_{n,\gamma}(i, j-2))\Big].\end{aligned} \tag{3.16}$$

Adding expressions (3.15) and (3.16) yields

$$\begin{aligned}\mu_2\Big[&\mu_1 R_1 - \mu_2 R_2 + \mu_1(p\tilde{v}_{n,\gamma}(i-1, j+1) + q\tilde{v}_{n,\gamma}(i-1, j) - \tilde{v}_{n,\gamma}(i, j)) \\ &+ \mu_2(\tilde{v}_{n,\gamma}(i, j) - \tilde{v}_{n,\gamma}(i, j-1))\Big] \\ &+ \mu_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1(p\tilde{v}_{n,\gamma}(i-1, j) + q\tilde{v}_{n,\gamma}(i-1, j-1) - \tilde{v}_{n,\gamma}(i, j-1)) \\ &+ \mu_2(\tilde{v}_{n,\gamma}(i, j-1) - \tilde{v}_{n,\gamma}(i, j-2))\Big].\end{aligned}$$

Each term above is non-negative by the inductive hypothesis.

Case 3. *Suppose that policies π_3 and π_5 have the provider work at station 2 and 1, respectively. In this case, let policies π_1, π_2, and π_4 have the provider 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 (3.11) are

$$\mu_1\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1(p\tilde{v}_{n,\gamma}(i-1,j+1) + q\tilde{v}_{n,\gamma}(i-1,j) - \tilde{v}_{n,\gamma}(i,j))$$
$$+ \mu_2(p\tilde{v}_{n,\gamma}(i-1,j+1) + q\tilde{v}_{n,\gamma}(i-1,j) - p\tilde{v}_{n,\gamma}(i-1,j) - q\tilde{v}_{n,\gamma}(i-1,j-1))\Big]. \tag{3.17}$$

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 (3.11) are

$$\mu_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1((p\tilde{v}_{n,\gamma}(i-1,j) + q\tilde{v}_{n,\gamma}(i-1,j-1)) - \tilde{v}_{n,\gamma}(i,j-1))$$
$$+ \mu_2(\tilde{v}_{n,\gamma}(i,j) - \tilde{v}_{n,\gamma}(i,j-1))\Big]. \tag{3.18}$$

Adding expressions (3.17) and (3.18) yields

$$\mu_1\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1(p\tilde{v}_{n,\gamma}(i-1,j+1) + q\tilde{v}_{n,\gamma}(i-1,j) - \tilde{v}_{n,\gamma}(i,j))$$
$$+ \mu_2(\tilde{v}_{n,\gamma}(i,j) - \tilde{v}_{n,\gamma}(i,j-1))\Big]$$
$$+ \mu_2\Big[\mu_1 R_1 - \mu_2 R_2 + \mu_1(p\tilde{v}_{n,\gamma}(i-1,j+1) + q\tilde{v}_{n,\gamma}(i-1,j) - \tilde{v}_{n,\gamma}(i,j))$$
$$+ \mu_2(\tilde{v}_{n,\gamma}(i,j) - \tilde{v}_{n,\gamma}(i,j-1))\Big].$$

Again, each term is non-negative by the inductive hypothesis. The result follows. ∎

Of course the issue with the policy that prioritizes station 1 is that it is also the policy that spends the least time (on average) at station 2; hardly a fair policy.

3.4.1 Proof of Theorem 3

3.4.1.1 Proof under the discounted reward criterion

3.4.1.1.1 Proof. Again, by a sample path argument. Fix $i, j \geq 1$ and start two processes on the same probability space. Process 1 assigns the provider to station 1, while Process 2 assigns it to station 2. After the first event, both processes follow an optimal policy. We show that the expected reward of Process 2, is at least as high as that of Process 1. The principle of optimality then yields the result.

If the first event is an arrival, both processes enter state $(i+1, j)$, we relabel this as the new starting state and restart the argument. Similarly, if there is an abandonment the two processes enter state $(i, j-1)$, again we relabel the starting states and continue. Now, with probability $\frac{\mu_1}{\lambda+j\beta+\mu_1+\mu_2}$ the service completion for station 1 occurs first. Process 1 accrues a reward and sees the service, Process 2 does not. The difference in the (expected) rewards is

$$B_1(i,j) := \Big(\frac{\mu_1}{\lambda + j\beta + \mu_1 + \mu_2}\Big)\Big(R_1 + pv_\alpha(i-1, j+1) + (1-p)v_\alpha(i-1, j) - v_\alpha(i,j)\Big)$$

If the service at station 2 completes first the difference is

$$B_2(i,j) := \Big(\frac{\mu_2}{\lambda + j\beta + \mu_1 + \mu_2}\Big)\Big(v_\alpha(i,j) - R_2 - v_\alpha(i, j-1)\Big).$$

Thus, if $B_1 + B_2 \leq 0$ the result holds. A little algebra yields that

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

for all $i, j \geq 1$. We continue to follow the sample paths of five processes (on the same probability space) to show (3.19) 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

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

Since π_1 and π_3 are potentially sub-optimal, (3.19) follows from (3.20).

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

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 had started in these states.

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

In this case, it follows that the remaining rewards in the left side of inequality (3.20) are

$$\begin{aligned} &\mu_2 R_2 - \mu_1 R_1 + \mu_2[v_\alpha(i, j-1) - v_\alpha(i, j)] \\ &\qquad + \mu_1[v_\alpha(i, j) - v_\alpha(i-1, j)]. \end{aligned} \tag{3.21}$$

Since (3.21) is greater than or equal to the expression on the left side of the inequality in (3.20), it follows that it is enough to show (3.20) 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 (3.20) for all other remaining cases.

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

From this point on assume $\pi_1 = \pi_2$, so that the first two terms in inequality (3.20) are the same. To complete the proof it suffices to show that the remaining rewards are non-negative

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

There are several subcases to consider.

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

After state transitions assume π_3 uses a stationary optimal policy. Thus, (3.22) becomes

$$\begin{aligned}
&\mu_2 R_2 - \mu_1 R_1 + \mu_1[v_\alpha^{\pi_3}(i, j-1) - (pv_\alpha^{\pi_4}(i-1, j-1) + qv_\alpha^{\pi_5}(i-1, j-1))] \\
&\quad = \mu_2 R_2 - \mu_1 R_1 + \mu_1[v_\alpha(i, j-1) - (pv_\alpha(i-1, j-1) + qv_\alpha(i-1, j-1))] \\
&\quad = \mu_2 R_2 - \mu_1 R_1 + \mu_1[v_\alpha(i, j-1) - v_\alpha(i-1, j-1)] \geq 0,
\end{aligned} \tag{3.23}$$

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 provider to work at station 1 and 2, respectively. Assume that π_3 also works at station 1.*

Assume that the next event is a service completion at station 1 for Process 4 (with probability $\frac{\mu_1}{\lambda+\mu_1+\mu_2+j\beta}$). Suppressing the denominator of the probability, the left hand side of inequality (3.22) becomes

$$\mu_1\Big[\mu_2R_2 - \mu_1R_1 + p\mu_1[qR_1 + v_\alpha^{\pi_3}(i-1,j) - (pv_\alpha^{\pi_4}(i-2,j+1) + qv_\alpha^{\pi_5}(i-1,j-1))] \\ + q\mu_1[qR_1 + v_\alpha^{\pi_3}(i-1,j-1) - (pv_\alpha^{\pi_4}(i-2,j) + qv_\alpha^{\pi_5}(i-1,j-1))]\Big]$$

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

$$\mu_1\Big[\mu_2R_2 - p\mu_1R_1 + p\mu_1[v_\alpha(i-1,j) - (pv_\alpha(i-2,j+1) + qv_\alpha(i-2,j))]\Big]. \tag{3.24}$$

If, however, the next event is a service completion at station 2 in Process 5, then the left hand side of inequality (3.22) becomes

$$\mu_2\Big[\mu_2R_2 - \mu_1R_1 + \mu_1[v_\alpha^{\pi_3}(i,j-1) - (pv_\alpha^{\pi_4}(i-1,j) + qR_2 + qv_\alpha^{\pi_5}(i-1,j-2))]\Big].$$

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

$$\mu_2\Big[\mu_2R_2 - \mu_1R_1 - q\mu_1R_2 + \mu_1[v_\alpha(i,j-1) - (pv_\alpha(i-1,j) + qv_\alpha(i-1,j-2))]\Big]. \tag{3.25}$$

Adding (3.24) and (3.25) (with a little algebra) we get

$$p\mu_1[\mu_2R_2 - \mu_1R_1 + \mu_1(v_\alpha(i-1,j) - (pv_\alpha(i-2,j+1) + qv_\alpha(i-2,j)))] \\ + \mu_2[\mu_2R_2 - \mu_1R_1 + \mu_1(v_\alpha(i,j-1) - (pv_\alpha(i-1,j) + qv_\alpha(i-1,j-1)) \\ + qv_\alpha(i-1,j-1) - qv_\alpha(i-1,j-2))].$$

The first expression is (3.22) evaluated at $(i-1, j+1)$. The second is (3.22) 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 provider 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 $\frac{\mu_2}{\lambda+\mu_1+\mu_2+j\beta}$), then the remaining rewards of the processes in (3.22) are (suppressing the denominator)

$$\mu_2\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_1(v_\alpha^{\pi_3}(i, j-1) - p v_\alpha^{\pi_4}(i-1, j-1) - pR_2 - q v_\alpha^{\pi_5}(i-1, j-1))\Big]$$
$$= \mu_2[\mu_2 R_2 - \mu_1 R_1 - p\mu_1 R_2 + \mu_1(v_\alpha^{\pi_3}(i, j-1) - v_\alpha(i-1, j-1))] \qquad (3.26)$$

If, however, the next event is a service completion at station 1 in Processes 3 and 5, then from (3.22) we get

$$\mu_2 R_2 - \mu_1 R_1 + p\mu_1[pR_1 + v_\alpha^{\pi_3}(i-1, j) - (p v_\alpha(i-1, j) + q v_\alpha(i-2, j))]$$
$$+ q\mu_1[pR_1 + v_\alpha^{\pi_3}(i-1, j-1) - (p v_\alpha(i-1, j) + q v_\alpha(i-2, j-1))]$$

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

$$\mu_1[\mu_2 R_2 - q\mu_1 R_1 + q\mu_1(v_\alpha^{\pi_3}(i-1, j-1) - p v_\alpha(i-2, j) - q v_\alpha(i-2, j-1))] \qquad (3.27)$$

Adding expressions (3.26) and (3.27) we get the following lower bound

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

The first expression above is (3.22) 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.

Case 4. *Suppose that policy π_1 has the provider work at station 1 and policies $\pi_i(i = 2, 4, 5)$ all have the provider work at station 2. In this case, let π_3 have the provider 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 (3.20) are

$$\mu_1\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_2(pv_\alpha(i-1, j) + qv_\alpha(i-1, j-1) - v_\alpha(i, j) + R_1) + \mu_1[v_\alpha(i, j) - pv_\alpha(i-1, j+1) - qv_\alpha(i-1, j)]\Big]. \tag{3.28}$$

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 (3.20) are

$$\mu_2[\mu_2 R_2 - \mu_1 R_1 - \mu_2 R_2 + \mu_1(v_\alpha(i, j-1) - pv_\alpha(i-1, j) - qv_\alpha(i-1, j-1))]. \tag{3.29}$$

Adding expressions (3.28) and (3.29) yields

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

Note that expression (3.30) is implied by the left hand side of inequality (3.20). That is, we may restart the argument from here.

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

If the first event is a service completion at station 1 in Processes 1 and 5, after which all processes follow an optimal control, then the remaining rewards in (3.20) are

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

If the first event is a service completion at station 2 in Processes 2-4, again after which optimal controls are used, then the remaining rewards from (3.20) are

$$\mu_2\Big[\mu_2 R_2 - \mu_1 R_1 - \mu_2 R_2 + \mu_1\big[v_\alpha(i, j-1) - p v_\alpha(i-1, j) + q R_2 - q v_\alpha(i-1, j)\big]\Big] \quad (3.32)$$

Adding (3.31) and (3.32) we get

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

After rearranging terms, we get

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

The first expression (with coefficient μ_1) is the left side of inequality (3.19). Similarly, so is the second expression (with coefficient $q\mu_1$) except that it is evaluated at $(i-1, j)$. In both cases, we may relabel the states and continue as though we had started in these states.

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

If the first event is a service completion at station 1 in Processes 1, 3, 4, and 5, after which all processes follow an optimal control, then the remaining rewards

in (3.20) are

$$\begin{aligned}\mu_1\Big[&\mu_2R_2-\mu_1R_1+\mu_2(pv_\alpha(i-1,j)+qv_\alpha(i-1,j-1)+R_1-v_\alpha(i,j))\\&+\mu_1(pv_\alpha(i-1,j+1)+qv_\alpha(i-1,j)-p(pv_\alpha(i-2,j+2)+qv_\alpha(i-2,j+1))\\&-q(pv_\alpha(i-2,j+1)+qv_\alpha(i-2,j)))\Big].\end{aligned}\tag{3.33}$$

If the first event is a service completion at station 2 in Process 2, again after which optimal controls are used, then the remaining rewards from (3.20) are

$$\mu_2\Big[\mu_2R_2-\mu_1R_1-\mu_2R_2+\mu_1(v_\alpha(i,j)-pv_\alpha(i-1,j+1)-qv_\alpha(i-1,j))\Big]\tag{3.34}$$

Adding (3.33) and (3.34) we get

$$\begin{aligned}&p\mu_1[\mu_2R_2-\mu_1R_1+\mu_2(v_\alpha(i-1,j)-v_\alpha(i-1,j+1))\\&\quad+\mu_1(v_\alpha(i-1,j+1)-pv_\alpha(i-2,j+2)-qv_\alpha(i-2,j+1))]\\&\quad+q\mu_1[\mu_2R_2-\mu_1R_1+\mu_2(v_\alpha(i-1,j-1)-v_\alpha(i-1,j))\\&\quad+\mu_2(v_\alpha(i-1,j)-pv_\alpha(i-2,j+1)-qv_\alpha(i-2,j))].\end{aligned}\tag{3.35}$$

The expression inside the first pair of brackets in the expression above is the left side of the inequality (3.19) evaluated at $(i-1,j+1)$. Similary, the expression inside the second pair of brackets in the expression above is the left side of inequality (3.19) evaluated at $(i-1,j)$. In both cases, we may relabel the states and continue as though we had started in these states.

Case 7. *Suppose that policies $\pi_i(i=1,4)$ have the provider work at station 1 and that policies $\pi_i(i=2,5)$ have the provider work at station 2. In this case, let policy π_3 have the provider work at station 1.*

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

in (3.20) are

$$\mu_1\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_2(pv_\alpha(i-1,j) + qv_\alpha(i-1,j-1) + R_1 - v_\alpha(i,j))$$
$$+ \mu_1\Big(pv_\alpha(i-1,j+1) + qv_\alpha(i-1,j) + R_1 - p[pv_\alpha(i-2,j+2)$$
$$+ qv_\alpha(i-2,j+1) + R_1] - qv_\alpha(i-1,j)\Big)\Big]. \tag{3.36}$$

If the first event is a service completion at station 2 in Processes 2 and 5, again after which optimal controls are used, then the remaining rewards from (3.20) are

$$\mu_2\Big[\mu_2 R_2 - \mu_1 R_1 - \mu_2 R_2 + \mu_1(v_\alpha(i,j) - pv_\alpha(i-1,j+1) - qv_\alpha(i-1,j-1) - qR_2)\Big] \tag{3.37}$$

Adding (3.36) and (3.37) we get

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

The expression inside the brackets in the expression above is the left side of the inequality (3.19) evaluated at $(i-1, j+1)$. In this case, we may relabel the states and continue as though we had started in these states.

Case 8. *Suppose policy π_2 has the provider work at station 1 and policies $\pi_i (i = 4, 5)$ have the provider work at station 2. In this case, let policies π_1 and π_3 have the provider work at station 1 and 2, respectively.*

If the first event is a service completion at station 1 in Processes 1-2, after which all processes follow an optimal control, then the remaining rewards

in (3.20) are

$$
\begin{aligned}
&\mu_1\Big[\mu_2R_2 - \mu_1R_1 + \mu_2(pv_\alpha(i-1,j) + qv_\alpha(i-1,j-1) - pv_\alpha(i-1,j+1) - qv_\alpha(i-1,j)) \\
&\qquad + \mu_1(v_\alpha(i,j) - pv_\alpha(i-1,j+1) - qv_\alpha(i-1,j))\Big] \\
&\quad = \mu_1\Big[\mu_2R_2 - \mu_1R_1 + \mu_2(pv_\alpha(i-1,j) + qv_\alpha(i-1,j-1) - pv_\alpha(i-1,j+1) - qv_\alpha(i-1,j)) \\
&\qquad + \mu_1(v_\alpha(i,j) - pv_\alpha(i-1,j+1) - qv_\alpha(i-1,j)) \\
&\qquad + \mu_2(v_\alpha(i,j-1) - v_\alpha(i,j)) - \mu_2(v_\alpha(i,j-1) - v_\alpha(i,j))\Big]. \qquad (3.38)
\end{aligned}
$$

If the first event is a service completion at station 2 in Processes 3-5, again after which optimal controls are used, then the remaining rewards from (3.20) are

$$
\begin{aligned}
&\mu_2\Big[\mu_2R_2 - \mu_1R_1 + \mu_2(v_\alpha(i,j-1) - v_\alpha(i,j)) \\
&\qquad + \mu_1(v_\alpha(i,j-1) - pv_\alpha(i-1,j) - qv_\alpha(i-1,j-1))\Big] \\
&\quad = \mu_2\Big[\mu_2R_2 - \mu_1R_1 + \mu_2(v_\alpha(i,j-1) - v_\alpha(i,j)) \\
&\qquad + \mu_1(v_\alpha(i,j-1) - pv_\alpha(i-1,j) - qv_\alpha(i-1,j-1)) \\
&\qquad + \mu_1(v_\alpha(i,j) - pv_\alpha(i-1,j+1) - qv_\alpha(i-1,j)) \\
&\qquad - \mu_1(v_\alpha(i,j) - pv_\alpha(i-1,j+1) - qv_\alpha(i-1,j))\Big] \qquad (3.39)
\end{aligned}
$$

Adding (3.38) and (3.39) we get

$$
\begin{aligned}
&\mu_1\Big[\mu_2R_2 - \mu_1R_1 + \mu_2(v_\alpha(i,j-1) - v_\alpha(i,j)) + \mu_1(v_\alpha(i,j) - pv_\alpha(i-1,j+1) - qv_\alpha(i-1,j))\Big] \\
&\quad + \mu_2\Big[\mu_2R_2 - \mu_1R_1 + \mu_2(v_\alpha(i,j-1) - v_\alpha(i,j)) + \mu_1(v_\alpha(i,j) - pv_\alpha(i-1,j+1) - qv_\alpha(i-1,j))\Big].
\end{aligned}
$$

The expression inside the first and second pair of brackets in the expression above is the left side of the inequality (3.19). In both cases, we may relabel the states and continue as though we had started in these states.

Case 9. *Suppose that policies π_2 and π_5 have the provider work at station 1 and that policy π_4 has the provider work at station 2. In this case, let policies π_1 and π_3 have the provider work at station 1.*

If the first event is a service completion at station 1 in Processes 1,2,3, and 5, after which all processes follow an optimal control, then the remaining rewards in (3.20) are

$$\begin{aligned}
\mu_1\Big[\mu_2 R_2 - \mu_1 R_1 + \mu_2\Big(&pv_\alpha(i-1,j) + qv_\alpha(i-1,j-1) \\
&- pv_\alpha(i-1,j+1) - qv_\alpha(i-1,j)\Big) \\
&+ \mu_1(pv_\alpha(i-1,j+1) + qv_\alpha(i-1,j) - pv_\alpha(i-1,j+1) \\
&- q[pv_\alpha(i-2,j+1) + qv_\alpha(i-2,j)] + pR_1)\Big]
\end{aligned} \tag{3.40}$$

If the first event is a service completion at station 2 in Process 4, again after which optimal controls are used, then the remaining rewards from (3.20) are

$$\begin{aligned}
\mu_2\Big[&\mu_2 R_2 - \mu_1 R_1 + \mu_2(v_\alpha(i,j-1) - v_\alpha(i,j)) \\
&+ \mu_1(v_\alpha(i,j) - pv_\alpha(i-1,j) - pR_2 - qv_\alpha(i-1,j))\Big]
\end{aligned} \tag{3.41}$$

Adding (3.40) and (3.41) we get

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

The expression inside the first pair of brackets in the expression above is the left side of the inequality (3.19) evaluated at $(i-1, j)$. The expression inside the second pair of brackets in the expression above is the left side of inequality (3.19). In both cases, we may relabel the states and continue as though we had started in these states.

Case 10. *Suppose that policies π_2 and π_4 have the provider work at station 1 and that policy π_5 has the provider work at station 2. In this case, let policies π_1 and π_3 have the provider work at station 1.*

If the first event is a service completion at station 1 in Processes 1-4, after which all processes follow an optimal control, then the remaining rewards in (3.20) are

$$\begin{aligned}
&\mu_1\Big[\mu_2R_2-\mu_1R_1+\mu_2(pv_\alpha(i-1,j)+qv_\alpha(i-1,j-1)-pv_\alpha(i-1,j+1)-qv_\alpha(i-1,j))\\
&\quad+\mu_1(pv_\alpha(i-1,j+1)+qv_\alpha(i-1,j)-p(pv_\alpha(i-2,j+2)+qv_\alpha(i-2,j+1))\\
&\quad-qv_\alpha(i-1,j)+qR_1)\Big]. \qquad (3.42)
\end{aligned}$$

If the first event is a service completion at station 2 in Process 5, again after which optimal controls are used, then the remaining rewards from (3.20) are

$$\begin{aligned}
&\mu_2\Big[\mu_2R_2-\mu_1R_1+\mu_2(v_\alpha(i,j-1)-v_\alpha(i,j))\\
&\qquad+\mu_1(v_\alpha(i,j)-pv_\alpha(i-1,j+1)-qv_\alpha(i-1,j-1)-qR_2)\Big]\\
&=\mu_2\Big[\mu_2R_2-\mu_1R_1+\mu_2(v_\alpha(i,j-1)-v_\alpha(i,j))\\
&\qquad+\mu_1(v_\alpha(i,j)-pv_\alpha(i-1,j+1)-qv_\alpha(i-1,j-1)-qR_2)\\
&\qquad+\mu_1(v_\alpha(i,j)-pv_\alpha(i-1,j+1)-qv_\alpha(i-1,j))\\
&\qquad-\mu_1(v_\alpha(i,j)-pv_\alpha(i-1,j+1)-qv_\alpha(i-1,j))\Big] \qquad (3.43)
\end{aligned}$$

Adding (3.42) and (3.43) we get

$$\begin{aligned}
&p\mu_1\Big[\mu_2R_2-\mu_1R_1+\mu_2(v_\alpha(i-1,j)-v_\alpha(i-1,j+1))\\
&\qquad+\mu_1(v_\alpha(i-1,j+1)-pv_\alpha(i-2,j+2)-qv_\alpha(i-2,j+1))\Big]\\
&\qquad+\mu_2\Big[\mu_2R_2-\mu_1R_1+\mu_2(v_\alpha(i,j-1)-v_\alpha(i,j))\\
&\qquad+\mu_1(v_\alpha(i,j)-pv_\alpha(i-1,j+1)-qv_\alpha(i-1,j))\Big].
\end{aligned}$$

The expression inside the first pair of brackets in the expression above is the left side of the inequality (3.19) evaluated at $(i-1, j+1)$. The expression inside the second pair of brackets in the expression above is the left side of inequality (3.19). In both cases, we may relabel the states and continue as though we had started in these states.

Case 11. *Assume that each policy* $\{\pi_1, \pi_2, \ldots, \pi_5\}$ *serves at station 1.*

If all processes see a service completion (at station 1), the remaining rewards in (3.20) are

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

The term with coefficient p (q) is again the left hand side of (3.19) evaluated at $(i-1, j+1)$ ($(i-1, j)$). Relabel the initial states and continue.

In every case save one 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. ∎

3.4.1.2 Proof in the average reward case

3.4.1.2.1 Proof. Using the same sample path argument provided in the discounted cost case, positive recurrence implies that the processes will reach case 3 for some i, j with probability 1.

The subcases following case 3 show that at the next step, either the relative position of the states starting from case 3 are maintained, but with i replaced by $i-1$, or, we move to subcase 3.1 for some non-negative integers i, j. Since the chain is positive recurrent, it follows that eventually either the processes move to case 3 with $i = 1$, or, they move to subcase 3.1 with $i = 1$. In both cases, let policy π_1 have the provider work at station 1.

In the first case, the remaining rewards of the processes after service completions are at least

$$\mu_2[v_\alpha(1, j-1) - (pv_\alpha(0, j-1) + qv_\alpha(0, j-2))].$$

The last expression is non-negative since the value function is monotone increasing in each coordinate.

In the second case, the remaining rewards of the processes after service completions is at least

$$\mu_2[v_\alpha(1, j-1) - v_\alpha(0, j-2)] + \mu_1[pv_\alpha(i-1, 1) + qv_\alpha(i-1, 0) - v_\alpha(i-1, 0)].$$

The last expression is also non-negative. It follows from this analysis that, in the total reward case, it is optimal to serve at station 2 whenever station 2 is not empty and whenever the controlled continuous-time Markov chain is positive recurrent. ∎

3.5 Numerical Study

In this section, we study numerically the control policies discussed thus far and compare them to other potential policies in practice. To do so, we recall that the Lutheran Medical Center (LMC) located in Brooklyn is a 462-bed academic teaching hospital. For the LMC's emergency room, it is reported that the average service time for phase 1 (triage) is approximately 7 minutes and that the average phase 2 (treatment) time is approximately 13 minutes. As a result, we fix exponential service times with rates $\mu_1 = \frac{60}{7} \approx 8.57$ per hour and $\mu_2 = \frac{60}{13} \approx 4.62$ per hour. We assume that the system accrues a reward of 20 after a service completion at station 2 (i.e. $R_2 = 20$), and that patients join the second station immediately after completing service at station 1 (i.e. that $p = 1$). We compare policies for two scenarios. In the first scenario $R_1 = 10$ so that $\mu_1 R_1 < \mu_2 R_2$. In the second scenario, $R_1 = 15$ so that $\mu_1 R_1 > \mu_2 R_2$.

Hospitals do not generally track the time that a patient spends waiting for treatment before leaving the system. As a result, we do not have direct estimates for β. We propose the following for estimating a range for β and then parameterize β over that range. Assuming that abandonment time strongly correlates with service time, let $\hat{\theta} = \frac{\mu}{\hat{\beta}}$, where $\hat{\beta}$ is our estimate of the abandonment rate. This is the average patience measured in units of average service time. We use $\hat{\theta}$ to estimate the abandonment rate. That is to say, vary $\hat{\theta}$ over a sensible range of values, and for each value of $\hat{\theta}$ in this range, determine the corresponding value of $\hat{\beta}$. For this study, we assume that $\hat{\theta} \in (5, 30)$. This implies that the mean abandonment (actually patience) time is from 5 to 30 times that of the mean service time (i.e. about one to 6.5 hours). This yields that the range for the abandonment rate β is approximately $(0.15, 0.92)$.

To justify the range we have proposed for θ, assume there is a linear relationship between the fraction of abandonments, $\mathbb{P}(\text{Ab})$, and average waiting time, $\mathbb{E}[W]$, (see for example **[?]** where it is shown that this relatonship holds for the Erlang A model):

$$\mathbb{P}(\text{Ab}) = \beta\,\mathbb{E}[W]. \tag{3.44}$$

The relation (3.44) between the average wait in queue, $\mathbb{E}[W]$, and the fraction of patients who left without being seen, $\mathbb{P}(\text{LWBS})$, provides a method of estimating

β as follows

$$\hat{\beta} = \frac{\mathbb{P}(\text{LWBS})}{\mathbb{E}[W]} = \frac{\%\text{LWBS}}{\text{Average wait}}. \tag{3.45}$$

Using the data in Table 1 of [**?**], we use (3.45) to obtain estimates for the abandonment rate $\hat{\beta}$ (see Table 3.1). The Emergency Severity Index (ESI), which is used in triage, has 5 levels of severity with levels 1 and 2 signifying patients requiring prompt attention and in need of hospitalization. The TTR is designed for patients who will not need to be hospitalized and generally only includes those in levels 3-5. We have extended this to include level 2 to expand the range of consideration for our study. If we compute the corresponding $\hat{\theta}$ estimates we get a range for $\hat{\theta}$ that is approximately (5,20). We extend this to (5,30) (which leads to (.29,.92)) which fall in the range of $(0.15, 0.92)$; a proper subset of the range for β we have proposed. In other words, we have proposed a range of values for θ that includes the range of values for $\hat{\beta}$ we have estimated using the data from [**?**] and an estimation procedure based on the Erlang-A model [**?**]. We consider a slightly wider interval to account for the possibility of a smaller hospital than that considered in [**?**].

	ESI 2	ESI 3	ESI 4	ESI 5
Population	27,538	65,773	39,878	10,509
%LWBS	1.70%	9.50%	4.70%	7.40%
Average wait (hr.)	1.0	1.9	1.3	1.3
Service time (hr.)	3.7	4.0	1.8	1.2
$\hat{\beta}$	0.02	0.05	0.04	0.06

Table 3.1: Estimation on abandonment rate

The arrival rate varies tremendously, mostly depending on the time of day. It has been reported that the arrival rate can range anywhere from 0.5 to 9.0 per hour. We assume that the arrival process is Poisson with rate λ, and consider $\lambda \in \{0.5, 1.5, 3, 4.5, 6.5, 8.5\}$. Let $S(\mu_1, \mu_2, \beta) := \frac{1}{\mu_1} + \frac{1}{\mu_2+\beta}$ and recall in Section 3.4, we showed that $\lambda < \frac{1}{S(\mu_1,\mu_2,\beta)}$ implies that the provider should prioritize station 2 with the intuition being that it is the policy that minimizes the number of abandonments. Since the policy that prioritizes station 1 leads to a stable system whenever $\lambda < \mu_1$, while the policy that prioritizes station 2, is not stable if $\lambda \geq \frac{1}{S(\mu_1,\mu_2,\beta)}$ we also study the region $\frac{1}{S(\mu_1,\mu_2,\beta)} \leq \lambda < \mu_1$. Note that if $\lambda \geq 4.5$, then the system under the policy that prioritizes station 2 is unstable.

The policy that prioritizes station 2 has the potential disadvantage that in stationarity it leaves station 1 to work at station 2 after **every** station 1 service. In other words, it is the policy that spends the least amount of time working at station 1 (on average) out of all of the policies considered here. For example, since triage is supposed to take place rather quickly to safeguard getting timely care to those who have truly emergent problems, it may be desirable to minimize the number of patients awaiting triage. As a result, using a threshold policy, for example, may be a sensible alternative to prioritizing station 2.

In what follows, we use a discrete event simulation study to compare the average reward of several policies for different values of λ and β, where λ approaches μ_1 and β ranges from values contained in the interval specified above (see Table 3.2 for a complete list of the parameter values). In each scenario for R_1, the simulation length is one year (24 hours per day) per replication with 30 replications (see B for Tables and corresponding confidence intervals). We compare the following policies, which are described in detail in Section 3.2:

1. Prioritize station 1 (P1);
2. Prioritize station 2 (P2);

3. Exhaustive policy (E);

4. Threshold policy (K) having a threshold values, which we denote by K, of $2, 5, 10, 15$, and 20.

Parameter Symbol	Value(s)
μ_1	8.57
μ_2	4.62
β	$0.15, 0.3, 0.5, 0.8$
R_2	20
p	1
R_1	$10, 15$
λ	$0.5, 1.5, 3, 4.5, 6.5, 8.5$

Table 3.2: List of Parameters and their values

3.5.1 Summary of Results from Numerical Study

We have assumed throughout that maximizing the average reward is our primary concern, but that some providers may prefer to forego some average reward in favor of having the provider prioritize station 1 more often. One reason for this is that minimizing the number of patients awaiting triage may be desirable given that triage is supposed to take place rather quickly to safeguard getting timely care to those who have truly emergent problems. Moreover, it seems that the model may have higher average rewards, but larger average total number in the system; and thus, longer wait times. This trade-off is also examined. The basic insights obtained from the numerical study follow with details in the following sections.

- It has already been shown that as long as the policy **P2** yields a stable Markov process, the average reward is maximized. The numerical study confirms this finding, but shows that when the system is **lightly loaded** there may not be significant differences between the optimal policy and several others. When the system is **highly loaded** the same trade-off can be made, but not without significant loss of average reward.

- When **P2** does not yield a stable Markov chain, the average waiting times are unbounded making this policy impractical. However, the average reward of the other candidate policies are not significantly different. A provider might look to the policy with the lowest average total number in the system. This policy is the policy that prioritizes station 1.

3.5.2 Average Reward for $\lambda < \frac{1}{S(\mu_1,\mu_2,\beta)}$

In this case, the policy that prioritizes station 2 is stable and thus, as was shown in Section 3.4, is also average reward optimal. As a result, we set the average reward of the "prioritize station 2" policy as a baseline (average reward displayed) and display the percent of this baseline for the remaining policies in Table B.1 for the case when $R_1 = 10$ and Table B.2 for the case when $R_2 = 15$.

3.5.2.1 Subcase I: ($R_1 = 10$)

Results are summarized in Table B.1. As is somewhat expected, the further we get from the optimal policy (by decreasing the threshold), the worse the average reward becomes. In **all** instances, the policy that prioritizes station 1 (**P1**), which corresponds to $K = 1$ performed the worst. The other threshold policies get progressively better as the threshold increases. The exhaustive policy (**E**) sits between the threshold policies with $K = 2$ and 5.

For each value of β tested, lightly loaded systems (with $\lambda \in \{0.5, 1\}$), yields that all policies are comparable to policy **P2**; all of the policies are within 5% of the optimal. In general, for a given value of β, all policies tend to get further from optimal as the load (λ) increases. For example, when $\beta = 0.15$, the system under policy **T (**$K = 5$**)** is 0.25% away from optimality when $\lambda = 0.5$, but it is 8.99% from optimality when $\lambda = 3$. Similarly, when $\beta = 0.15$ and $\lambda = 3$, the system under policies **P1**, **T (**$K = 2$**)**, **E**, are 11.98%, 11.08%, and 8.82%, away from optimal, respectively. This is not surprising since all policies prioritizing station 2 emphasize the need to minimize abandonments; higher loads would expect more abandonments. The system under policy **T (**$K = 20$**)** is only 0.34% away from optimality when $\lambda = 0.5$ and is only 3.62% away from optimality when $\lambda = 3$. In short, the higher threshold appears to alleviate most of the loss of optimality.

The system under policies **T(**$K = 10$**)**, **T (**$K = 15$**)**, and **T (**$K = 20$**)** perform relatively well for all reported values of λ and β. No instance was more than 6% away from optimal. Furthermore, in 20 out of the total 24 instances the system under policies **T (**$K = 10$**)**, **T(**$K = 15$**)**, and **T(**$K = 20$**)** were strictly within 2% of the optimal value, and in the worst case scenario, when $\beta = 0.8$ and $\lambda = 3$, the system under policy **T(**$K = 20$**)** was within 3% of the optimal. Lastly, as is to be expected, the average reward of the system under threshold policy improves as K increases so that, for example, the system under policy **T (**$K = 20$**)** is the closest to optimal.

3.5.2.2 Subcase II: ($R_1 = 15$)

Results are summarized in Table B.2. Exactly the same observations that were made for the data summarized in Table B.1 apply in this case as well.

3.5.3 Average Reward for $\lambda \geq \frac{1}{S(\mu_1,\mu_2,\beta)}$

In this case the system under the policy that prioritizes station 2 is unstable while that which prioritizes station 1 is stable. Thus, while the system under policy **P2** is overloaded when $\lambda = 4.5$, **P1** is lightly loaded. As λ increases to 8.5, the system under the policy that prioritizes station 1 becomes highly loaded. For each instance we determine the average reward for each individual policy. This data is displayed in Table B.3 in the case when $R_1 = 10$ and in Table B.4 in the case when $R_1 = 15$.

3.5.3.1 Subcase I: ($R_1 = 10$)

The average reward for each policy is displayed in Table B.3. When the system using **P2** is slightly overloaded ($\lambda = 4.5$), the results are similar to the previous case; threshold policies improve as K increases. The system using the exhaustive policy (**E**) still lies between **T(**$K = 2$**)** and **T (**$K = 5$**)**. It is also the case that when $\lambda = 4.5$, the average reward rates generally tended to decrease as β increased; more abandonments yields less rewards.

For the remaining values of λ, especially when $\lambda \geq 6.5$, the average reward of the system for all of the policies except **P2** are approximately the same which (when combined with the results of Section 3.5.2.1) indicates some concavity as

λ increases (see Figure 3.1). Since the average queue-length in each case except **P2** is finite (we have not pursued stability conditions for threshold policies), this indicates that enough time is spent at station 1 to stabilize the system, but then those customers are eventually allowed to abandon the system adding nothing to the reward.

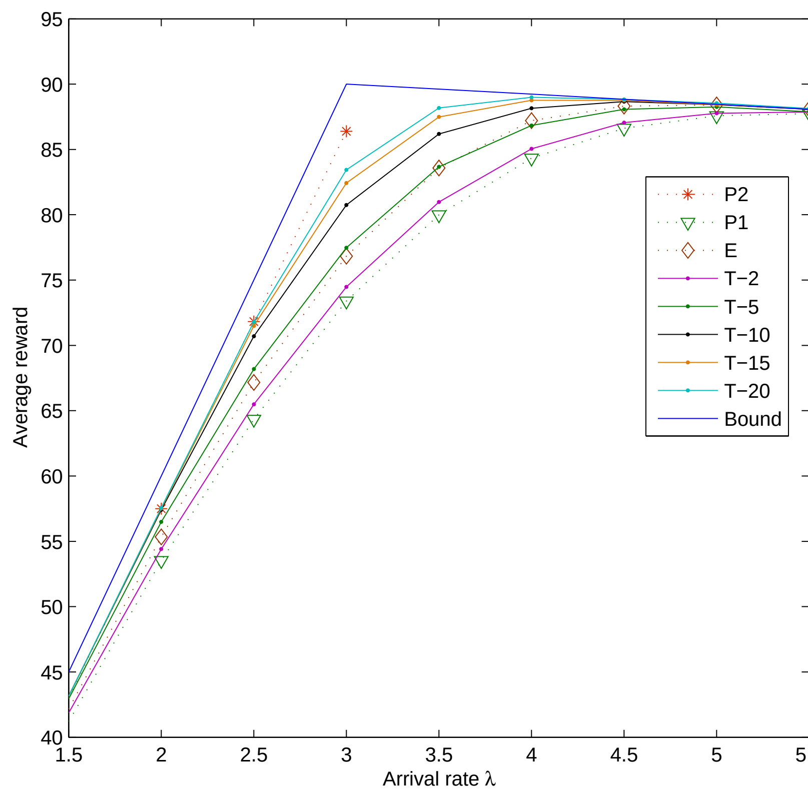

Figure 3.1: Average Reward Concavity in λ ($\beta = 0.3, R_1 = 10$)

3.5.3.2 Subcase II: ($R_1 = 15$)

The average reward for each policy is displayed in Table B.4. Exactly the same observations that were made for the data summarized in Table B.3 apply in this case with one exception. It is no longer the case that the values for the average reward for system using policy **P2** are an upper bound for the values of the average reward of the other policies. In fact, **P2** is the policy with lowest average reward in all of the instances described in this case. Moreover, since the system is unstable, the threshold policies (and the exhaustive policy) are very close to the policy that prioritizes station 1. This adds credibility to the conjecture that if $\mu_1 R_1 \geq \mu_2 R_2$ and $\frac{1}{S(\mu_1,\mu_2,\beta)} \leq \lambda < \mu_1$ that there exists parameters under which it is optimal (or near optimal) to prioritize station 1. This is left as future study.

3.5.4 Practical Considerations

If $\lambda < \frac{1}{S(\mu_1,\mu_2,\beta)}$, then the policy that prioritizes station 2 is both stable and optimal. This is ideal from the patient's perspective, since once diagnosed, the patient immediately proceeds to phase-2 service. This is also the policy, most likely to be implemented in a TTR program. On the other hand, it may be desirable to minimize the number of patients awaiting triage. This is because triage is supposed to take place rather quickly to safeguard getting timely care to those who have truly emergent problems. In this case, prioritizing station 1 to some extent, by using a threshold policy, say, may make more sense than using a policy that prioritizes station 2. Moreover, prioritize 2 leads to an unstable system whenever $\lambda \geq \frac{1}{S(\mu_1,\mu_2,\beta)}$. In this case, if waiting times are of concern, alternative service disciplines must be used. In this section, we study numerically two important practical measures of performance: average total number in system and the average waiting time in station 1.

3.5.4.1 Average total number in system ($\lambda < \frac{1}{S(\mu_1,\mu_2,\beta)}$)

Note from the results shown in Table B.5 that, as expected, for each value of β, the average total number in system for each policy is non-decreasing in λ while, for each value of λ, the average total number in system is non-increasing in β.

There are not significant differences in the average total number in system when the systems are lightly loaded, but when $\lambda = 3$ the system that prioritizes station 2 is highly loaded and thus has long wait times. The other systems suffer from this issue in a non-decreasing fashion. That is, the average total number in system for the threshold policies increases as K increases (when $\lambda = 3$).

3.5.4.2 Average total number in system ($\lambda \geq \frac{1}{S(\mu_1,\mu_2,\beta)}$)

As expected, for each value of β, the average total number in system for each policy increases as λ increases and that for each value of λ, and the average total number in system decreases whenever β increases. See Table B.6. Of course, since **P2** is unstable the average total number in system is largest for this policy. What is interesting is the policy that is closest to **P2** (with threshold level K = 20) has total number in systems that are not far from **P1** when **P1** is highly loaded, but can be twice or sometimes almost three times the total number in system when **P1** is lightly loaded.

3.5.4.3 Average number of patients in station 1 and in station 2:

Results are summarized in Tables B.7 and B.8. Note from the results shown in Table B.7 that, as expected, the average number in station 1 increases as the threshold increases (with **P1** corresponding to $K = 1$ and **P2** corresponding to $K = \infty$). This is because service at station 1 is further neglected as the threshold increases. Also as expected, is that the average number in station 1 increases in λ and is relatively unaffected by the abandonment rate, β, and that the average number of patients at station 2 decreases in β and increases in λ.

Two features of the data stand out. First, is that when policy **P2** is highly loaded (but still stable) with $\lambda = 3$, $\beta = .15$, the average number of patients in station 1 is at least 33, but if we use policy **T (**$K = 20$**)**, that number drops to less than 7. This savings is at the cost of a single person increase at station 2 (see Table B.8).

The second feauture that stands out is that when $\lambda \geq 3$, the average number of patients in station 1 for policy **P1** is significantly less than than that of policy **T (**$K = 5$**)**, and hence, policies **T (**$K = 10$**)**, **T(**$K = 15$**)**, and **T (**$K = 20$**)**. For example, when $\lambda = 3, \beta = 0.8$, the average number of patients in station 1 is 0.54 whereas the average number of patients in station 1 under policy **T (**$K = 5$**)** is more than twice that, 1.48. On the other hand, when $\lambda \geq 3$, the average number of patients in station 2 is relatively the same for all of the policies, except **P2**. An immediate implication from this is that if it is of intererest to prioritize station 1 (for the reasons alluded to above), then **P1** (or **T (**$K = 5$**)**) are good practical candidates, since the average number of patients in station 1 is relatively low, while the average number of patients in station 2 is relatively the same as that of the other policies, with the exception of **P2**.

Finally, it definitely makes sense that the queues are large with lambda = 8.5. If we use **P1**, which spends the most time at station 1, we still get a highly loaded but stable system since $\mu_1 = 8.57$. Time away from station 1 via the threshold policies make station 1 even more highly loaded.

3.5.4.4 Average sojourn time at station 1:

The average sojourn time for each policy is displayed in Table B.9. Exactly the same observations that were made for the data summarized in Table B.7 apply in this case.

3.6 Chapter Summary

In this chapter, we consider the question of how to facilitate the delivery of care by a single medical service provider in a two-phase stochastic service system. This is motivated by patient care in health care service systems like the triage and treatment process in the ER, and in particular, by the TTR Program at the LMC in Brooklyn, NY. We model the triage and treatment operation as a single server two-stage tandem queueing system. We then formulate a CTMDP for this queueing model to determine service disciplines that maximize discounted expected reward and long run average reward. In contrast to existing models, ours has unbounded transitions rates due to potential abandonments by patients waiting for or in treatment (phase-two service). This means, in particular, that the model cannot be analyzed by first uniformizing the process. We use the continuous time dynamic programming optimality equations and sample path arguments to determine the optimal service discipline under each criterion. Our main result is that we provide sufficient conditions under which prioritizing phase-two treatment is optimal, which as we have already pointed out, is beneficial from both the patient's and provider's perspective. Moreover, this is the service discipline that providers at the LMC TTR program currently use. In the discounted reward case, the condition is a c-μ-like rule while in the long run average reward case, prioritizing station 2 is optimal regardless of the reward structure as long as this policy yields a stable Markov process. Finally, we give a condition for when to prioritize station 1.

We develop a new class of service policies, which we call K-level threshold policies as reasonable alternatives to the two priority disciplines. In particular, a 1 (∞)-level threshold policy is the same as prioritizing phase-one (two). Using data reported from the TTR program at the LMC, we compare several service disciplines in a simulation study. Results show that when prioritizing phase-two service is stable and when the system is lightly loaded there are not significant differences between the optimal policy and several others, but when the system is highly loaded alternative disciplines can lead to significant loss of average reward. When prioritizing phase-two service yields an unstable system, the average reward of the other candidate policies are not significantly different and a provider might consider policies with the lowest average total number in the system.

CHAPTER 4
CONCLUSION

In this thesis, we have proposed frameworks for allocating healthcare resources in two different settings. The first is disaster response, where we proposed a framework for allocating EMS vehicles in response to large-scale events. The second is patient flow in the ER, where we address how to prioritize the work of medical service providers in a two-phase stochastic service system, motivated in part by the LMC's TTR program. Our broad contribution was to use Operations Research concepts and tools to develop these frameworks. In particular, we modeled each operation as a multi-server queueing system. For the EMS problem, we used classical results from the queueing theory literature to determine the number of vehicles that can be made available to the affected region from the unaffected one. In addition, we used Markov decision processes to determine how EMS vehicles can be constantly re-allocated within the affected region to help bring the workload back to its normal day-to-day levels. Lastly, we used a knapsack model to determine how many vehicles will be allocated to different areas in the affected region. For the ER triage and treatment process, we also developed a Markov decision process formulation and then used this formulation in combination with sample-path arguments to determine the optimal dynamic schedule for medical providers for several different scenarios. In both problems, we compared policies obtained from the MDP formulation with other heuristics in a corresponding numerical study.

There are several avenues for further research. In the EMS problem, we have presented but one suggested heuristic, which works well under the two criteria studied. To be certain, there might be others that would serve well under this and other criteria. Another possibility is to extend the current study to analyze control policies when there are costs associated with moving vehicles from one locale to another. This could involve analyzing the structure of optimal control policies using dynamic programming techniques, sample path methods, and/or simulation. One other possibility is to use Approximate Dynamic Programming frameworks to develop good control policies for the the general *L-station Clearing System Model* and comparing these with the buddy system heuristics we have proposed and other relevant heuristics in a numerical study. Lastly, since in practice EMS organizations are unlikely to reallocate vehicles between different counties after every call completion, another approach is to use fluid models instead of dynamic programming make these decisions. Fluid models would likely lead to simple (one-time) but highly competitive and useful allocation rules that can be constructed in higher dimensions with relatively little trouble. At the very least, we hope that the research presented in Chapter 2 can serve as a further catalyst to adopt systematic methods for resource allocation in the face of large scale emergency events.

In the triage and treatment problem, one extension would be to replace the homogeneous Poisson process with a non-homogeneous one or with a Markov Modulated Poisson Process. Arrivals to many Emergency Departments are, in fact, time-dependent. Another consideration is to analyze control policies assuming that service is non-preemptive so that once a provider begins triage (treatment), (s)he must remain in that phase until service is completed before (s)he can re-allocated. Moreover, the health condition for ailing patients may deteriorate the longer they wait for treatment. As a result, the exponential assumption may be inappropriate and one might consider the same model but with generally distributed service times. Lastly, a simulation study that analyzes additional service disciplines, like K-limited, gated, and K-decrementing policies may also be of interest.

While the aforementioned work is directly related to the problems discussed in Chapters 2 and 3, we are also pursuing other research projects related to healthcare. Here we highlight one, related to patient flow in the ED. In particular, we consider how should health care operations managers control the admission of patients with different needs, or severity levels, into a health care service facility with limited resources such as beds, examination rooms, or medical equipment. A specific example is the ED where admission control can be done by either dispatching certain patients after triage or by re-directing incoming ambulance arrivals to nearby hospitals. The latter phenomena is known as ambulance diversion, and if properly executed, has the potential for reducing waiting times in the ED by re-routing patients from crowded hospitals to less crowded ones. We have modeled the admission control problem using a Markov decision process formulation. As we saw in Chapter 3, patients in health care service systems, such as the ED, leave the facility without receiving medical care. Furthermore, arrivals to many health care service systems, like the ED, are time-dependent. As a result, we have assumed that the arrival process is a non-homogeneous Poisson process while allowing patients to leave without being seen. Both are key features that have not been adequately addressed in the literature. The goal is to develop and analyze admission control policies using dynamic programming, sample-path methods, and simulation that have the potential to improve patient flow.

APPENDIX A

CHAPTER 2 APPENDIX

A.1 Dictionary of Symbols and Variables

Parameter Symbol	Definition
N	number of cities
λ_k, μ_k, n_k	arrival rate, service rate, and capacity at city k
$L_q^{n_k}$	Long-run average number of jobs waiting to be assigned
L^{n_k}	Average number of jobs in system
$W_q^{n_k}$	Average wait before being assigned a vehicle
W_{n_k}	Average wait until service is completed
L	Number of cities in affected region
$\vec{l}$	Vector of available vehicles
$\vec{m}$	Initial vector of calls above normalcy
$\vec{v}$	Vector of current vehicle configuration
$v_{N_k}(\vec{m}, \vec{v})$	Total cost of returning system to normalcy
h_k	Holding cost per unit time at city k
$c(\cdot)$	Cost function for moving one vehicle
N_D	Number of vehicles available for re-allocation
M	Total number of vehicles available for assignment
$A(M)$	Set of available assignments when total number of vehicles available for assignment is M
B	Number of cities in city group

Table A.1: Dictionary of Symbols and Variables

APPENDIX B

CHAPTER 3 APPENDIX

β	λ	P2	P1	E	T (K=2)	T (K=5)	T (K=10)	T (K=15)	T (K=20)
0.15	0.5	15 (0)	98% (0)	99% (0)	99% (0)	100% (0)	100% (0)	100% (0)	100% (0)
	1.5	44 (0)	98% (0)	99% (0)	98% (0)	100% (0)	100% (0)	100% (0)	100% (0)
	3	88 (0)	88% (0)	91% (0)	89% (0)	91% (0)	94% (0)	95% (0)	96% (0)
0.3	0.5	14 (0.083)	99% (0.083)	100% (0.076)	100% (0.079)	100% (0.11)	101% (0.082)	100% (0.082)	100% (0.082)
	1.5	43 (0.16)	96% (0.16)	98% (0.14)	97% (0.15)	100% (0.14)	100% (0.16)	100% (0.16)	100% (0.16)
	3	86 (0.18)	85% (0.18)	89% (0.15)	86% (0.14)	90% (0.16)	93% (0.16)	95% (0.13)	97% (0.13)
0.5	0.5	14 (0.082)	99% (0.082)	100% (0.1)	99% (0.08)	100% (0.067)	100% (0.09)	100% (0.079)	101% (0.088)
	1.5	42 (0.15)	94% (0.15)	97% (0.14)	96% (0.14)	99% (0.16)	100% (0.13)	100% (0.13)	100% (0.13)
	3	84 (0.26)	83% (0.26)	87% (0.13)	84% (0.13)	89% (0.12)	94% (0.14)	96% (0.14)	97% (0.17)
0.8	0.5	13 (0.071)	98% (0.071)	99% (0.093)	99% (0.077)	100% (0.071)	100% (0.098)	100% (0.09)	100% (0.084)
	1.5	41 (0.15)	92% (0.15)	96% (0.11)	94% (0.1)	99% (0.14)	100% (0.14)	100% (0.11)	100% (0.14)
	3	81 (0.23)	81% (0.23)	86% (0.14)	83% (0.13)	89% (0.16)	94% (0.17)	97% (0.16)	98% (0.19)

Table B.1: $R_1 = 10$, $\lambda < \frac{1}{S(\mu_1,\mu_2,\beta)}$: Percent of the baseline reward with **Prioritize 2** as baseline (95% Confidence Interval – note numbers above 100% are statistically indistinguishable from baseline).

β	λ	P2	P1	E	T (K=2)	T (K=5)	T (K=10)	T (K=15)	T (K=20)
0.15	0.5	17 (0.074)	99% (0.074)	99% (0.07)	99% (0.084)	100% (0.066)	99% (0.069)	99% (0.079)	99% (0.063)
	1.5	51 (0.15)	98% (0.15)	99% (0.12)	99% (0.13)	100% (0.14)	100% (0.12)	100% (0.12)	100% (0.12)
	3	100 (0.22)	90% (0.22)	93% (0.13)	91% (0.13)	93% (0.15)	95% (0.16)	96% (0.15)	97% (0.14)
0.3	0.5	17 (0.1)	98% (0.1)	99% (0.071)	99% (0.083)	100% (0.11)	99% (0.075)	99% (0.079)	99% (0.095)
	1.5	51 (0.15)	96% (0.15)	98% (0.17)	97% (0.11)	100% (0.18)	100% (0.18)	100% (0.17)	100% (0.21)
	3	100 (0.26)	87% (0.26)	90% (0.14)	88% (0.17)	91% (0.14)	94% (0.18)	96% (0.18)	97% (0.16)
0.5	0.5	16 (0.088)	98% (0.088)	99% (0.081)	100% (0.1)	100% (0.099)	100% (0.096)	100% (0.096)	101% (0.083)
	1.5	50 (0.17)	95% (0.17)	98% (0.17)	96% (0.15)	99% (0.17)	100% (0.14)	100% (0.14)	100% (0.16)
	3	99 (0.23)	85% (0.23)	89% (0.16)	87% (0.17)	91% (0.16)	95% (0.15)	97% (0.17)	98% (0.21)
0.8	0.5	16 (0.099)	98% (0.099)	100% (0.094)	100% (0.081)	100% (0.11)	100% (0.087)	100% (0.094)	100% (0.088)
	1.5	48 (0.17)	94% (0.17)	97% (0.15)	95% (0.13)	99% (0.17)	100% (0.15)	100% (0.16)	100% (0.14)
	3	96 (0.22)	84% (0.22)	88% (0.17)	86% (0.18)	91% (0.23)	95% (0.19)	97% (0.17)	99% (0.17)

Table B.2: $R_1 = 15$, $\lambda < \frac{1}{S(\mu_1,\mu_2,\beta)}$. Percent of the baseline reward with **Prioritize 2** as baseline (95% Confidence Interval – note numbers above 100% are statistically indistinguishable from baseline).

β	λ	P2	P1	E	T (K=2)	T (K=5)	T (K=10)	T (K=15)	T (K=20)
0.15	4.5	90 (0)	88 (0)	89 (0)	88 (0)	89 (0)	89 (0)	89 (0)	89 (0)
	6.5	90 (0)	87 (0)	87 (0)	87 (0)	87 (0)	87 (0)	87 (0)	87 (0)
	8.5	90 (0)	86 (0)	86 (0)	86 (0)	86 (0)	86 (0)	86 (0)	86 (0)
0.3	4.5	90 (0.13)	87 (0.13)	88 (0.12)	87 (0.18)	88 (0.14)	89 (0.099)	89 (0.11)	89 (0.15)
	6.5	90 (0.14)	87 (0.14)	87 (0.12)	87 (0.16)	87 (0.15)	87 (0.14)	87 (0.13)	87 (0.15)
	8.5	90 (0.12)	86 (0.12)	86 (0.12)	86 (0.12)	86 (0.1)	86 (0.12)	86 (0.11)	86 (0.1)
0.5	4.5	90 (0.19)	84 (0.19)	87 (0.12)	85 (0.12)	87 (0.12)	88 (0.13)	89 (0.16)	89 (0.17)
	6.5	90 (0.13)	87 (0.13)	87 (0.13)	87 (0.16)	87 (0.13)	87 (0.13)	87 (0.14)	87 (0.14)
	8.5	90 (0.14)	86 (0.14)	86 (0.12)	86 (0.12)	86 (0.11)	86 (0.13)	86 (0.11)	86 (0.12)
0.8	4.5	90 (0.16)	81 (0.16)	85 (0.13)	82 (0.14)	85 (0.11)	87 (0.15)	88 (0.14)	88 (0.16)
	6.5	90 (0.16)	86 (0.16)	87 (0.14)	87 (0.18)	87 (0.15)	87 (0.14)	87 (0.16)	87 (0.15)
	8.5	90 (0.15)	86 (0.15)	86 (0.14)	86 (0.14)	86 (0.13)	86 (0.13)	86 (0.13)	86 (0.14)

Table B.3: Average Reward when $R_1 = 10$, $\lambda \geq \frac{1}{S(\mu_1,\mu_2,\beta)}$ (95% Confidence Interval).

β	λ	P2	P1	E	T (K=2)	T (K=5)	T (K=10)	T (K=15)	T (K=20)
	4.5	105 (0.17)	111 (0.17)	111 (0.17)	111 (0.12)	111 (0.14)	111 (0.13)	111 (0.15)	111 (0.14)
0.15	6.5	105 (0.15)	120 (0.15)	120 (0.13)	120 (0.14)	120 (0.11)	120 (0.12)	120 (0.13)	120 (0.14)
	8.5	105 (0.14)	128 (0.14)	128 (0.12)	128 (0.11)	128 (0.15)	128 (0.13)	128 (0.1)	128 (0.11)
	4.5	105 (0.21)	109 (0.21)	111 (0.15)	110 (0.15)	110 (0.18)	111 (0.18)	111 (0.18)	111 (0.19)
0.3	6.5	106 (0.21)	120 (0.21)	120 (0.11)	120 (0.14)	120 (0.14)	120 (0.13)	120 (0.15)	120 (0.13)
	8.5	105 (0.19)	128 (0.19)	128 (0.16)	128 (0.15)	128 (0.14)	128 (0.14)	128 (0.12)	128 (0.15)
	4.5	106 (0.18)	107 (0.18)	109 (0.14)	107 (0.12)	109 (0.15)	111 (0.13)	111 (0.14)	111 (0.15)
0.5	6.5	106 (0.14)	120 (0.14)	120 (0.15)	120 (0.17)	120 (0.14)	120 (0.17)	120 (0.18)	120 (0.18)
	8.5	106 (0.15)	128 (0.15)	128 (0.16)	128 (0.13)	128 (0.12)	128 (0.16)	128 (0.16)	128 (0.15)
	4.5	106 (0.17)	103 (0.17)	107 (0.14)	105 (0.13)	108 (0.13)	110 (0.17)	110 (0.12)	111 (0.16)
0.8	6.5	106 (0.21)	119 (0.21)	120 (0.15)	119 (0.14)	120 (0.17)	120 (0.14)	120 (0.11)	120 (0.13)
	8.5	106 (0.14)	128 (0.14)	128 (0.17)	128 (0.16)	128 (0.14)	128 (0.17)	128 (0.14)	128 (0.15)

Table B.4: Average Reward when $R_1 = 15$, $\lambda \geq \frac{1}{S(\mu_1,\mu_2,\beta)}$ (95% Confidence Interval).

β	λ	P2	P1	E	T (K=2)	T (K=5)	T (K=10)	T (K=15)	T (K=20)
	0.5	0.19	0.20	0.19	0.19	0.19	0.19	0.19	0.19
0.15	1.5	0.85	0.93	0.89	0.90	0.85	0.85	0.85	0.85
	3	33.76	4.71	5.54	4.78	5.17	6.01	7.22	8.54
	0.5	0.18	0.19	0.18	0.19	0.18	0.18	0.18	0.18
0.3	1.5	0.82	0.84	0.82	0.83	0.82	0.82	0.82	0.81
	3	18.91	3.34	3.96	3.42	3.81	4.80	5.94	7.15
	0.5	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
0.5	1.5	0.78	0.76	0.77	0.75	0.77	0.79	0.79	0.78
	3	11.29	2.59	3.07	2.68	3.10	4.07	5.13	6.11
	0.5	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
0.8	1.5	0.73	0.68	0.70	0.68	0.72	0.73	0.73	0.73
	3	7.29	2.07	2.42	2.16	2.59	3.50	4.41	5.22

Table B.5: Average total number in system $\left(\lambda < \frac{1}{S(\mu_1,\mu_2,\beta)}\right)$.

β	λ	P2	P1	E	T (K=2)	T (K=5)	T (K=10)	T (K=15)	T (K=20)
	4.5	6277.69	16.67	26.40	17.17	18.50	21.05	23.61	26.31
0.15	6.5	14986.89	39.01	61.47	39.39	40.88	43.42	45.95	48.49
	8.5	23811.51	145.27	164.03	165.23	171.74	175.42	177.44	173.15
	4.5	6042.12	9.17	13.81	9.58	10.89	13.25	15.92	18.68
0.3	6.5	14786.41	21.13	32.52	21.67	23.12	25.67	28.12	30.58
	8.5	23569.65	118.80	132.80	119.09	123.54	136.41	138.11	142.08
	4.5	5691.72	6.20	8.90	6.52	7.65	9.97	12.53	15.21
0.5	6.5	14436.71	13.86	20.77	14.42	15.87	18.38	20.88	23.39
	8.5	23190.66	109.69	140.81	105.50	108.58	110.80	116.17	107.02
	4.5	5195.24	4.50	6.16	4.79	5.81	8.00	10.45	12.92
0.8	6.5	13911.75	9.94	14.17	10.42	11.80	14.29	16.73	19.18
	8.5	22603.52	97.23	123.92	115.79	114.16	118.23	114.13	106.48

Table B.6: Average total number in system $\left(\lambda \geq \frac{1}{S(\mu_1,\mu_2,\beta)}\right)$.

β	λ	P2	P1	E	T (K=2)	T (K=5)	T (K=10)	T (K=15)	T (K=20)
0.15	0.5	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)
	1.5	0.5 (0.00)	0.2 (0.00)	0.4 (0.00)	0.3 (0.00)	0.5 (0.00)	0.5 (0.00)	0.5 (0.00)	0.5 (0.00)
	3	33.1 (0.00)	0.5 (0.00)	2.4 (0.00)	0.9 (0.00)	1.9 (0.00)	3.6 (0.00)	5.2 (0.00)	6.8 (0.00)
	4.5	6277 (0.00)	1.1 (0.00)	11.1 (0.00)	1.6 (0.00)	3.1 (0.00)	5.6 (0.00)	8.2 (0.00)	10.9 (0.00)
	6.5	14986 (0.00)	3.2 (0.00)	25.7 (0.00)	3.6 (0.00)	5.1 (0.00)	7.6 (0.00)	10.1 (0.00)	12.6 (0.00)
	8.5	23811 (0.00)	89.1 (0.00)	107.9 (0.00)	109.0 (0.00)	115.5 (0.00)	119.1 (0.00)	121.2 (0.00)	116.9 (0.00)
0.3	0.5	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)
	1.5	0.5 (0.01)	0.2 (0.00)	0.4 (0.00)	0.3 (0.00)	0.5 (0.00)	0.5 (0.01)	0.5 (0.01)	0.5 (0.01)
	3	18.3 (6.20)	0.5 (0.00)	1.7 (0.02)	0.8 (0.01)	1.7 (0.01)	3.2 (0.03)	4.7 (0.05)	6.0 (0.08)
	4.5	6041 (51.07)	1.1 (0.01)	6.1 (0.08)	1.6 (0.01)	3.0 (0.01)	5.5 (0.01)	8.2 (0.02)	11.0 (0.02)
	6.5	14786 (62.43)	3.1 (0.04)	14.5 (0.14)	3.7 (0.04)	5.2 (0.04)	7.7 (0.04)	10.2 (0.06)	12.7 (0.04)
	8.5	23569 (44.67)	90.7 (14.39)	104.6 (8.70)	91.0 (18.66)	95.4 (20.01)	108.3 (19.89)	110.0 (19.61)	113.9 (20.87)
0.5	0.5	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)
	1.5	0.5 (0.00)	0.2 (0.00)	0.3 (0.00)	0.3 (0.00)	0.4 (0.00)	0.5 (0.01)	0.5 (0.01)	0.5 (0.01)
	3	10.7 (1.71)	0.5 (0.00)	1.4 (0.02)	0.8 (0.00)	1.6 (0.01)	2.9 (0.02)	4.2 (0.06)	5.3 (0.07)
	4.5	5691 (56.14)	1.1 (0.01)	4.1 (0.04)	1.5 (0.01)	2.8 (0.01)	5.3 (0.01)	7.9 (0.02)	10.6 (0.02)
	6.5	14436 (58.37)	3.1 (0.03)	10.0 (0.11)	3.6 (0.04)	5.1 (0.05)	7.6 (0.05)	10.1 (0.05)	12.6 (0.06)
	8.5	23190 (69.02)	127.1 (12.21)	123.9 (12.10)	88.6 (13.92)	91.7 (13.22)	93.9 (18.13)	99.3 (20.37)	90.2 (15.26)
0.8	0.5	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)
	1.5	0.5 (0.01)	0.2 (0.00)	0.3 (0.00)	0.3 (0.00)	0.4 (0.00)	0.5 (0.00)	0.5 (0.01)	0.5 (0.01)
	3	6.7 (0.59)	0.5 (0.00)	1.2 (0.01)	0.8 (0.00)	1.5 (0.01)	2.7 (0.02)	3.7 (0.04)	4.6 (0.07)
	4.5	5195 (48.27)	1.1 (0.01)	3.0 (0.03)	1.5 (0.01)	2.7 (0.01)	5.0 (0.01)	7.5 (0.02)	10.0 (0.03)
	6.5	13911 (61.65)	3.1 (0.03)	7.4 (0.06)	3.6 (0.04)	5.0 (0.03)	7.5 (0.04)	10.0 (0.05)	12.4 (0.06)
	8.5	22603 (59.16)	86.7 (23.32)	113.4 (23.35)	105.2 (12.66)	103.6 (17.95)	107.7 (15.13)	103.6 (18.16)	95.9 (13.57)

Table B.7: Average total number at station 1 (95% Confidence Interval).

β	λ	P2	P1	E	T (K=2)	T (K=5)	T (K=10)	T (K=15)	T (K=20)
0.15	0.5	0.11 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	1.5	0.31 (0)	1 (0)	1 (0)	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	3	0.63 (0)	4 (0)	3 (0)	4 (0)	3 (0)	2 (0)	2 (0)	2 (0)
	4.5	0.64 (0)	16 (0)	15 (0)	16 (0)	15 (0)	15 (0)	15 (0)	15 (0)
	6.5	0.64 (0)	36 (0)	36 (0)	36 (0)	36 (0)	36 (0)	36 (0)	36 (0)
	8.5	0.64 (0)	56 (0)	56 (0)	56 (0)	56 (0)	56 (0)	56 (0)	56 (0)
0.3	0.5	0.1 (0.00085)	0 (0.0011)	0 (0.0013)	0 (0.001)	0 (0.00087)	0 (0.00085)	0 (0.00085)	0 (0.00085)
	1.5	0.31 (0.0015)	1 (0.0068)	0 (0.0046)	1 (0.005)	0 (0.0039)	0 (0.0017)	0 (0.0015)	0 (0.0016)
	3	0.61 (0.0022)	3 (0.049)	2 (0.025)	3 (0.046)	2 (0.042)	2 (0.042)	1 (0.05)	1 (0.039)
	4.5	0.64 (0.00063)	8 (0.092)	8 (0.097)	8 (0.1)	8 (0.098)	8 (0.091)	8 (0.11)	8 (0.097)
	6.5	0.64 (0.00067)	18 (0.14)	18 (0.13)	18 (0.16)	18 (0.16)	18 (0.14)	18 (0.17)	18 (0.14)
	8.5	0.64 (0.00066)	28 (0.14)	28 (0.12)	28 (0.12)	28 (0.12)	28 (0.12)	28 (0.12)	28 (0.13)
0.5	0.5	0.098 (0.00084)	0 (0.00098)	0 (0.00087)	0 (0.0013)	0 (0.00095)	0 (0.0008)	0 (0.00079)	0 (0.00075)
	1.5	0.29 (0.0014)	1 (0.0053)	0 (0.0041)	0 (0.0048)	0 (0.0028)	0 (0.0013)	0 (0.0014)	0 (0.0012)
	3	0.59 (0.0015)	2 (0.023)	2 (0.024)	2 (0.021)	1 (0.024)	1 (0.022)	1 (0.025)	1 (0.021)
	4.5	0.63 (0.00086)	5 (0.058)	5 (0.057)	5 (0.051)	5 (0.051)	5 (0.044)	5 (0.044)	5 (0.05)
	6.5	0.63 (0.00055)	11 (0.048)	11 (0.058)	11 (0.049)	11 (0.067)	11 (0.068)	11 (0.067)	11 (0.083)
	8.5	0.63 (0.00075)	17 (0.065)	17 (0.078)	17 (0.069)	17 (0.065)	17 (0.056)	17 (0.063)	17 (0.063)
0.8	0.5	0.092 (0.00059)	0 (0.0012)	0 (0.001)	0 (0.0011)	0 (0.00086)	0 (0.00091)	0 (0.00092)	0 (0.00072)
	1.5	0.28 (0.0014)	0 (0.0043)	0 (0.0026)	0 (0.003)	0 (0.0017)	0 (0.0012)	0 (0.0014)	0 (0.0013)
	3	0.55 (0.0017)	2 (0.012)	1 (0.014)	1 (0.012)	1 (0.011)	1 (0.013)	1 (0.01)	1 (0.012)
	4.5	0.61 (0.00047)	3 (0.028)	3 (0.031)	3 (0.023)	3 (0.029)	3 (0.027)	3 (0.034)	3 (0.034)
	6.5	0.61 (0.00062)	7 (0.035)	7 (0.035)	7 (0.04)	7 (0.038)	7 (0.032)	7 (0.036)	7 (0.035)
	8.5	0.61 (0.00072)	11 (0.041)	11 (0.034)	11 (0.036)	11 (0.04)	11 (0.036)	11 (0.039)	11 (0.027)

Table B.8: Average total number at station 2 (95% Confidence Interval).

β	λ	**P2**	**P1**	**E**	**T (K=2)**	**T (K=5)**	**T (K=10)**	**T (K=15)**	**T (K=20)**
0.15	0.5	0.17 (0)	0.12 (0)	0.15 (0)	0.14 (0)	0.17 (0)	0.17 (0)	0.17 (0)	0.17 (0)
	1.5	0.36 (0)	0.14 (0)	0.25 (0)	0.2 (0)	0.32 (0)	0.35 (0)	0.36 (0)	0.36 (0)
	3	11 (0)	0.18 (0)	0.79 (0)	0.29 (0)	0.63 (0)	1.2 (0)	1.7 (0)	2.3 (0)
	4.5	1395 (0)	0.25 (0)	2.5 (0)	0.36 (0)	0.68 (0)	1.2 (0)	1.8 (0)	2.4 (0)
	6.5	2309 (0)	0.49 (0)	4 (0)	0.55 (0)	0.78 (0)	1.2 (0)	1.6 (0)	1.9 (0)
	8.5	2804 (0)	10 (0)	13 (0)	13 (0)	14 (0)	14 (0)	14 (0)	14 (0)
0.3	0.5	0.16 (0.0014)	0.12 (0.0006)	0.15 (0.0012)	0.14 (0.001)	0.16 (0.0014)	0.16 (0.0014)	0.16 (0.0014)	0.16 (0.0014)
	1.5	0.34 (0.0036)	0.14 (0.00058)	0.24 (0.0019)	0.2 (0.00066)	0.31 (0.0022)	0.34 (0.0036)	0.34 (0.0036)	0.34 (0.0035)
	3	6.1 (2)	0.18 (0.00085)	0.58 (0.006)	0.28 (0.0013)	0.58 (0.0026)	1.1 (0.0075)	1.6 (0.015)	2 (0.026)
	4.5	1343 (9.5)	0.24 (0.0012)	1.3 (0.015)	0.35 (0.0014)	0.66 (0.0016)	1.2 (0.0028)	1.8 (0.0043)	2.4 (0.0075)
	6.5	2274 (6.7)	0.48 (0.0054)	2.2 (0.019)	0.57 (0.0058)	0.79 (0.0056)	1.2 (0.0057)	1.6 (0.0075)	1.9 (0.0055)
	8.5	2771 (3.3)	11 (1.7)	12 (1)	11 (2.2)	11 (2.3)	13 (2.3)	13 (2.3)	13 (2.4)
0.5	0.5	0.16 (0.0014)	0.12 (0.00081)	0.15 (0.00098)	0.14 (0.0011)	0.16 (0.0012)	0.16 (0.0013)	0.16 (0.0013)	0.16 (0.00098)
	1.5	0.33 (0.0029)	0.14 (0.00064)	0.22 (0.002)	0.19 (0.00072)	0.3 (0.0022)	0.33 (0.003)	0.33 (0.0029)	0.33 (0.0032)
	3	3.6 (0.56)	0.18 (0.00074)	0.47 (0.0052)	0.26 (0.001)	0.53 (0.0026)	0.98 (0.0061)	1.4 (0.017)	1.8 (0.022)
	4.5	1265 (11)	0.25 (0.0014)	0.91 (0.0071)	0.34 (0.0015)	0.63 (0.0018)	1.2 (0.0025)	1.8 (0.0048)	2.4 (0.006)
	6.5	2222 (6.5)	0.48 (0.0045)	1.5 (0.015)	0.56 (0.0064)	0.79 (0.0065)	1.2 (0.0071)	1.6 (0.0072)	1.9 (0.0085)
	8.5	2730 (4.9)	15 (1.4)	15 (1.4)	10 (1.6)	11 (1.5)	11 (2.1)	12 (2.4)	11 (1.8)
0.8	0.5	0.16 (0.0015)	0.12 (0.00086)	0.14 (0.0011)	0.14 (0.00098)	0.16 (0.0014)	0.16 (0.0013)	0.16 (0.0011)	0.16 (0.0013)
	1.5	0.3 (0.0032)	0.14 (0.0006)	0.21 (0.0012)	0.18 (0.00078)	0.28 (0.002)	0.3 (0.0025)	0.3 (0.0033)	0.3 (0.0028)
	3	2.3 (0.19)	0.18 (0.0008)	0.39 (0.0034)	0.25 (0.001)	0.49 (0.0016)	0.89 (0.0068)	1.2 (0.011)	1.5 (0.021)
	4.5	1153 (9.5)	0.25 (0.0016)	0.67 (0.0045)	0.33 (0.0012)	0.6 (0.002)	1.1 (0.003)	1.7 (0.005)	2.2 (0.007)
	6.5	2141 (6.6)	0.48 (0.0045)	1.1 (0.0092)	0.56 (0.0059)	0.78 (0.0051)	1.2 (0.0055)	1.5 (0.0073)	1.9 (0.0076)
	8.5	2664 (4.7)	10 (2.7)	13 (2.7)	12 (1.5)	12 (2.1)	13 (1.8)	12 (2.1)	11 (1.6)

Table B.9: Average sojourn time at station 1 (95% Confidence Interval).