MODELS FOR DECISION-MAKING AND PERFORMANCE EVALUATION IN EMERGENCY MEDICAL SERVICE SYSTEMS

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

Kenneth Chun-Ling Chong

August 2016

MODELS FOR DECISION-MAKING AND PERFORMANCE EVALUATION IN EMERGENCY MEDICAL SERVICE SYSTEMS

Kenneth Chun-Ling Chong, Ph.D.
Cornell University 2016

We propose four models for evaluating and benchmarking performance in Emergency Medical Service (EMS) systems. We begin by studying tiered EMS systems, whose fleets are comprised of ambulances providing either Advanced Life Support (ALS) or Basic Life Support (BLS), with the goal of studying the effects of vehicle mix (that is, the combination of ALS and BLS ambulances deployed) on system performance. The ideal choice of vehicle mix has been the subject of some debate in the medical literature, and we contribute to this debate by developing a framework with which quantitative comparisons between vehicle mixes can be made.

Noting that vehicle mix affects how resources are allocated to emergency calls, we build two stylized, but complementary models to study decision-making in tiered EMS systems. First, to examine how ambulances would be dispatched to incoming calls of varying severity, we consider the problem of routing and admission control in a loss system featuring two classes of servers and arriving jobs. We formulate this problem as a Markov decision process, and study its theoretical properties. Second, to study how ambulances would be deployed to base locations to maximize coverage, we formulate an integer program. By combining insights from our models, we conclude there are rapidly-diminishing marginal returns associated with biasing towards all-ALS fleets.

Next, we consider systems practicing ambulance redeployment, the strategic relocation of idle ambulances in real time to improve responsiveness to future calls. Finding an optimal redeployment policy is computationally intractable, and a number of heuristic policies have been proposed. We complement this stream of literature by proposing two upper bounds on the performance of optimal redeployment policies (which, in turn, serve as benchmarks for existing heuristic policies). First, we adapt an existing upper bound so that it is provably valid for EMS providers who operate loss systems. Second, we develop a new upper bound based upon a perfect information relaxation of the EMS provider's problem (in which the decisions-maker is clairvoyant), which we tighten by penalizing policies that use future information to make decisions. We evaluate our bounds through extensive computational experiments, and find that they are reasonably tight in small-scale systems.

BIOGRAPHICAL SKETCH

Kenneth was born on November 6, 1988 in San Francisco, California, and grew up in the city. He attended the University of California, Berkeley to pursue a degree in engineering, but realized his motivation for doing so had more to do with math than with actually wanting to build something. With that in mind, he completed his B.S. in Industrial Engineering and Operations Research in May 2010 (and his B.A. in German for some reason). He began his Ph.D at Cornell three months later, and despite his slowness, will hopefully make it out alive.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my parents, Harvey and Janice, for their love and unwavering support, and for the sacrifices they made for the sake of (my brother Calvin's and) my education. I am also indebted to my advisors, Mark Lewis and Shane Henderson, for the patience to put up with me over the years. They have been a pleasure to work with, and have helped me to grow both as an operations researcher and as a person. My minor advisors, Huseyin Topaloglu and Peter Frazier, have been fantastic teachers, and in particular, I would like to thank Huseyin for his collaboration and valuable input on the final chapters of my thesis.

Finally, I am grateful to the Ph.D students that I have gotten to know during my time in the School of ORIE, and not simply because they (perhaps unwillingly) consumed any baked goods I left in the kitchen. In addition to Chek Hin Choi, Andrew Daw, Emily Fischer, Nanjing Jian, Dave Lingenbrink, Steve Pallone, Wei Qian, Patrick Steele, Mutiara Sondjaja, Xiao Ting Zhao, and Zhengyi Zhou, I would like to thank:

James Davis and Julian Sun, for being James Davis and Julian Sun,

James Dong, for knowing literally everything,

Jake Feldman, for being a fellow foodie and 49ers fan,

Daniel Fleischman, for being an amazing teacher and partner-in-crime during College Puzzle Challenges, and for his magnificent balls (brigadeiro),

Cory Girard, for being an endless source of puns and accents,

Weici Hu, for getting me to go to the gym, and constantly reminding me of my receding hairline,

Ravi Kumar, for many hours of thoughtful discussion, Venus Lo, for persuading me not to pursue a career in baking, and

Chaoxu Tong, for letting me know that if I could somehow make it to 20 beers, that anything is possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Emergency Medical Service (EMS) providers are tasked with responding to 911 calls, providing medical treatment on scene, and when necessary, transporting patients to hospitals to receive more definitive care. This task has become increasingly difficult in recent years, as many providers operate in environments characterized by rising call volumes, increasing medical costs, and frequent delays (such as those resulting from traffic congestion and hospital diversion). As a result, maintaining a high level of service often requires the careful and effective management of ambulances and medical personnel.

Considerable attention has been given in the operations research literature to how this can be achieved. Work in this vein spans nearly half a century, for which Brotcorne et al. [18], Goldberg [37], Green and Kolesar [38], Ingolfsson [44], Mason [58], and Swersey [91] provide excellent overviews. This stream of literature has primarily focused upon three aspects of ambulance operations: the strategic decision of *deploying* ambulances to base locations, the tactical decision of *dispatching* ambulances to arriving calls, and the operational decision of *redeploying* idle ambulances in real time in response to improve the system's responsiveness to future calls. Each level of decision-making has been studied extensively, and a wide range of models have been proposed to support EMS providers. We contribute to this growing body of literature in two ways.

1.1 The Vehicle Mix Decision

In Chapters 2 and 3, we study decision-making in tiered EMS systems, which operate fleets that can roughly divided into those providing Advanced Life

Support (ALS), and those providing Basic Life Support (BLS). These types of ambulances can be differentiated by the equipment and personnel on board. ALS ambulances staff paramedics, who receive more extensive medical training, and thus can administer medicine and perform more advanced medical procedures on scene. They are essential components of many EMS systems, because paramedic intervention can have a measurable effect on patient outcomes, particularly in life-or-death situations. However, ALS ambulances are more expensive to operate, and a significant fraction of emergency calls only require BLS responses.

Operating a tiered EMS system necessitates selecting a *vehicle mix*— that is, a configuration of ALS and BLS ambulances to deploy. The ideal choice of vehicle mix has been a topic of some debate in the medical literature, with the primary issue being whether or not BLS ambulances should be included in a fleet at all. Proponents of all-ALS fleets cite the risk (both to the patient and to the EMS provider) associated with providing only a BLS response to a call requiring a paramedic, if ALS ambulances are not available. Proponents of mixed fleets argue that BLS ambulances allow providers to operate larger fleets, thus decreasing overall response times. Both sides of the debate allude to a fundamental trade-off that EMS providers face: that between decreasing the likelihood of inadequately treating high-priority calls, and improving the system's overall responsiveness.

In Chapters 2 and 3, we contribute to this debate by quantifying this trade-off, through models that allow quantitative comparisons among vehicle mixes to be made. This presents two technical challenges. First, comparing fleets requires the use of a performance measure; we do so by associating rewards with calls receiving timely service, and defining performance to be the long-run average reward collected by the system. While this may appear unnatural, we demonstrate in Chapter 3 this framework is fairly general, and can accommodate a wide range of realistic performance measures. Second, changing the composition of the fleet affects how the fleet would be managed: specifically, how ambulances are deployed to base locations and dispatched to incoming calls, which complicates the task of computing performance. We instead gather insights by studying two stylized, but complementary models of decision-making in tiered EMS systems.

Motivated by the question of how dispatchers allocate resources to calls of varying severity, we begin in Chapter 2 with a stylized model of decision-making in this setting. Our model is one of routing and admission control in a loss system featuring two types of servers ("ALS" and "BLS") and two classes of arriving jobs ("high-priority" and "low-priority"). We impose constraints on how jobs can be routed through the system, to mimic the constraints under which dispatchers operate. For instance, we require that high-priority jobs always be admitted whenever a server (of either type) is idle, so as to provide these jobs with some level of service (medical treatment) instead of diverting them from the system. We assign class-dependent and server-dependent rewards to each admitted job, and seek an admission control and routing policy that maximizes the long-run reward collected by the system. We formulate this problem as a Markov Decision Process (MDP), study its theoretical properties, and provide a partial characterization of the optimal policy.

Our MDP model cannot be directly used to guide dispatchers in practical contexts, as by modeling an EMS system as a loss system, we assume away the role that geographical factors (specifically, the locations of ambulances) play in decision-making. Nonetheless, we study it for two reasons. First, the MDP is theoretically interesting in its own right (independently of its applications to the vehicle mix problem), as the corresponding value function, in general, lacks many of the structural properties, such as convexity and supermodularity, on which standard techniques rely. This necessitates the use of more unconventional techniques to identify structure in our MDP. Second, we are more interested in the objective function value attained by an "optimal policy", rather than the policy itself, as it provides some indication of how a system operating under a given vehicle mix would perform.

To supplement our MDP model, we formulate an Integer Program (IP) in Chapter 3 that studies the complementary problem of making deployment decisions in a tiered EMS system. Here, the goal is to locate a fixed number of ALS and BLS ambulances so as to maximize a long-run measure of the system's ability to respond to high-priority and low-priority calls (which also serves as a proxy for system performance). We draw conclusions about the effects of vehicle mix on system performance by combining the insights we draw from both models. This approach may appear flawed, as we consider dispatching and deployment decisions in separate models, and in practice, these decisions interact. One may argue that a unified model that jointly optimizes with respect to dispatching and deployment decisions would result in more reliable conclusions. However, such a model would likely be computationally intractable. Furthermore, we demonstrate in Chapter 3 that both models yield similar qualitative results, and any discrepancies in our quantitative results can be reconciled without much difficulty. Thus, our two models reinforce one another, and suggest that we would obtain similar results from a single, unified model.

Returning to the question of how vehicle mix affects system performance, we subject our two models to extensive numerical experiments on a large-scale EMS system. Our primary finding is that there are rapidly-diminishing marginal returns associated with biasing one's fleet towards all-ALS. That is, there are a wide range of tiered systems that perform comparably to (and in some cases, outperform) an all-ALS system. This finding is robust to changes to our input parameters— that is, to modifications to call arrival rates, the relative frequency of high-priority and low-priority calls, the EMS provider's reward structure, and the relative operating costs of ALS and BLS ambulances. From these results, we draw the conclusion that EMS providers should not have a strong incentive to change their vehicle mixes, provided that their fleets can adequately respond to high-priority calls. Alternatively, an EMS provider tasked with selecting a vehicle mix should make the decision based upon local considerations, instead of any notion of an "ideal" vehicle mix.

1.2 Upper Bounds for Ambulance Redeployment

In Chapters 4 and 5, we turn our attention to EMS systems practicing *ambulance redeployment* (or system status management), which entails strategically relocating idle ambulances in real time so as to improve the system's ability to respond to future call arrivals. The problem of finding an optimal way to do so is computationally intractable; while this problem, in theory, can be formulated as a stochastic dynamic program, this approach succumbs to the curse of dimensionality. To find good redeployment policies, a number of heuristic approaches have been proposed, but a natural question to ask is how close these policies perform to optimal.

To address this question, we consider the problem of upper bounding the performance that is attained by an optimal redeployment policy. Such an upper bound has two practical applications. First, it serves as a benchmark for heuristic policies; a heuristic that attains a performance close to the upper bound is very likely to be near-optimal. Second, it can be used to determine whether implementing ambulance redeployment in an EMS system is worthwhile, as the practice remains controversial, due to the added burden it places on medical personnel. Specifically, if EMS providers do not consider the resulting increase in performance substantial, even if the upper bound were attainable, then implementation can safely be ruled out.

We propose two methods for developing upper bounds. In Chapter 4, we revisit an existing bound by Maxwell et al. [61], which is valid for systems in which calls arriving when all ambulances are busy can queue. We adapt his bound so that it is provably valid for loss systems, in which arriving calls can be redirected to an external service (such as the fire department or a neighboring EMS) during periods of congestion. This adaptation is nontrivial, and requires the introduction of novel ideas. In Chapter 5, we consider an "information relaxation" of the stochastic dynamic programming formulation of the problem. Specifically, we assume that the decision-maker is clairvoyant, and knows the arrival times, locations, and service requirements of all calls at the start of the horizon. This relaxation allows us to overcome the curse of dimensionality (by solving an equivalent integer program), but results in an upper bound that is loose. We then attempt to tighten the upper bound by penalizing, in the spirit of Brown et al. [20], policies that use future information to make decisions. We develop a fairly natural and intuitive class of penalties, and demonstrate that finding the penalty yielding the tightest upper bound can be found by solving a convex optimization problem.

We evaluate our upper bounds through extensive computational experiments on a wide range of systems of varying sizes, to investigate where our bounds perform well, as well as where they tend to be loose. Our "loss system" bound from Chapter 4 is relatively tight in small-scale to moderately-sized systems, but struggles in large-scale (city-size) instances, where it barely outperforms the perfect information upper bound. Our "information penalty" bound from Chapter 5 appears to be effective in small-scale systems, particularly when resources cannot be effectively pooled. However, it appears to be dominated by our loss system bound, as we could not find a problem instance for which the comparison is favorable. Nonetheless, we lay down a theoretical framework upon which an improved penalty (and thus, a tighter upper bound) could potentially be developed.

CHAPTER 2

TWO-CLASS ROUTING AND ADMISSION CONTROL WITH STRICT PRIORITIES

2.1 Introduction

A common decision faced by operators of service systems is that of dynamically allocating system resources to incoming demand. Complicating matters is the fact that in such systems, customers and servers are often heterogeneous. Servers may have different capabilities, or receive varying levels of training. Similarly, different types of customers may have varying needs, or take priority over other customer classes. This situation arises, for instance, in telecommunications [4, 13, 71], healthcare [10, 65], and rental systems [32, 73, 79]. Decision-making in this setting has typically been modeled in the literature as a problem of admission control or routing in a queueing system, canonical examples of which include the models by Harrison [40] and Miller [67].

However, there may be restrictions in the decisions that can be made in certain system states. One example of this (the original motivation for the model we present in this paper) arises in Emergency Medical Service (EMS) systems. In this setting, arriving calls (customers) are categorized into priorities, based upon severity, and ambulances (servers) are staffed by personnel who receive varying levels of medical training. EMS providers are primarily evaluated by their responsiveness to calls of the highest priority: emergencies for which patients' lives are potentially at stake (such as cardiac arrest). As a result, decision-making with these types of calls is typically forced, in that they receive an immediate response if resources are available. Moreover, if multiple ambulances can respond to such a high-priority call, EMS providers prefer to dispatch one staffed by paramedics (who receive the highest level of medical training), as doing so can have a measurable effect on patient outcomes. (See, for instance, Bakalos et al. [7] or Jacobs et al. [47] for discussions of these effects.) However, EMS providers are often required to provide a minimum level of service to lower-priority calls. As a result, decision-makers face a trade-off between keeping resources available to serve higher-priority calls, and allocating these resources to adequately serve lower-priority calls.

We study this situation through a loss system featuring two types of servers (Type A and Type B), that is tasked with processing two classes of jobs: Type H (or high-priority) jobs and Type L (or low-priority) jobs, in which decision-making with Type H jobs is forced, and Type L jobs are subject to admission control. Although this model cannot be directly used to guide decision making in practical contexts (as the locations of ambulances must also be taken into account), it is theoretically interesting in its own right. This is because the element of forced decision-making in our model presents significant technical challenges. In particular, standard techniques for characterizing the optimal admission control policy, which involve formulating our model as an MDP and analyzing the corresponding value function, are inadequate. Although there are conditions under which standard techniques suffice, in general, the value function associated with our model lacks classical structural properties, such as convexity or supermodularity. Moreover, this lack of structure can be directly attributed to our requirement that decision-making with Type H jobs is forced.

To address these difficulties, we instead identify situations in which we can provably recover structure in the optimal policy, and proceed in two directions.

First, we provide a sufficient (but not necessary) condition on our model inputs under which our value function is convex and supermodular. Second, we demonstrate that when we restrict attention to a certain intuitive class of policies, the optimal policy in this class is "monotone" in a way that we later specify, the proof of relies on a novel argument using renewal theory. One may wonder whether the technical challenges our model presents can be circumvented, for instance by devising an effective heuristic policy, but numerical experiments suggest that in service systems such as ours, that there is value in taking into account heterogeneity of both servers and jobs.

The remainder of this chapter is organized as follows. After a review of relevant literature in Section 2.2, we explicitly formulate our model as a Markov decision process in Section 2.3, and present the corresponding optimality equations. We identify basic structural properties in our value function in Section 2.4, which we leverage in Section 2.5 to identify conditions under which the optimal policy can be characterized. Following a brief numerical study in Section 2.6, we conclude in Section 2.7.

2.2 Literature Review

Markov decision processes have been used to study real-time decision making in EMS systems, for which the goal is to identify policies for dispatching ambulances to emergency calls, or to relocate idle ambulances to improve the system's responsiveness to future call arrivals. See, for instance, the models by Berman [10, 11], Jarvis [49], McLay and Mayorga [65], and Zhang [101]. These models are detailed, and take into account the effects of ambulance locations (and heterogeneity) when making decisions, and thus, can be used to guide decision-making in practical contexts. Although our model lacks this applicability, we consider a more nuanced objective function that incorporates the system's responsiveness to high-priority and low-priority calls, as well as the level of service that different types of ambulances can provide to these calls. It can also be used to draw basic insights; we use this model in [24] to study the effects of fleet composition (the mixture of "Type A" and "Type B" ambulances deployed) on the performance of EMS systems. Our choice to model the system as a loss system is partly due to tractability, and partly due to the fact that some systems divert calls to an external service (such as the fire department) during periods of congestion.

Routing and admission control in queueing systems has been widely studied in the literature. Surveys, such as those by Stidham [86] and by Stidham and Weber [87], provide excellent overviews of work in this area; we restrict our attention to models closely related to our own— specifically, models featuring multiple classes of arriving jobs. Perhaps the most influential model of this type is due to Miller [67], who studies the problem of admission control to a loss system featuring homogeneous servers and n job classes, as well as class-dependent rewards. Extensions to Miller's model have been studied (see, for instance, Carrizosa et al. [22], Feinberg and Reiman [31], and Lewis et al. [55]), and his model has also been adapted to study telecommunications systems and call centers (examples of which include Altman et al. [4], Bhulai and Koole [13], Blanc et al. [14], Gans and Zhou [33], and Örmeci and van der Wal [71]). Our model differs from those previously mentioned in two respects. First, it features two types of servers, and we allow rewards to depend on the type of server to which jobs are assigned. Second, one of our arrival streams is uncontrolled, as Type H jobs must be admitted whenever possible. While Blanc et al. [14] also consider a model with a similar element of forced decision-making, and Bhulai and Koole [13] and Gans and Zhou [33] impose a service level constraint on Type H jobs, they do so in a system with homogeneous servers.

Our model also closely resembles the N-network, which can be viewed as a variant of our model in which $R_{HA} = R_{HB}$, Type L jobs cannot be routed to Type A servers, and jobs can queue. See, for instance, Bell and Williams [9], Down and Lewis [29], and Harrison [40]. While our model does not feature a queue, we allow for servers to be flexible, in that they can serve both types of jobs, and discourage certain routing decisions through the reward structure we impose.

Finally, we draw a connection between our model and a stream of literature relating to capacity management in rental systems, a relatively small subfield of revenue management in which individual resources are not perishable (as is the case, for instance, with seats on a particular flight itinerary), but can be reused to generate revenue from multiple customers. In this setting, servers are viewed as resources that can be rented to customers for random (exponentially distributed) durations of time. When resources are scarce, there is a decision as to which resources (if any) to make available to arriving customers. This problem has been studied, for instance, by Gans and Savin [32], Savin et al. [79], and Papier and Thonemann [73], all of whom model resources as homogeneous. In this context, our model can be viewed as that of a rental system with two classes of resources, in which product substitutions can be made during periods of high demand. For instance, assigning a Type L job with a Type A server may correspond to upgrading a low-priority customer, whereas assigning a Type H job to a Type B server may correspond to downgrading (and compensating) a high-priority arrival.

2.3 Model Formulation

Consider a system operating N_A Type A and N_B Type B servers. Type H and Type L jobs arrive according to independent Poisson processes with rates λ_H and λ_L, respectively. An arriving Type H job must be admitted into the system if at least one server is idle, and if this is the case, must be processed by a Type A server if one is available. Routing a Type H job to a Type B server is less desirable, but is permitted when all Type A servers are busy, to provide the job with some level of service during periods of congestion. Type L jobs can be processed by either type of server, but can be diverted from the system upon arrival, so as to reserve system resources for future Type H jobs. If a job (of either type) is admitted (with either type of server), it leaves the system after a time that is exponentially distributed with rate μ. Jobs that are diverted or that arrive when all servers are busy immediately leave the system.

Let R_{HA} and R_{HB} denote the reward associated with assigning a Type H job to a Type A and Type B server, respectively. Similarly, let R_L denote the reward associated with admitting a Type L job (and assigning it to either type of server). We assume $R_{HA} \geq \max\{R_{HB}, R_L\}$, but we make no assumptions about the relative ordering of R_{HB} and R_L. Our goal is to find an admission control policy for low-priority jobs that maximizes the expected long-run reward collected by the system, and consider both the discounted reward and the long-run average reward criteria. We use this objective to quantify the level of service that the system is able to provide. By setting R_{HA} to be the largest reward in our model, we prioritize serving Type H jobs adequately, and we can model the trade-off between serving Type H jobs and Type L jobs through the value of R_L relative to those of R_{HA} and R_{HB}.

We formulate the problem described as a Markov decision process (MDP). Let $\mathcal{S} = \{0, 1 \ldots, N_A\} \times \{0, 1 \ldots, N_B\}$ be the state space, where $(i, j) \in \mathcal{S}$ denotes the state in which i Type A servers and j Type B servers are busy. To determine the actions $\mathcal{A}(i, j)$ that are available when the system is in state (i, j), it suffices to only consider arriving Type L jobs. If either $i < N_A$ or $j < N_B$, two actions can be taken: admitting (action 1) or rejecting (action 0) the next Type L job, if one arrives during the next decision epoch. If $j < N_B$, action 1 entails assigning the call to a Type B server. Although we do not allow Type L calls to be assigned to Type A servers in this situation, this is without loss of optimality; we prove this in Proposition 2.4.3 below. If $j = N_B$, but $i < N_A$, then action 1 entails a Type A service. Finally, if $i = N_A$ and $j = N_B$, only a dummy action (action 0) can be taken.

Because interevent times are exponentially distributed with a rate that is bounded above by $\Lambda := \lambda_H + \lambda_L + (N_A + N_B)\mu$, our MDP is uniformizable in the spirit of Lippman [56] and Serfozo [81], and we can consider an equivalent process in discrete time. Without loss of generality, assume $\Lambda = 1$. We define a policy to be a sequence of decision rules $\pi = \{\pi^0, \pi^1, \ldots\}$, where $\pi^k : \mathcal{S} \to \{0, 1\}$ specifies a deterministic action to be taken during the k^{th} decision epoch, given the state of the system. Let Π be the set of all such policies. Given a fixed policy $\pi \in \Pi$ and an initial state (i, j), let S_k^π be the state of the system at the start of the k^{th} decision epoch, and A_k^π be the action $\pi^k(S_k^\pi)$ selected by policy π at this time. Considering first the case of discounted rewards. We define the *expected total discounted reward* collected by π to be

$$v_\alpha^\pi(i, j) = \mathbb{E}\left[\sum_{k=0}^{\infty} \alpha^k r(S_k^\pi, A_k^\pi) \,\middle|\, S_0^\pi = (i, j)\right], \tag{2.1}$$

where $r(s, a)$ is the reward collected when the system is in state $s \in \mathcal{S}$ and action $a \in \mathcal{A}(s)$ is taken, and $\alpha \in [0, 1)$ is the discount factor. This quantity is well-defined for each initial state (i, j), since $0 \leq r(s, a) \leq R_{HA}$ for each s and a. Next, we define $v_\alpha(i, j) = \sup_{\pi \in \Pi} v_\alpha^\pi(i, j)$. Because state and action spaces are finite, the supremum is attained, and so $v_\alpha(i, j)$ denotes the total discounted reward obtained by an optimal policy, given a system initialized in state (i, j). Theorem 6.2.6 of Puterman [75] implies that the value function v_α is the unique solution to the optimality equations:

$$v_\alpha(i,j) = T_\alpha v_\alpha(i, j) \tag{2.2}$$

$$\begin{aligned}
:= \lambda_H \Big[& \mathbf{1}_{\{i<N_A\}} \big[R_{HA} + \alpha v_\alpha(i+1, j))\big] \\
& + \mathbf{1}_{\{i=N_A, j<N_B\}} \big[R_{HB} + \alpha v_\alpha(i, j+1)\big] \\
& + \mathbf{1}_{\{i=N_A, j=N_B\}} \alpha v_\alpha(i, j)\big]\Big] \\
+ \lambda_L \Big[& \mathbf{1}_{\{j<N_B\}} \max\big\{R_L + \alpha v_\alpha(i, j+1), \alpha v_\alpha(i, j)\big\} \\
& + \mathbf{1}_{\{i<N_A, j=N_B\}} \max\big\{R_L + \alpha v_\alpha(i+1, j),\, \alpha v_\alpha(i, j)\big\} \\
& + \mathbf{1}_{\{i=N_A, j=N_B\}} \alpha v_\alpha(i, j)\Big] \\
+ i\mu\, & \alpha v_\alpha(i-1, j) + j\mu\, \alpha v_\alpha(i, j-1) + (N_A + N_B - i - j)\mu\, \alpha v_\alpha(i, j).
\end{aligned} \tag{2.3}$$

The first term on the right-hand side of (2.3) corresponds to the case when a Type H arrival occurs in the next decision epoch. The reward collected and the resulting transition depends on the system state; we capture this dependence using indicators for brevity. The remaining four terms correspond to cases where a Type L arrival, a Type A service completion, a Type B service completion, and a dummy transition occur, respectively. Note that the optimality equations (2.3) imply that it is optimal to accept a Type L arrival when either $j < N_B$ and $R_L > \alpha[v_\alpha(i, j) - v_\alpha(i, j+1)]$ or $i < N_A$, $j = N_B$, and $R_L > \alpha[v_\alpha(i, j) - v_\alpha(i+1, j)]$.

We also consider a finite-horizon analogue of this problem, in which we terminate the decision process after n decision epochs. Define the functions $v^{\pi}_{n+1,\,\alpha}$ and $v_{n,\alpha}$ analogously to v^{π}_{α} and v_{α}, but with the sum in (2.1) terminating at n instead of ∞. Here, we allow $\alpha = 1$. The optimality equations for this problem can be constructed analogously by replacing v_{α} on the left-hand side of (2.3) with $v_{n,\alpha}$, and occurrences of v_{α} on the right-hand side of (2.3) with $v_{n-1,\alpha}$ (and specifying the boundary condition $v_{0,\alpha}(i,\, j) = 0$ for all i and j).

Given any initial state $(i,\, j)$ and a policy π, we define the *long-run average reward* attained to be $J^{\pi} = \lim_{n\to\infty} v_{n,\,1}(i,\, j)/n$. By Theorem 8.3.2 of Puterman [75], J^{π} is well-defined and independent of $(i,\, j)$, as the Markov chain induced by π is irreducible. To see this, suppose $(i,\, j),\ (i',\, j') \in \mathcal{S}$. Then state $(i',\, j')$ can be reached from state $(i,\, j)$ under π via $i + j$ consecutive service completions, followed by $i' + j'$ Type H arrivals. Next, define $J = \sup_{\pi\in\Pi} J^{\pi}$, the long-run average reward attained by an optimal policy. By Theorem 8.4.3 of [75], this can be found by solving the optimality equations

$$J + h(i,\, j) = T_1 h(i,\, j) \tag{2.4}$$

for J and $h(\cdot)$, where we defined the operator T_{α} in (2.2). One property of h that we use extensively is

$$h(i, j) - h(i', j') = \lim_{n\to\infty} \left[v_{n,1}(i, j) - v_{n,1}(i', j')\right], \tag{2.5}$$

which is proven, for instance, in Section 8.2.1 of Puterman [75].

2.4 Basic Structural Properties

We begin our analysis by identifying structural properties of the value functions v_{α}, $v_{n,\alpha}$, and h. The first such property we examine confirms two intuitive notions: that additional idle servers are beneficial to the system, and that an idle

Type A server is preferable to an idle Type B server (which, in turn, implies that we prefer to assign Type L calls to Type B servers than to Type A servers).

Lemma 2.4.1. *For all $\alpha \in [0,1)$, and $f = v_\alpha$, $v_{n,\alpha}$, or h (depending on the optimality criterion), we have*

1. $f(i,\, j) - f(i+1,\, j) \geq 0 \qquad i = 0,\, \ldots,\, N_A - 1,\, j = 0,\, \ldots,\, N_B,$
2. $f(i,\, j) - f(i,\, j+1) \geq 0 \qquad i = 0,\, \ldots,\, N_A,\, j = 0,\, \ldots,\, N_B - 1,$ *and*
3. $f(i,\, j+1) - f(i+1,\, j) \geq 0 \qquad i = 0,\, \ldots,\, N_A - 1,\, j = 0,\, \ldots,\, N_B - 1.$

Proof. We show Statement 2 holds via a sample path argument; the proofs of Statements 1 and 3 are similar. Suppose $f = v_\alpha$; the case where $f = v_{n,\alpha}$ follows via a nearly identical proof, from which the case where $f = h$ follows by leveraging Equation (2.5). Fix $\alpha \in [0,1)$, $i \in \{0,\, \ldots,\, N_A\}$, and $j \in \{0,\, \ldots,\, N_B - 1\}$. We construct two processes on the same probability space. Process 2 begins in state $(i,\, j+1)$ and follows the optimal policy π^*, whereas Process 1 starts in $(i,\, j)$ and uses a potentially suboptimal policy π that imitates the actions taken by Process 2.

By construction, arrivals occur simultaneously in both processes, and any job admitted by Process 2 is also admitted by Process 1. Furthermore, any service completion occurring in Process 1 also occurs in Process 2. Thus, both process move in "parallel" (in that the same changes in state occur simultaneously in both processes) until one of the following events occurs:

1. Process 2 sees a service completion that is not observed by Process 1.
2. A Type H arrival when Process 1 is in state $(N_A,\, N_B - 1)$, and Process 2 is in state $(N_A,\, N_B)$.

Either event causes both processes to *couple*, in that they transition into the same state, and behave identically from this time onward. Let Δ be a random variable denoting the difference in reward collected by the two processes until coupling occurs. By Equation (2.1), $\mathbb{E}\Delta = v_\alpha^\pi(i,\, j) - v_\alpha(i,\, j+1)$. Thus, it suffices to show that $\mathbb{E}\Delta \geq 0$, as this implies $v_\alpha(i,\, j) - v_\alpha(i,\, j+1) \geq v_\alpha^\pi(i,\, j) - v_\alpha(i,\, j+1) = \mathbb{E}\Delta \geq 0$. Indeed, prior to the coupling event, both processes observe the same transitions and collect the same rewards. When coupling occurs, Process 1 collects a reward at least as large as that by Process 2. This implies $\Delta \geq 0$ pathwise, and so $\mathbb{E}\Delta \geq 0$, as desired. □

We next establish two upper bounds: one on the benefit of an idle Type B server, and one on the benefit associated with substituting an idle Type B server with an idle Type A server.

Lemma 2.4.2. *For all $\alpha \in [0, 1)$, and $f = v_\alpha$, $v_{n,\alpha}$, or h, we have that*

1. $f(i,\, j+1) - f(i+1,\, j) \leq R_{HA} - R_{HB}$ $\quad i \in \{0, \ldots, N_A - 1\}$, $j \in \{0, \ldots, N_B\}$
2. $f(i,\, j) - f(i,\, j+1) \leq \max\{R_{HB},\, R_L\}$ $\quad i \in \{0, \ldots, N_A\}$, $j \in \{0, \ldots, N_B - 1\}$

Proof. We show Statement 1 holds via a sample path argument; the proof of Statement 2 is similar. As in the proof of Lemma 2.4.1, it suffices to show that the above properties hold when $w = v_\alpha$. Fix $\alpha \in [0, 1)$, $i \in \{0,\, \ldots,\, N_A - 1\}$, and $j \in \{0,\, \ldots,\, N_B - 1\}$. Construct two processes on the same probability space. Process 1 begins in state $(i,\, j+1)$ and follows the optimal policy π^*, whereas Process 2 begins in state $(i+1,\, j)$, and imitates the actions taken by Process 1.

There is a Type A server that is idle in Process 1, but busy in Process 2, and a Type B server that is busy in Process 1, but idle in Process 2. We construct our probability space so that both units complete service simultaneously. Both processes move in parallel until one of the following occurs:

1. The coupled Type A server (in Process 1) and Type B server (in Process 2) complete service.
2. A Type H arrival occurs when Process 1 is in state $(N_A - 1, j' + 1)$, and Process 2 is in state (N_A, j'), for some $j' \in \{0, 1, \ldots, N_B - 1\}$. (In this case, Process 1 admits the job with a Type A server, and Process 2 admits the job with a Type B server.)
3. A Type L arrival occurs when Processes 1 and 2 are in states $(i' - 1, N_B)$ and $(i', N_B - 1)$, respectively, for some $i' \in \{1, 1, \ldots, N_A\}$ and Process 1 admits the job with a Type A server. (In this case, Process 2 admits the job with a Type B server.)

Let Δ be the difference in reward collected by the two processes until coupling occurs. Since $\mathbb{E}\Delta = v_\alpha(i, j + 1) - v_\alpha^\pi(i + 1, j)$, it suffices to show that $\mathbb{E}\Delta \leq R_{HA} - R_{HB}$. We observe that $\Delta(\omega) = 0$ on all paths ω in which events 1 or 3 occur, and $R_{HA} - R_{HB}$ on paths in which event 2 occurs (modulo the effects of discounting). Thus $\Delta \leq R_{HA} - R_{HB}$ pathwise. □

Lemma 2.4.2 has two implications on the structure of optimal policies.

Proposition 2.4.3. *If a Type H job arrives when both types of servers are available, then it is preferable to assign a Type A server. Moreover, if $R_{HB} \leq R_L$, then it is optimal to admit low-priority jobs when at least one Type B server is idle.*

Proof. We prove the first claim; the second follows using a similar argument. It suffices to show that our claim holds under the discounted reward criterion. Let (i, j) be such that $i < N_A$ and $j < N_B$, and suppose, contrary to the optimality equations (2.3), that we can serve an arriving Type H job in this state with a Type L server. It is preferable to assign a Type H server if $R_{HA} + \alpha v_\alpha(i+1,\, j) \geq R_{HB} + \alpha v_\alpha(i,\, j+1)$, which by Statement 1 of Lemma 2.4.2, always holds. □

2.5 Optimal Policy

If $N_B = 0$, we can use a result by Miller [67] to characterize the optimal policy.

Proposition 2.5.1 (Miller 1969)**.** *If $N_B = 0$, then there exists an optimal policy with the property that if it admits Type L jobs when i servers are busy, then it also does so when $i' < i$ servers are busy.*

That is, *threshold-type* policies are optimal. When $N_B > 0$, it is reasonable to conjecture that a multi-dimensional analogue to this class of policies is optimal:

Definition 2.5.2. *A policy is of* ***monotone switching curve*** *type if there exists a monotone curve $s(\cdot)$ dividing the state space into two connected regions, one in which action 0 is taken, and one in which action 1 is taken.*

Threshold-type policies are special cases of monotone switching curve policies. We analyze the cases $R_L \leq R_{HB}$ and $R_L > R_{HB}$ separately.

2.5.1 The Case $R_L \leq R_{HB}$

If $R_L \leq R_{HB}$, the optimal policy is fairly structured, and our main result in this section as follows:

Theorem 2.5.3. *If $R_L \leq R_{HB}$, there exists an optimal monotone switching curve policy with slope of at least -1 under both the discounted reward and long-run average reward criteria.*

A monotone switching curve policy can be viewed as one that keeps some number of Type B servers in reserve to respond to future Type H arrivals, and grows this reserve as more Type A servers become busy. A bound on the slope of the switching curve implies the size of this reserve does not change dramatically in response to “small” changes in system state. In particular, if a Type A server becomes free or busy, the size of the reserve can change by at most one. To prove Theorem 2.5.3, we use the fact that our value functions v_α, $v_{n,\alpha}$, and h have the following structural properties:

Lemma 2.5.4. *If $R_L \leq R_{HB}$, then for all $\alpha \in [0, 1)$ and $f = v_\alpha$, $v_{n,\alpha}$, or h, we have*

1. *(Convexity in j) For all $i \in \{0,\, 1, \ldots,\, N_A\}$ and $j \in \{0,\, 1, \ldots,\, N_B - 2\}$,*

$$f(i,\, j) - w(i,\, j+1) \leq f(i,\, j+1) - f(i,\, j+2) \tag{2.6}$$

2. *(Supermodularity) For all $i \in \{0,\, 1, \ldots,\, N_A - 1\}$ and $j \in \{0,\, 1, \ldots,\, N_B - 1\}$,*

$$f(i,\, j) - w(i,\, j+1) \leq f(i+1,\, j) - f(i+1,\, j+1). \tag{2.7}$$

3. *(Convexity in i when $j = N_B$) For all $i \in \{0,\, 1, \ldots,\, N_A - 2\}$,*

$$f(i,\, N_B) - f(i+1,\, N_B) \leq f(i+1,\, N_B) - w(i+2,\, N_B). \tag{2.8}$$

4. *(Slope property) For $0 \leq i \leq N_A - 1$ and $0 \leq j \leq N_B - 2$,*

$$f(i+1,\, j) - f(i+1,\, j+1) \leq f(i,\, j+1) - f(i,\, j+2). \tag{2.9}$$

The proof, which we defer to Appendix A, follows by demonstrating that properties (2.6)–(2.9) hold when $w = v_{n,\alpha}$, via a straightforward induction argument on n, then reasoning as in Lemma 2.4.1 to show that the same properties hold for the value functions v_α and h.

Proof of Theorem 2.5.3. We consider only the discounted reward criterion, as the proof for the long-run average reward criterion is nearly identical. Consider an optimal policy π^*, and let $(i,\, j+1)$ be a state in which π^* takes action 1— that is, admits Type L jobs into the system. (If such a state does not exist, then our claim trivially holds.) If $j+1 < N_B$, then the optimality equations (2.3) imply $R_L + \alpha v_\alpha(i,\, j+2) \geq \alpha v_\alpha(i,\, j+1) \iff R_L \geq \alpha[v_\alpha(i,\, j+1) - v_\alpha(i,\, j+2)]$. Statement 1 of Lemma 2.5.4 implies that $R_L \geq \alpha[v_\alpha(i,\, j) - v_\alpha(i,\, j+1)]$, and so it is also optimal to take action 1 in state $(i,\, j)$. If $j+1 = N_B$, action 1 routes the Type L job to a Type A server, and

$$\begin{aligned} R_L + \alpha v_\alpha(i+1,\, j+1) \geq \alpha v_\alpha(i,\, j+1) &\iff R_L \geq \alpha[v_\alpha(i,\, j+1) - v_\alpha(i+1,\, j+1)] \\ &\implies R_L \geq \alpha[v_\alpha(i,\, j) - v_\alpha(i+1,\, j)] \\ &\implies R_L \geq \alpha[v_\alpha(i,\, j) - v_\alpha(i,\, j+1)], \end{aligned}$$

where the second line follows by Statement 2 of Lemma 2.5.4, and the third by Statement 3 of Lemma 2.4.1. Thus, it is again optimal to take action 1 in state $(i,\, j)$. Similar reasoning yields the conclusion for all states $(i,\, j')$ where $j' < j$.

Now consider a state $(i+1,\, j)$ at which π^* admits Type L jobs (assuming without loss of generality that one exists). If $j < N_B$, then $R_L \geq \alpha[v_\alpha(i+1,\, j) - v_\alpha(i+1,\, j+1)]$, and Statement 2 of Lemma 2.5.4 implies $R_L \geq \alpha[v_\alpha(i,\, j) - v_\alpha(i,\, j+1)]$. If $j = N_B$, then $R_L \geq \alpha[v_\alpha(i+1,\, j) - v_\alpha(i+2,\, j)]$, and Statement 3 of Lemma 2.5.4 implies that $R_L \geq \alpha[v_\alpha(i,\, j) - v_\alpha(i+1,\, j)]$. In either case, it is optimal to admit Type L jobs in state $(i,\, j)$. Similar reasoning can be used to show that Type L jobs are also admitted in state $(i',\, j)$ where $i' < i$.

Thus, if π^* admits Type L jobs in state $(i,\, j)$, then it also does so in all states $(i',\, j')$ for which $i' \leq i$ and $j' \leq j$. For each $i \in \{0,\, 1,\, \ldots,\, N_A\}$, define the function $s(i) = \max\{j : \pi^*(i,\, j) = 1\}$; we claim this function is nonincreasing.

Indeed, if this is not the case, then there exists an i for which $s(i+1) > s(i)$, implying that for some j, the policy admits Type L jobs in state $(i+1,\, j+1)$, but not in state $(i,\, j+1)$. Contradiction.

To prove that s has slope of at least -1, it suffices to show that if π^* admits Type L jobs when the system is in state $(i,\, j+1)$, then it also does so in state $(i+1,\, j)$. Let $(i,\, j+1)$ be such a state, and suppose first that $j+1 < N_B$. Then $R_L \geq \alpha[v_\alpha(i,\, j+1) - v_\alpha(i,\, j+2)]$, and Statement 4 of Lemma 2.5.4 implies $R_L \geq \alpha[v_\alpha(i+1,\, j) - v_\alpha(i+1,\, j+1)]$. If $j+1 = N_B$, then Lemma 2.4.1 implies $R_L \geq \alpha[v_\alpha(i,\, j+1) - v_\alpha(i+1,\, j+1)] \geq \alpha[v_\alpha(i+1,\, j) - v_\alpha(i+1,\, j+1)]$. $\square$

We conclude this section by noting that when $R_L = R_{HB}$, the optimal policy is simpler to characterize, as Lemma 2.4.2 implies that this policy must admit Type L jobs whenever Type B servers are idle. Combining this insight with Theorem 2.5.3 implies the existence of an optimal threshold-type policy.

2.5.2 The Case $R_L > R_{HB}$

As in the case when $R_L = R_{HB}$, we can leverage Lemma 2.4.2, and consider decision-making only in states $(i,\, N_B)$, where $i < N_A$. We conjecture that a threshold-type policy is optimal here. However, we cannot reason as in Section 2.5.1, as the value function v_α is, in general, neither convex nor supermodular. We demonstrate this with an example. While we specifically consider the discounted reward criterion, we can adopt the example below to the case of long-run average rewards.

Example 2.5.5. Consider a system where $\lambda_A = 40$, $\lambda_B = 30$, $R_{HA} = 1$, $R_{HB} = 0.1$, $R_L = 0.9$, $N_A = 2$, $N_B = 28$, $\mu = 1$, and $\alpha = 0.995$. Policy iteration yields

$v_\alpha(2,\,28) = 28.479$, $v_\alpha(1,\,28) = 29.545$, $v_\alpha(2,\,27) = 28.914$, $v_\alpha(1,\,27) = 30.125$, and $v_\alpha(0,\,28) = 30.620$, Thus,

$$0.580 = v_\alpha(1,\,27) - v_\alpha(1,\,28) > v_\alpha(2,\,27) - v_\alpha(2,\,28) = 0.435 \quad \textit{and}$$

$$1.075 = v_\alpha(0,\,28) - v_\alpha(1,\,28) > v_\alpha(1,\,28) - v_\alpha(2,\,28) = 1.066,$$

and so v_α is neither supermodular nor convex. □

Example 2.5.5 establishes that the value functions v_α and h, in general, are unstructured. It can be shown that this lack of structure can be directly attributed to our assumption that decision-making with Type H jobs is forced. In particular, if the decision-maker could subject Type H jobs to admission control, then we return to a setting in which standard techniques suffice to characterize optimal policies:

Proposition 2.5.6. *Consider a modified system in which Type H jobs are subject to admission control, in that they can be rejected upon arrival in any system state. Let* $\tilde{v}_\alpha$*,* $\tilde{v}_{n,\alpha}$*, and* $\tilde{h}$ *denote the value functions associated with optimal policies in this setting. If* $R_L > R_{HB}$*, then for all* $\alpha \in [0, 1)$*, these functions are convex (that is, convex in* i*, and convex in* j *when* $i = N_A$*) and supermodular, and there exists an optimal monotone switching curve policy for Type H jobs.*

The proof, which we again defer to Appendix A, is similar to that used to prove Lemma 2.5.4, in that we prove that the value function $v_{n,\alpha}$ (and consequently, v_α and h) has certain structural properties via induction on n. Although forced-decision making complicates the analysis of our model, to ensure the existence of an optimal threshold-type policy, it is only necessary to show that the value functions $v_{n,\alpha}$, v_α, and h satisfy the "single-crossing" property

$$v_\alpha(i-1,\,N_B) - v_\alpha(i,\,N_B) \le R_L \implies v_\alpha(i-1,\,N_B) - v_\alpha(i,\,N_B) \le R_L \qquad (2.10)$$

for all $i \in \{0, \ldots, N_A - 2\}$. Although we conjecture this holds, we have been unable to develop a proof. Nor have we found a counterexample, as extensive numerical experiments on a wide range of problem instances have all yielded optimal threshold-type policies. We proceed by imposing additional assumptions that allow us to identify structure, and conclude this section with two results in this vein.

2.5.2.1 A Sufficient Condition for Convexity

Example 2.5.5 is an unrealistically overloaded system in which the arrival rate greatly exceeds the system's service capacity. It may not be of practical interest to study such systems, even if threshold policies can be shown to be optimal. By restricting our attention to more reasonable parameter values, we identify conditions under which the value functions v_α and h are convex, implying the optimality of threshold-type policies. One such condition is the following:

Proposition 2.5.7. *Fix $\alpha \in [0, 1)$. If*

$$R_L \leq R_{HB} + \frac{\mu}{\lambda_L} R_{HB} + \frac{\mu}{\lambda_H}\left(1 + \frac{\mu}{\lambda_L} + \frac{\lambda_H}{\lambda_L}\right) R_{HA}, \tag{2.11}$$

then for all $i \in \{0, 1 \ldots, N_A - 2\}$, $j \in \{0, 1, \ldots, N_B\}$, and $n \geq 0$, the value functions v_α, $v_{n,\alpha}$, and h are convex in i.

The proof, which we provide in the Appendix, is a sample path argument. The intuition is that $v_\alpha(i, j) - v_{n,\alpha}(i+1, j)$ can be viewed as the expected difference in rewards collected by two stochastic processes defined on the same probability space— one initialized in state (i, j), the other in state $(i + 1, j)$— until coupling occurs; call this expectation $\mathbb{E}\Delta$. The quantity $v_\alpha(i+1, j) - v_{n,\alpha}(i+2, j)$ can be interpreted in a similar fashion; call the corresponding expectation $\mathbb{E}\Delta'$.

There may be sample paths on which $\Delta > \Delta'$, and if the probability of this collection of paths is too large, we may have $\mathbb{E}\Delta > \mathbb{E}\Delta'$; this is the case in Example 2.5.5. Condition (2.11) guards against this possibility. While this condition is certainly sufficient to prove convexity, it is certainly not necessary; numerical experiments suggest that convexity holds for a wide range of parameter values violating inequality (2.11).

2.5.2.2 Threshold Policies

Policies without threshold structure is unappealing from a practical point of view. Thus, it may be reasonable to omit such policies from consideration. Restricting attention to the set of threshold-type policies may still yield value functions that are neither convex or supermodular. The optimal policy in Example 2.5.5 never admits Type L jobs when $j = N_B$, and thus is of threshold type.

Nonetheless, the optimal policy in this setting satisfies a monotonicity property, in that the optimal choice of threshold is nonincreasing in R_{HA}, nondecreasing in R_{HB}, and nondecreasing in R_L. More formally, for $i \in \{-1, 0, \dots, N_A\}$, let π_i denote the threshold-type policy that admits Type L jobs in all states (i', N_B) where $i' \leq i$. (Policy π_{-1} never assigns Type A servers to Type L jobs.) Since the set of threshold-type policies is finite, there exists an optimal policy, but it may not be unique; we break ties by selecting the policy with the highest threshold.

Proposition 2.5.8. *Consider a system with rewards R_{HA}, R_{HB}, and R_L, and let π_{i^*} denote the largest threshold-type policy that is optimal. Suppose we modify the system so that it has rewards R'_{HA}, R'_{HB}, and R'_L, where $R'_{HA} \geq R_{HA}$, $R'_{HB} \leq R_{HB}$, and $R'_L \leq R_L$. Let π_{ℓ^*} be the largest threshold-type policy that is optimal in the modified system. Then $\ell^* \leq i^*$.*

The proof, which we again defer to Appendix A, involves a novel application of renewal theory. We define two stochastic processes on the same probability space (under the original reward structure), one using the optimal policy π_{i^*}, and one using a suboptimal policy π_ℓ, where $\ell > i^*$. We initialize both systems in some state $(i_0,\, j_0)$, and define renewal epochs to be the points in time at which both processes find themselves in state $(i_0,\, j_0)$. When we consider the difference in rewards collected by the two processes during a single renewal epoch (this suffices, due to the Renewal Reward Theorem), we find that the gap widens when we increase R_{HA}, decrease R_{HB}, or decrease R_L. Thus, any policy with a larger threshold than i^* remains suboptimal when when we modify the reward structure in this way.

Proposition 2.5.8 is intuitive, as if we modify rewards in our system so as to more heavily prioritize Type A responses to Type H jobs, or to decrease the importance of serving Type L jobs, we would be less willing to assign Type A servers to Type L jobs.

2.6 Computational Study

In this section, we compare our optimal policy to three heuristic policies:

- A *myopic* policy that admits every incoming job, regardless of the system state (as long as servers are available),
- A *single threshold* policy that admits both types of jobs whenever $i < N_A$, but admits Type L jobs according to a threshold policy when $i = N_A$.
- A *diagonal threshold* policy that always admits Type H jobs whenever possible, but rejects Type L jobs if the total number of busy servers exceeds a threshold t— that is, in all states $(i,\, j)$ where $i + j > t$.

The primary motivation for the latter two policies is that they are two-dimensional analogues of the threshold-type policies that were shown to be optimal in Miller [67]. The single threshold policy considers a one-dimensional cross-section of the state space, whereas the diagonal threshold policy does not distinguish between busy Type A and busy Type B servers.

We evaluate the performance of these three policies on a system with ten Type A and Type B servers, each of which operate at a rate of $\mu = 1$. Without loss of generality, we assume $R_{HA} = 1$, and consider two reward regimes: one that favors Type H jobs ($R_{HB} = 0.6$, $R_L = 0.4$), and one that favors Type L jobs ($R_{HB} = 0.4$, $R_L = 0.6$). We also consider performance under varying levels of congestion, ranging from a severely underloaded system (in which $\lambda_H = \lambda_L = 0.5$) to a severely overloaded system (in which $\lambda_H = \lambda_L = 15$). In each of the MDP instances we consider, we solve for the optimal policy numerically using policy iteration. We find the optimal single threshold policy by computing the stationary distribution of the Markov chain induced by the policy with threshold t (from which long-run average reward can be easily calculated), and finding the threshold in the set $\{0, 1, \ldots, N_A\}$ that maximizes reward. We compute the optimal diagonal threshold policy in a similar fashion. Our findings are summarized in Figure 2.1.

When the system is lightly loaded, our heuristic policies perform optimally, as there is no need to reserve servers for Type H jobs when the system is rarely congested. As the system load increases, the optimal policy performs noticeably better, as routing decisions affect performance primarily during periods of congestion. The greedy policy, unsurprisingly, performs the poorest, particularly as arrival rates increase. In moderately-loaded systems, the myopic policy ap-

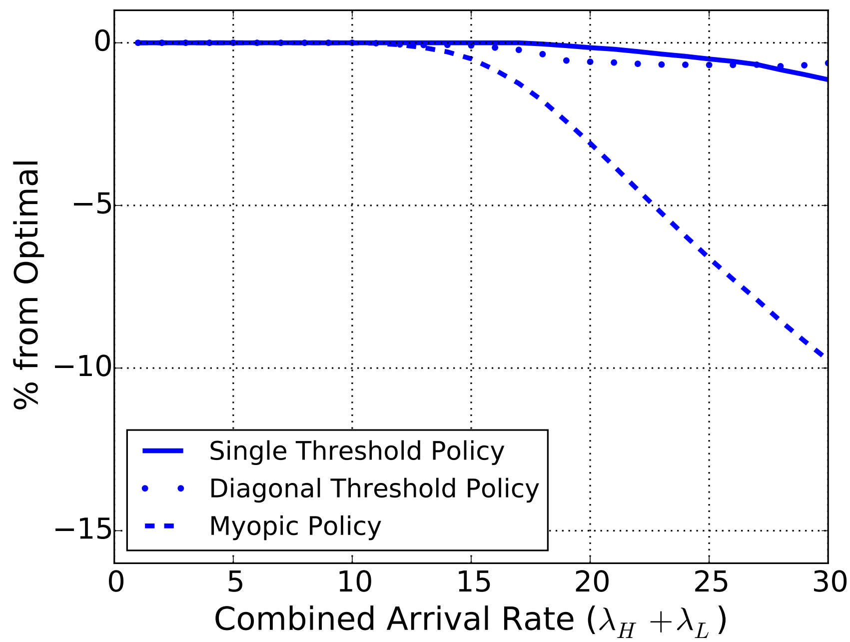

Figure 2.1: Optimality gap for three heuristic admission control policies, as a function of arrival rate, when $R_{HB} = 0.6$, $R_L = 0.4$

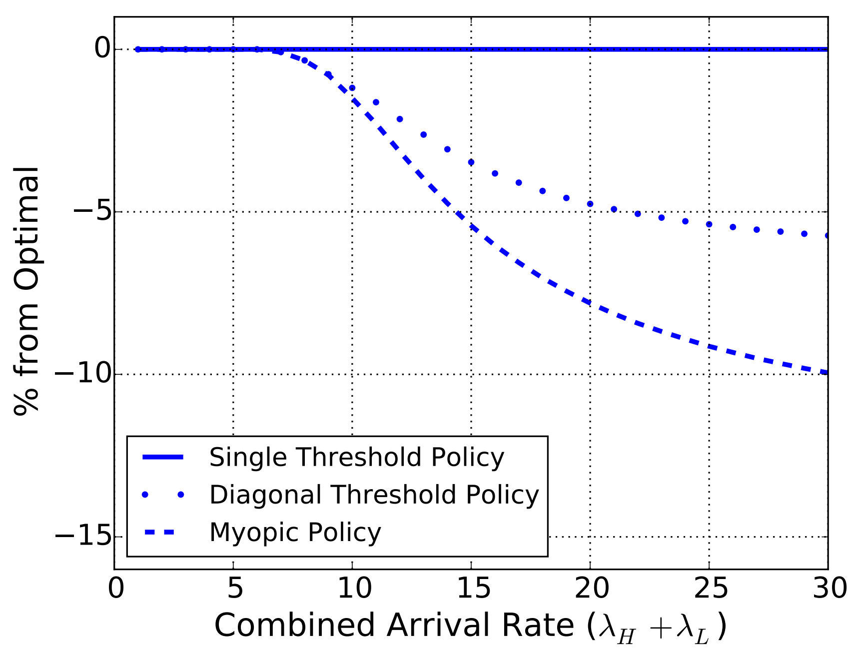

Figure 2.2: Optimality gap for three heuristic admission control policies, as a function of arrival rate, when $R_{HB} = 0.6$, $R_L = 0.4$

pears to struggle more when $R_L < R_{HB}$. This is also not surprising, given that setting $R_L < R_{HB}$ decreases the number of states in which it is optimal for the decision-maker to behave greedily.

The single threshold policy is optimal when $R_{HB} \leq R_L$, but this is a consequence of Proposition 2.4.3. When $R_{HB} > R_L$, it performs comparably to the optimal policy, but also struggles in more heavily loaded systems. The diagonal threshold policy performs quite well when $R_{HB} > R_L$, likely due to the fact that the structure of the policy closely mimics the monotone switching curve (with slope of at least -1) that was shown to be optimal in Theorem 2.5.3. However, it performs noticeably poorer when $R_{HB} < R_L$, as the policy becomes too conservative in terms of admitting Type L jobs.

Although we could consider a combination policy that uses either the single threshold or diagonal threshold policy, depending on the reward structure, the above experiments demonstrate that there is value in taking the heterogeneity of resources into account when making decisions in our system. We observe similar behavior when we consider different systems (that are less "balanced" in terms of the relative values of N_A and N_B, and of λ_H and λ_L), and perform sensitivity analyses, but we omit the corresponding results for brevity.

2.7 Conclusion

In this chapter, we considered the problem of routing and admission control in a service system featuring two classes of arriving jobs and two types of servers, where one arrival stream is uncontrolled, and our reward structure influences the desirability of each of the four possible routing decisions. We seek optimal policies under both the discounted long-run reward and long-run average reward criteria. We characterize the optimal policy by formulating the problem as a Markov decision process. When $R_L \leq R_{HB}$, we prove the existence of an optimal monotone switching curve policy with a slope of at least -1. When $R_L > R_{HB}$, we conjecture that threshold-type policies are optimal, but encounter difficulties because the value function, in general, is neither convex nor supermodular. These difficulties can be attributed directly to our assumption that decision-making with Type H jobs is forced. We instead prove a sufficient (but not necessary) condition for convexity, and show that when we restrict attention to the set of threshold-type policies, the optimal policy is monotone in our choice of reward structure. Numerical experiments suggest there is value in taking heterogeneity of servers into account when making decisions.

We propose two directions for future research. The first involves finding, for the case $R_L > R_{HB}$, stronger conditions under which threshold-type policies are optimal. This may entail strengthening the sufficient condition (2.11) presented in Section 2.5.2.1, via a more refined analysis of sample paths, as our upper bound on the probability of the collection of "bad" paths is somewhat loose. Alternatively, this may involve proving our conjecture that the value functions v_α and h satisfy the single-crossing property (2.10). A second direction for future research is to study a model in which jobs can be placed in buffers. This is especially relevant in emergency medical service systems (the original motivation for our model), as Type L emergency calls may be queued during periods of congestion until ambulances finish with more urgent calls. Incorporating buffers into our model would likely increase the dimension of the state space. Although it may still be possible to study such a system analytically, this may entail could involve numerically finding near-optimal heuristic policies.

CHAPTER 3

THE VEHICLE MIX DECISION IN EMERGENCY MEDICAL SERVICE SYSTEMS

3.1 Introduction

Ambulances in Emergency Medical Service (EMS) systems can be differentiated by the medical personnel on board, and thus, by the types of treatment they can provide. These personnel can be roughly divided into two groups: Emergency Medical Technicians (EMTs) and paramedics. EMTs use non-invasive procedures to maintain a patient's vital functions until more definitive medical care can be provided at a hospital, while paramedics are also trained to administer medicine and to perform more sophisticated medical procedures on scene. The latter may be necessary for stabilizing a patient prior to transport to a hospital, or for slowing the deterioration of the patient's condition en route. Thus, they may improve the likelihood of survival for patients suffering from certain life-threatening pathologies, such as cardiac arrest, myocardial infarction, and some forms of trauma. See, for instance, Bakalos et al. [7], Gold [36], Isenberg and Bissell [46], Jacobs et al. [47], McManus et al. [66], and Ryynänen et al. [77] for discussions of the effects of paramedic care on patient outcomes.

An ambulance staffed by at least one paramedic are referred to as an Advanced Life Support (ALS) unit (we use the terms "ambulance" and "unit" interchangeably), whereas an ambulance staffed solely by EMTs is known as a Basic Life Support (BLS) unit. While ALS ambulances are more expensive to operate (due to higher personnel and equipment costs) they are viewed as essential components of most EMS systems, as such systems are frequently evaluated by their ability to respond to life-threatening emergencies.

The selection of a *vehicle mix*— that is, a combination of ALS and BLS ambulances to deploy— has been a topic of some debate in the medical community, with the primary issue being whether BLS units should be included in a fleet at all. Proponents of all-ALS systems, such as Ornato et al. [72] and Wilson et al. [99], cite the risk of sending a BLS unit to a call requiring a paramedic, either due to errors in the dispatch process or to system congestion. An ALS unit may also need to be brought on scene before transport can begin, thus diverting system resources, and allowing the patient's condition to deteriorate. Proponents of *tiered systems*, such as Braun et al. [17], Clawson [26], Slovis et al. [85] and Stout et al. [89], argue that BLS ambulances enable EMS providers to operate larger fleets, leading to decreased response times. This may be more desirable, as they observe that a significant fraction of calls do not require an ALS response. Both sides of the debate allude to a trade-off inherent in the vehicle mix decision— that between improving a system's responsiveness and reducing the risk of inadequately responding to calls.

A number of secondary considerations may influence the decision-making process. For instance, Henderson [43] notes that dispatchers in a tiered system must also assess whether an ALS or BLS response is needed, complicating triage, and delaying the response to a call. As another example, Braun et al. [17] and Stout et al. [89] observe that EMS providers may have difficulty hiring and training the number of paramedics needed to operate an all-ALS system, as well as providing these paramedics with opportunities to hone and maintain their skills. While many of these issues are quantifiable, they have received less attention in the literature, and we do not consider them here.

In this paper, we contribute to this debate by constructing models under which quantitative comparisons among vehicle mixes can be made. To do so, we consider a system in which the EMS provider has a fixed annual operating budget B, and for which annual operating costs for ALS and BLS ambulances are C_A and C_B, respectively (where $C_A \geq C_B$). We define a vehicle mix to be an integer pair (N_A, N_B)— corresponding to the number of ALS and BLS ambulances deployed, respectively—for which $C_A N_A + C_B N_B \leq B$. Let $f(N_A, N_B)$ denote a scalar measure of system performance that is attained when the EMS provider deploys the vehicle mix (N_A, N_B).

Our notion of performance, which we formally describe in Chapter 3, is an aggregate measure of the system's ability to respond both to life-threatening and less urgent calls. Regardless of how we define it, performance depends on two closely-related decisions: *dispatching* decisions, the policy by which ambulances are assigned to emergency calls in real time, and *deployment* decisions, the base locations to which ambulances are stationed. These decisions, in turn, are affected by an EMS provider's choice of vehicle mix. Thus, it is reasonable for $f(N_A, N_B)$ to be the output of an optimization procedure. However, we do not intend for any "optimal solutions" we obtain to directly aid decision-making in practical contexts. Doing so would require tailoring our model to a specific EMS system, and we are more interested in obtaining general insights.

While we could construct a model that jointly optimizes with respect to both dispatching and deployment decisions, such a model may not be tractable. To obtain general insights, we require a procedure that can be quickly applied to large collections of problem instances. Thus, instead of studying one sophisticated model, we base our analysis upon two more stylized models. If we make the (admittedly large) assumption that the locations of arriving calls and ambulances can be ignored, we need only optimize with respect to dispatching decisions, and the resulting problem can be modeled as a Markov decision process (MDP); here, we use the model we formulated in Chapter 2. Alternatively, if we assume that the EMS provider operates under a fixed dispatching policy, the resulting problem of deployment can be modeled as an integer program (IP). Both models treat $(N_A,\, N_B)$ as input, and we take $f(N_A,\, N_B)$ in each case to be the objective function value associated with an optimal solution.

We contend that there is value in using both models to study the vehicle mix problem. An analysis conducted using the MDP model alone would leave open the question of whether taking geographical factors into account would lead to qualitatively different results. Furthermore, an analysis conducted solely using the IP model largely omits the real-time decision-making aspects of the problem, and raises similar concerns. Thus, we view our two models as complementary to one another. Numerical experiments indicate that our models yield similar qualitative insights, but different quantitative results. However, we demonstrate below that these discrepancies can be reconciled, allowing us to draw stronger conclusions than is possible with either model alone.

Our paper's primary finding is that there are rapidly diminishing marginal returns associated with biasing a fleet towards all-ALS, suggesting there is a wide range of vehicle mixes that perform comparably to (and, in some cases, outperform) an all-ALS fleet. This can be interpreted in several ways. For systems that already operate all-ALS fleets, our analysis suggests that there is not a compelling reason to switch to a different vehicle mix. For systems that operate a mixture of ALS and BLS ambulances, the benefit that can be attained by a conversion to an all-ALS fleet may not justify the costs involved. With respect to the vehicle mix debate, we contend that the most appropriate fleet for a given system should depend more heavily on the secondary considerations described above, or at least, that these considerations can be weighted more heavily in the decision-making process without significantly affecting performance.

The remainder of this chapter is organized as follows. Following a literature review in Section 3.2, we describe and formally construct in Section 3.3 our MDP model for ambulance dispatching in a tiered EMS system. We perform a computational study on this model in Section 3.4, which is based upon a large-scale EMS system loosely modeled after Toronto EMS. In Section 3.5, we formulate our IP model for ambulance deployment, and conduct a numerical study in Section 3.6 similar to that in Section 3.4. We conclude and discuss future research directions in Section 3.7.

3.2 Literature Review

There is a sizable body of literature relating to the use of operations research models to guide decision-making in EMS systems. We do not give a detailed overview here, and instead refer the reader to surveys, such as those by Brotcorne et al. [18], Goldberg [37], Green and Kolesar [38], Henderson [43], Ingolfsson [44], Mason [58], McLay [63], and Swersey [91]. We draw primarily from two streams of literature.

The first stream of literature we consider pertains to dynamic models, which are used to analyze real-time decisions faced by an EMS provider. Such models typically assume that ambulances have already been deployed to bases, and instead consider how to dispatch these ambulances to incoming calls, or alternatively, how to redeploy idle ambulances to improve coverage of future demand. The first such model is due to Jarvis [49], who constructs an MDP for dispatching in a small-scale EMS, while McLay and Mayorga [64] consider a variant in which calls can be misclassified. Work relating to ambulance redeployment was initiated through a series of papers by Berman [10, 11, 12], and Zhang [101] conducts a more refined analysis for the case of a single-ambulance fleet. The aforementioned models succumb to the curse of dimensionality, and so to analyze large-scale systems, Maxwell et al. [59] and Schmid [80] develop redeployment policies using approximate dynamic programming. While our model is less detailed than those cited above, it captures the essential features of a tiered EMS system, and is amenable to sensitivity analysis, and can thus be used to obtain quick insights. Our model can also be extended to include a fairly wide range of system dynamics, such as call queueing, without significantly increasing the size of the state space; see Appendix B.

The second stream of literature we consider relates to integer programming models for ambulance deployment, canonical examples of which include Toregas et al. [92], Church and ReVelle [25], and Daskin [27]. These models base their objective functions upon some measure of the system's responsiveness to emergency calls. This can be quantified via the proportion of emergency calls that survive to hospital discharge, as in Erkut et al. [30] and in Mayorga et al. [62], or more commonly, via the concept of coverage: the long-run average number of calls to which an ambulance can be dispatched within a given time threshold. These models have been extended to study the problem of deploying multiple types of emergency vehicles; see, for instance, Charnes and Storbeck [23], Mandell [57], and McLay [63]. Our IP model is most similar in spirit to that of Daskin [27], in that we use a similar notion of coverage in our objective function, and is also closely related to that of McLay [63], who considers calls of varying priority. However, our IP takes into account more of the nuances of dispatching in a tiered EMS system, by affording some flexibility in the dispatching policy. We also consider a more general notion of coverage that quantifies the system's ability to respond both to high-priority and low-priority calls.

Closely related is a stream of literature pertaining to descriptive models of EMS systems, which aims to develop accurate and detailed performance measures for a system operating under a given set of deployment decisions. Larson's [53] hypercube model and its variants, such as those by Jarvis [50] and Larson [54], are perhaps the most influential of this kind. Simulation has been widely used since Savas [78]; see, for instance, Henderson and Mason [42]. While descriptive models allow for more thorough comparisons between candidate deployment decisions, they are not as amenable to optimization.

There is a related body of literature on the flexible design of manufacturing and service systems. The seminal work in this area is due to Jordan and Graves [51], who observe that much of the benefit associated with a "fully flexible" system (which, in their case, represents the situation in which all plants in manufacturing system can produce every type of product) can be realized by a strategically-configured system with limited flexibility. A similar principle has been shown to hold for call centers Wallace and Whitt [97], as well as for more general queueing systems Gurumurthi and Benjaafar [39], Tsitsiklis and Xu [93]. Theoretical justifications thereof are provided in Akşin and Karaesmen [2] and Simchi-Levi and Wei [82]. In relating our work to this body of literature, we can define "flexibility" as the fraction of the budget that an EMS provider expends on ALS ambulances, but our paper's conclusions do not directly follow from this work. This is because servers in our model are geographically distributed, and their locations affect the system's ability to respond to incoming demand. The models we formulate provide a way to quantify these changes, and in turn, to study their effects on performance.

3.3 An MDP-Based Dispatching Model

3.3.1 Setup

Consider an EMS system operating N_A ALS and N_B BLS units. Incoming emergency calls are divided into two classes: urgent, *high-priority* calls for which the patient's life is potentially at risk, and less urgent *low-priority* calls. We assume that high-priority and low-priority calls arrive according to independent Poisson processes with rates λ_H and λ_L, respectively, and that they only require single-ambulance responses.

Service times are exponentially distributed with rate μ, independent of the priority of the call and of the type of ambulance dispatched. While we use the exponential distribution for tractability, our assumption of a single μ may be reasonable in large-scale systems, as on-scene treatment times are typically small relative to those for other components of the emergency response, such as travel and hospital drop-off times. Arrivals occurring when all ambulances are busy do not queue, but instead leave the system without receiving service. This is consistent with what occurs in practice, as EMS providers may redirect calls to external services, such as a neighboring EMS or the fire department, during periods of severe congestion, but we revisit this assumption in Section 3.3.3.

Dispatches to high-priority calls must be made whenever an ambulance is available, but a BLS unit can be sent instead if all ALS units are busy, so as to provide the patient with some level of medical care. In this case, we assume that the BLS unit can adequately treat the high-priority call, but that such a dispatch is undesirable, in a way that we clarify below; we also revisit this assumption in Section 3.3.3. Similarly, dispatches to low-priority calls must be made if a BLS ambulance is available, but if this is not the case, the dispatcher may either respond with an ALS unit (if one is available), or redirect the call to an external service (to reserve system resources for potential future high-priority calls).

Let R_{HA} and R_{HB} be the rewards associated with an ALS response and a BLS response to a high-priority call, respectively, and R_L be the reward for a response (of either type) to a low-priority call. We assume $R_{HA} = 1$ (without loss of generality) and $R_{HA} \geq \max\{R_{HB}, R_L\}$, but make no assumptions about the relative ordering of R_{HB} and R_L, as this may depend, for instance, on the skill gap between EMTs and paramedics, or on the incentives of the EMS in question. This is an unconventional modeling choice, but it yields an objective function that incorporates the system's responsiveness to both high-priority and low-priority calls. Typically, these two goals conflict, and we can adjust their relative weights by changing the reward structure. Our framework is also flexible. By letting R_{HA}, for instance, be the probability of patient survival when an ALS ambulance responds to a high-priority call (and defining R_{HB} and R_L similarly), we can mimic the reward structure adopted by McLay and Mayorga [64]. If an EMS provider is concerned solely with high-priority calls, we can let $R_{HA} = 1$, $R_{HB} = 0$, and $R_L = 0$. More generally, these rewards can represent a measure of the utility that the EMS provider derives from a successful dispatch. Nevertheless, identifying suitable choices for R_{HA}, R_{HB}, and R_L may be difficult, and we discuss this issue further in Section 3.4. We seek a policy that maximizes the long-run average reward collected by the system.

3.3.2 MDP Formulation

Our MDP model, as specified in Section 3.3.1, is nearly identical to the one we formulated in Section 2.3, but with the added requirement that BLS responses must be provided to low-priority calls whenever possible. To avoid repetition, we refer the reader to our MDP formulation in Section 2.3, and only specify how the model we build in this chapter differs.

This time, there is only a decision to make when all BLS ambulances are busy, and at least one ALS ambulances is avaialble, and so we define the action space to be $\mathcal{A} = \bigcup_{(i,\,j)\in\mathcal{S}} A(i,\,j)$, where

$$A(i,\,j) = \begin{cases} \{0,1\} & \text{if } i < N_A \text{ and } j = N_B, \\ \{0\} & \text{otherwise.} \end{cases}$$

As before, Action 1 dispatches an ALS unit to the next arriving low-priority call, while Action 0 redirects the call (or performs a dummy action) instead.

Our MDP is once again uniformizable in the spirit of Lippman [56] and Serfozo [81]; let $R((i,\,j),a)$ denote the expected reward collected over a single uniformized time period, given that the system begins the period in state (i,j), and the dispatcher takes action $a \in A(i,j)$. We have

$$R((i,\,j),a) = \begin{cases} \lambda_H R_{HA} + \lambda_L R_L & \text{if } i < N_A,\ j < N_B,\ a = 0, \\ \lambda_H R_{HB} + \lambda_L R_L & \text{if } i = N_A,\ j < N_B,\ a = 0, \\ \lambda_H R_{HA} & \text{if } i < N_A,\ j = N_B,\ a = 0, \\ \lambda_H R_{HA} + \lambda_L R_L & \text{if } i < N_A,\ j = N_B,\ a = 1, \\ 0 & \text{if } i = N_A,\ j = N_B,\ a = 0. \end{cases} \tag{3.1}$$

Let $P\big((i',\,j')\,|\,(i,\,j),\,a\big)$ denote the one-stage transition probabilities from state $(i,\,j)$ to state $(i',\,j')$ under action $a \in A(i,j)$. There are several cases to consider, as the system dynamics change slightly at the boundary of the state space. For brevity, we consider only the case when $0 < i < N_A$ and $j = N_B$, in which case

$$P\big((i',\, j') \,|\, (i,\, j),\, a\big) = \begin{cases} \lambda_H + \mathbf{I}(a=1)\lambda_L & \text{if } (i',\, j') = (i+1, j), \\ i\mu & \text{if } (i',\, j') = (i-1, j), \\ j\mu & \text{if } (i',\, j') = (i,\, j-1), \\ 1 - \lambda_H - \mathbf{I}(a=1)\lambda_L - (i+j)\mu & \text{if } (i',\, j') = (i,\, j). \end{cases}$$

The first transition corresponds to an arrival of a high-priority call (or a low-priority call, if the dispatcher performs Action 1), the second and third to service completions by ALS and BLS units, respectively, and the fourth to dummy transitions due to uniformization.

Defining, without loss of optimality, a policy to be a stationary, deterministic mapping $\pi : \mathcal{S} \to \{0, 1\}$ that assigns an action to every system state, we can find an optimal policy by solving for a set of optimality equations analogous to (2.4) in Section 2.3. We do this numerically using a policy algorithm: specifically, the one specified in Section 8.6.1 of Puterman [75].

3.3.3 Extensions to the MDP

Our MDP model can be modified to relax some of the above assumptions without dramatically increasing the size of the state space. In Appendix B.1, we formulate an extended MDP in which low-priority calls can be placed in queue during periods of congestion, and in which ALS units may need to be brought on scene to assist BLS first responses to high-priority calls. For our computational work in Section 3.4, we study only the base model we formulated above, as our extended model yields very similar numerical results; see Appendix B.1.6.

3.4 Computational Study of the MDP

We consider a hypothetical system loosely modelled after that operated in Toronto, Canada. We use the term "loosely" because we select inputs to our model using a dataset obtained from Toronto EMS, but we assume that the interarrival and service time distributions do not vary with time. Thus, the results we obtain below my be indicative of— but not necessarily predict— how the vehicle mixes we consider below would perform in practice.

3.4.1 Experimental Setup

Our dataset contains records of all ambulance dispatches occurring within the Greater Toronto Area between January 1, 2007 and December 31, 2008; we restrict our attention to calls originating from the City of Toronto. Emergency calls are divided into eight priority levels, two of which require a "lights and sirens" response. We treat calls belonging to these two priority levels as high-priority, and all other emergency calls as low-priority. Estimating arrival rates by taking long-run averages over the two-year period for calls originating within the City of Toronto, we obtain $\lambda_H = 8.1$ and $\lambda_L = 13.1$ calls per hour. We define service time as the length of the interval beginning with an ambulance dispatch and ending with the call being cleared (either on scene or following drop-off at a hospital). Because the mean service times for low-priority and high-priority calls do not differ substantially in the dataset (by less than 5%), our assumption of a single service rate for all calls is reasonable. We set $\mu = 3/4$ per hour, corresponding to a mean service time of 80 minutes.

To start our analysis, we let $R_{HA} = 1$, $R_{HB} = 0.5$, and $R_L = 0.6$; we investigate below the sensitivity of our findings to the latter two quantities. To estimate C_A and C_B, we assume that an ambulance requires three crews to operate 24 hours per day, that an ALS crew consists of two paramedics, that a BLS crew consists of 2 EMTs, and that ALS and BLS vehicles cost $110,000 and $100,000 to equip and operate annually, respectively. Assuming that annual salaries for paramedics and EMTs are $90,000 and $70,000 per year, respectively, we obtain C_A = 650,000 and C_B = 520,000, which we normalize to $C_A = 1.25$ and $C_B = 1$.

To determine our system's operating budget B, we assume that the average utilization of ambulances in our system is 0.4. Since $\lambda_H + \lambda_L = 21.2$ and $\mu = 0.75$, we would need approximately 70 ambulances to achieve the desired utilization. Assuming our system operates an all-ALS fleet (as in Toronto), we obtain $N_A = 70$ and $B = 87.5$. Restricting our attention to fleets that use as much of the budget as possible, we evaluate vehicle mixes in the set

$$\Gamma = \left\{ (N_A,\, N_B) : N_A \leq 70 \text{ and } N_B = \lfloor 87.5 - 1.25 N_A \rfloor \right\}. \tag{3.2}$$

3.4.2 Findings

Using the inputs specified above, we construct an MDP instance for each vehicle mix (N_A, N_B) in the set Γ specified in (3.2). We solve each instance numerically using policy iteration, and record the long-run average reward attained under the corresponding optimal policy. Plotting the resulting values with respect to N_A, we obtain Figure 3.1 below.

The curve in Figure 3.1 increases fairly steeply when N_A is small, indicating there is significant benefit associated with including ALS ambulances in a fleet. However, for larger values of N_A, the curve plateaus. This suggests that the marginal benefit associated with continuing to increase N_A diminishes

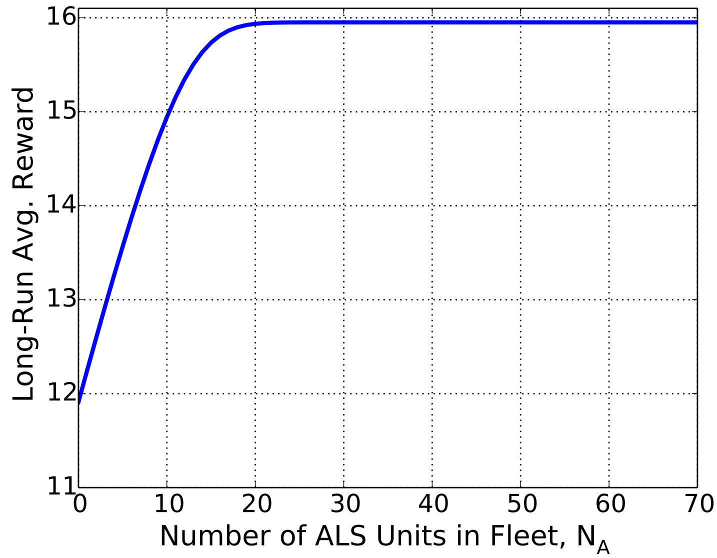

Figure 3.1: Long-run average reward attained by the optimal dispatching policy, as a function of N_A.

rapidly. One might suspect that the marginal benefits decrease too rapidly. Indeed, the long-run average reward attainable under any dispatching policy is upper bounded by $\lambda_H R_{HA} + \lambda_L R_L = 15.96$, but every vehicle mix for which $N_A \geq 20$ performs within 0.1% of this upper bound. We would not expect to find systems attaining this level of service in practice.

This behavior can be attributed to resource pooling. In formulating our MDP, we implicitly assumed that any ambulance can respond to any incoming call. In practice, most ambulances will be too far from a particular call to respond in time, and so only a subset of the fleet be effectively pooled. Thus, the MDP-based system can more readily respond to calls during periods of congestion; see, for instance, Whitt [98] for a formal discussion of this phenomenon.

To offset the effects of resource pooling, two alternatives include accelerating call arrivals, or reducing the number of ambulances in the system by decreasing the budget B; we adopt the latter approach here. Shrinking the fleet allows us to more easily see the effects of the vehicle mix decision under an optimal dispatching strategy, and in turn, draw insights from our MDP. This is because we subject the system to periods of congestion, which occur in practice, and represent the situations in which vehicle mix may have the greatest impact on system performance.

It is not at all obvious to what extent the fleet should be shrunk. We proceed by first selecting a service level, which we define as the long-run fraction of time in which at least one ambulance is available. If we assume that the EMS provider operates an all-ALS fleet, and adopts a dispatching policy in which low-priority calls are not redirected unless all ambulances are busy, then we can model the system as an $M/M/N_A/N_A$ queue, and find the blocking probability under a given budget using the Erlang loss formula. We consider five different budgets: 48.75, 46.25, 43.75, 41.25, and 37.50, which allow for all-ALS fleets of size 39, 37, 35, 33, and 30, respectively. These fleets can provide service levels of 0.990, 0.980, 0.965, 0.943, and 0.898, respectively, and operate under utilizations of 0.846, 0.808, 0.779, 0.749, and 0.717, respectively. Solving the corresponding MDP instances, and plotting the resulting five curves yields Figure 3.4.2 below.

While the curves in Figure 3.4.2 maintain the same basic structure observed in Figure 3.1, they taper off for larger values of N_A, suggesting that in certain situations, an all-ALS fleet may be detrimental. This is intuitive, as all-ALS systems tend to operate smaller fleets than their tiered counterparts, and a larger fleet may be preferable in heavily congested systems. To explore this idea further, we examine two related performance measures: the level of service provided to high-priority and to low-priority calls. Figure 3.3 plots these performance measures for the case $B = 43.75$.

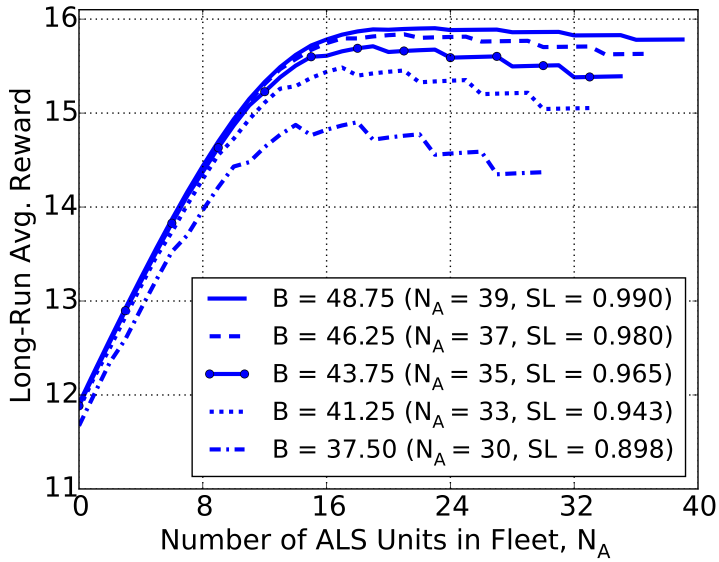

Figure 3.2: Long-run average reward as a function of vehicle mix, for several reduced values of the budget B.

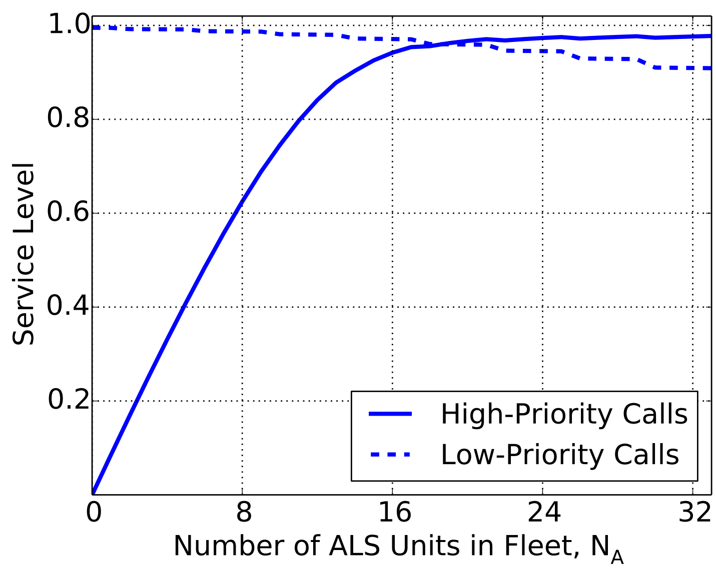

Figure 3.3: Long-run proportion of calls receiving an appropriate dispatch as a function of N_A, when $B = 41.25$.

As we expect, increasing N_A improves the system's responsiveness to high-priority calls, but worsens the system's responsiveness to low-priority calls. This results in a trade-off that is influenced by the relative importance of responding to high-priority and low-priority calls, as captured by the rewards R_{HA}, R_{HB}, and R_L. In this case, the marginal improvement attained by increasing N_A is eventually offset by the loss in ability to respond to low-priority calls. Because Figures 3.1 and 3.4.2 more succinctly describe this trade-off, we restrict our attention to plots of long-run average reward for the remainder of the paper.

3.4.3 Sensitivity Analysis

We next study the robustness of our findings by constructing curves analogous to those in Figures 3.1 and 3.4.2, but for problem instances in which we vary our model's input parameters. We also study the extended MDP model briefly described in Section 3.3.3. In the experiments that follow, we set $B = 43.75$.

We begin with a sensitivity analysis with respect to C_A, the annual cost of deploying an ALS unit. Figure 3.4 depicts five curves, each similar in spirit to that in Figure 3.1, but for values of C_A ranging from 1.05 to 1.45. Although we do not expect ALS units to cost 45% more than BLS units to operate in practice, we choose a wide range of values for illustrative purposes. Not surprisingly, operating an all-ALS fleet can be suboptimal when C_A is large. Perhaps the more interesting observation is that when C_A is small, an all-ALS fleet is not necessarily the obvious choice. As we observed in Figure 3.3, an EMS provider would consider a fleet with more ALS ambulances if there is a strong incentive to provide ALS responses to high-priority calls. However, a high level of service can be achieved without an all-ALS fleet, and the marginal benefit associated with increasing N_A shrinks rapidly— even when ALS ambulances are inexpensive.

Next, we consider the robustness of our findings to our rewards. Because we assume $R_{HA} = 1$, we need only perform a sensitivity analysis with respect

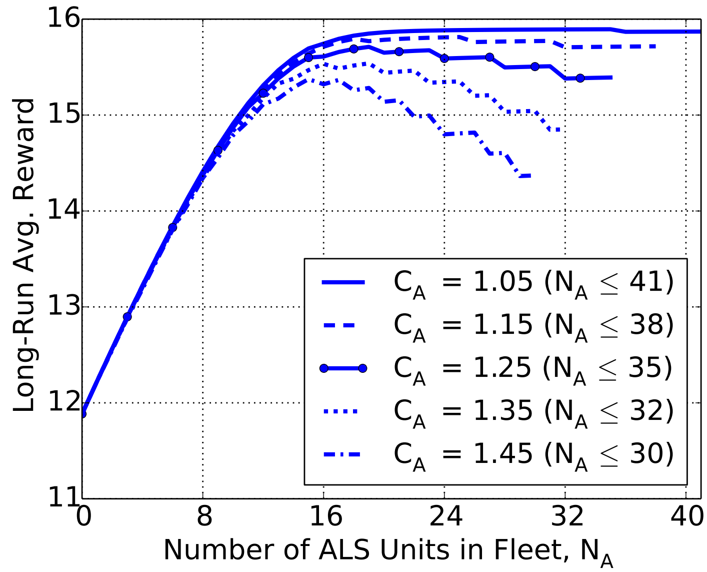

Figure 3.4: Long-run average reward as a function of vehicle mix, for various choices of C_A.

to R_{HB} and R_L. We restrict our attention to the all-ALS fleet $(35, 0)$ and the tiered system $(19, 20)$: the optimal vehicle mix from our "base case" analysis in Figure 3.4.2. We consider a collection of 625 MDP instances where R_{HB} and R_L take on one of 25 values in the set $\{0.02, 0.06, \ldots, 0.94, 0.98\}$. Plotting the relative difference in long-run average reward collected by the two systems for each instance, we obtain Figure 3.5.

In each problem instance, the tiered system outperforms the all-ALS fleet. This is unsurprising for instances where R_{HB} and R_L are close to 1 (as emergency calls become effectively indistinguishable, and the tiered system deploys eight more ambulances), but is counterintuitive for instances when R_{HB} and R_L are small. Even when performance is determined almost entirely by the system's responsiveness to high-priority calls, the all-ALS fleet does not outperform the tiered system. This suggests that the tiered system can adequately respond to these calls, which is consistent with Figure 3.3. We also observe

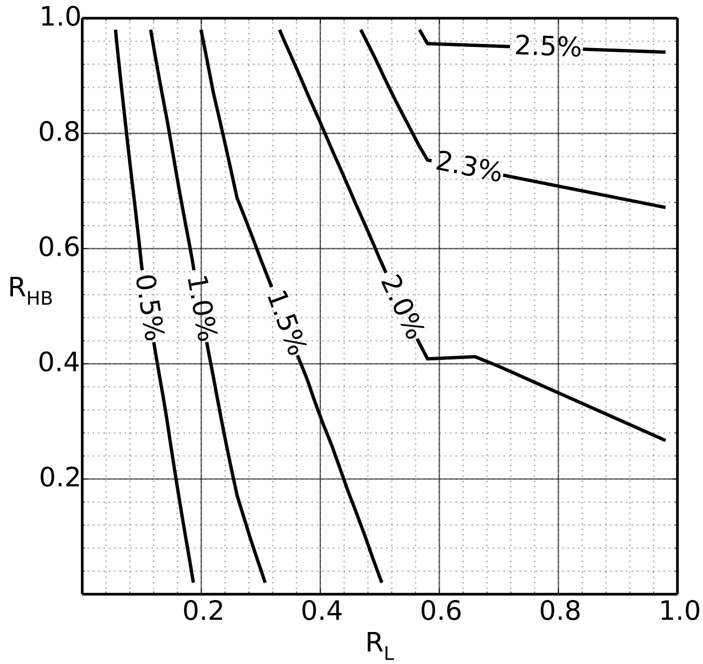

Figure 3.5: Contour plot of the relative difference between the long-run average reward collected by the tiered system $(19, 20)$ and that by the all-ALS fleet $(35, 0)$, for various values of R_{HB} and R_L.

that the two systems perform comparably in all of our problem instances; the largest observed difference was roughly 2.53%. We obtain similar results when comparing the all-ALS fleet to other tiered systems, although we omit the corresponding plots of these results in the interest of brevity. This suggests our findings are insensitive to our choice of rewards, which is encouraging, as selecting appropriate values of R_{HB} and R_L is difficult.

We conclude this section with a brief examination of the extended MDP model alluded to in Section 3.3.3. This model allows low-priority calls to be placed in a finite buffer, and considers the possibility that high-priority calls receiving a BLS first response may later receive treatment from an ALS ambulance. Plotting the "base case" curves for both the original and extended models, we obtain Figure 3.6 below. We truncate the curves at $N_A = 1$, as an all-BLS fleet cannot adequately treat high-priority calls in the extended model.

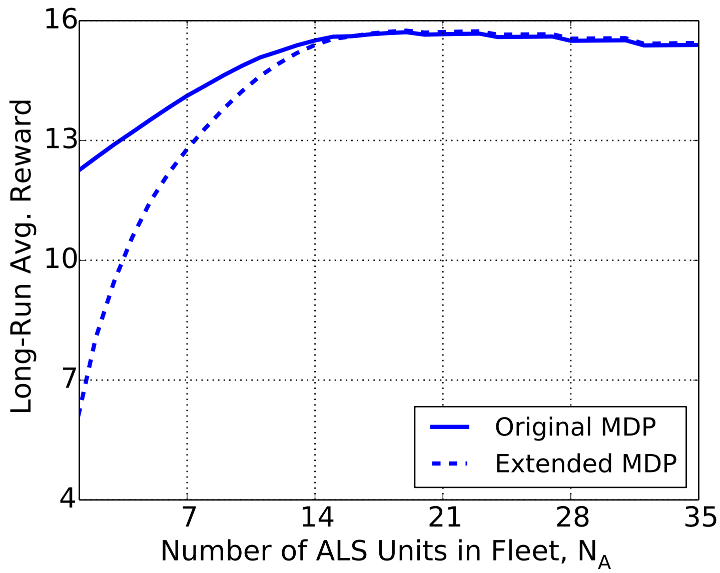

Figure 3.6: Long-run average reward as a function of vehicle mix, for the original and extended MDP models (see Section 3.3.3 and Appendix B.1).

Fleets operating too few ALS ambulances perform noticeably worse in the extended model. This is because high-priority calls will primarily be assigned to BLS ambulances, and many of these ambulances will be forced to idle until ALS units become free. However, for vehicle mixes under which high-priority demand can be adequately met, our two models are roughly in agreement. We conduct a more detailed analysis of the extended model in Appendix B.1.6.

3.4.4 Discussion

Taken together, our numerical experiments suggest that a wide range of tiered systems perform comparably to all-ALS fleets. The relatively small gap in long-run average reward that we observe between the two types of systems appears to be robust to changes in operating costs, arrival patterns, reward structure, and changes to system dynamics. In Section 3.6, we perform a similar set of numerical experiments, to determine whether we draw similar conclusions when we study our IP model, in which we mitigate the effects of resource pooling by geographically dispersing the fleet.

3.5 An IP-Based Deployment Model

3.5.1 Formulation

Consider an EMS system whose service area is represented by a connected graph $G = (N, E)$, where N is a set of demand nodes and E is a set of edges. High-priority and low-priority calls originate from node $i \in N$ at rates λ_i^H and λ_i^L, respectively. The EMS provider operates a fleet of N_A ALS and N_B BLS ambulances, which can be deployed to a set of base locations $\bar{N} \subseteq N$. For convenience we set $\bar{N} = N$, but this assumption can easily be relaxed.

We define t_{ij} as the travel time along the shortest path between nodes i and j. A call originating from node i can only be treated by an ambulance based at node j if $t_{ij} \leq T$, where T is a prespecified response time threshold. This gives rise to the *neighborhoods* $C_i = \{j \in N : t_{ij} \leq T\}$ — the set of bases from which an ambulance can promptly respond to a call originating from node i. If a ALS and b BLS units are deployed within C_i, we say that node i is *covered* by a ALS and b BLS units.

Let p_A denote the *busy probability* associated with each ALS ambulance: the long-run fraction of time that a given ALS unit is not available for dispatch; we define p_B similarly for BLS ambulances. We treat p_A and p_B as model inputs, and discuss our procedure for approximating these quantities in Section 3.5.2.

We assume that ambulances of the same type operate under the same utilization, and that all ambulances are busy independently of one another. Thus, if node i is covered by a ALS units and b BLS units, then $(p_A)^a\,(p_B)^b$ is the long-run proportion of time that the system cannot respond to calls originating from that node. Calls to which an ambulance cannot be immediately dispatched are redirected to an external service. We revisit these assumptions in Section 3.5.3.

As in the MDP model, we allow BLS units to be dispatched to high-priority calls, and ALS units to be dispatched to low-priority calls, but we do not require that ALS ambulances respond to every low-priority call arriving when all BLS ambulances are busy. Let ϕ denote the long-run proportion of low-priority calls receiving an ALS response in this situation. This quantity does not specify how decisions are made in real time, but provides a succinct measure of the system's willingness, in the long run, to dispatch ALS ambulances to low-priority calls. As with p_A and p_B, we assume ϕ to be given, and discuss in Section 3.5.2 how it can be approximated. Finally, we define rewards R_{HA}, R_{HB}, and R_L as before.

We build our objective function as follows. Suppose that node i is covered by a ALS and b BLS ambulances, and consider the level of coverage provided to low-priority calls at that node. With probability $1-(p_B)^b$, a BLS unit can be dispatched. Conditional on this not being the case, an ALS unit is available with probability $1-(p_A)^a$, but a dispatch only occurs with probability ϕ. Thus, the expected reward collected by the system from a single low-priority call is

$$R_L(a,b) = R_L[1-(p_B)^b + \phi\,(p_B)^b\,(1-(p_A)^a)]. \tag{3.3}$$

Similar reasoning yields that the system collects, in expectation, a reward

$$R_H(a,b) := R_{HA}\,(1-(p_A)^a) + R_{HB}\,(p_A)^a(1-(p_B)^b) \tag{3.4}$$

from a single high-priority call. This implies that the system obtains reward from node i at a rate $\lambda_i^H R_H(a,b) + \lambda_i^L R_L(a,b)$ per unit time. We want to deploy ambulances such that the sum of this quantity over all nodes in N is maximized. Let x_i^A and x_i^B be the number of ALS units and BLS units stationed at node $i \in N$, respectively, and let y_{iab} take on the value 1 if node $i \in N$ is covered by exactly a ALS units and b BLS units, and 0 otherwise. We thus obtain the formulation

$$\max \quad \sum_{i \in N} \lambda_i^H \sum_{a=0}^{N_A} \sum_{b=0}^{N_B} y_{iab} R_H(a,b) + \sum_{i \in N} \lambda_i^L \sum_{a=0}^{N_A} \sum_{b=0}^{N_B} y_{iab} R_L(a,b) \tag{IP}$$

$$\text{s.t.} \quad \sum_{i \in N} x_i^A \leq N_A \tag{3.5}$$

$$\sum_{i \in N} x_i^B \leq N_B \tag{3.6}$$

$$\sum_{a=0}^{N_A} a \sum_{b=0}^{N_B} y_{iab} \leq \sum_{j \in C_i} x_j^A \qquad \forall i \in N \tag{3.7}$$

$$\sum_{b=0}^{N_B} b \sum_{a=0}^{N_A} y_{iab} \leq \sum_{j \in C_i} x_j^B \qquad \forall i \in N \tag{3.8}$$

$$\sum_{a=0}^{N_A} \sum_{b=0}^{N_B} y_{iab} \leq 1 \qquad \forall i \in N \tag{3.9}$$

$$x_i^A \in \{0, 1, \ldots, N_A\} \qquad \forall i \in N \tag{3.10}$$

$$x_i^B \in \{0, 1, \ldots, N_B\} \qquad \forall i \in N \tag{3.11}$$

$$y_{iab} \in \{0, 1\} \qquad \forall i \in N, a, b \tag{3.12}$$

Constraints (3.5) and (3.6) state that at most N_A ALS units and N_B BLS units can be deployed. Constraints (3.7), (3.8), and (3.9) link the $x-$variables to the $y-$variables, by ensuring for each node i that if $\sum_{j \in C_i} x_j^A = \bar{a}$ and $\sum_{j \in C_i} x_j^B = \bar{b}$, then $y_{iab} = 1$ iff $a = \bar{a}$ and $b = \bar{b}$. This holds because the coefficients associated with the $y-$variables are strictly increasing in a and b. Finally, constraints (3.10), (3.11), and (3.12) restrict the our decision variables to integer values.

3.5.2 Approximating p_A, p_B, and ϕ

To approximate the quantities p_A, p_B, and ϕ, we use the inputs of our integer program to construct an instance of our MDP form Section 3.3.1, and examine the stationary distribution induced by an optimal dispatching policy. The MDP model takes as input the arrival rates λ_H and λ_L, and a service rate μ. We set $\lambda_H = \sum_i \lambda_i^H$ and $\lambda_L = \sum_i \lambda_i^L$, and for our computational work in Section 3.6, we again use $\mu = 0.75$. Let ν be the stationary distribution of the Markov chain induced by an optimal policy. We approximate the busy probabilities p_A and p_B using the average utilizations of ALS and BLS ambulances, respectively:

$$p_A \approx \frac{1}{N_A} \sum_{(i,j)\in\mathcal{S}} i\nu(i,\, j) \qquad \text{and} \qquad p_B \approx \frac{1}{N_B} \sum_{(i,j)\in\mathcal{S}} j\nu(i,\, j)\ . \tag{3.13}$$

To approximate ϕ, we could similarly use the quantity

$$\phi \approx \frac{\displaystyle\sum_{i=0}^{N_A-1} \nu(i,\, N_B) \cdot \mathrm{I}\big(A(i,\, N_B) = 1 \text{ under the optimal policy}\big)}{\displaystyle\sum_{i=0}^{N_A-1} \nu(i,\, N_B)}\ . \tag{3.14}$$

The denominator of (3.14) denotes the long-run fraction of time that all ALS ambulances are busy and at least one BLS ambulance is available. The numerator gives the long-run fraction of time in which the dispatcher sends an ALS ambulance to a low-priority call. Equations (3.13) and (3.14) are approximations, as p_A, p_B, and ϕ depend on how ambulances are located within the system. Nevertheless, these approximations allow us to capture, to an extent, the dependencies of these parameters on vehicle mix and on dispatching decisions.

3.5.3 Extensions to the IP

Perhaps the most significant assumptions we make in formulating (IP) are that calls do not queue, that ambulances are busy independently of one another, and that these probabilities do not depend on location. The former assumption can be relaxed by estimating p_A, p_B, and ϕ from the output of an MDP that includes call queueing, such as that in Appendix B.1, and by modifying the objective function to take queued calls into account.

The independence assumption can be relaxed using correction factors, which adjust the probabilities obtained under this assumption by a multiplicative constant to account for dependence. This idea is due to Larson [54], and has been used, for instance, in Ingolfsson et al. [45] and McLay [63] to formulate IP-based models for ambulance deployment. In Appendix B.2, we formulate an extended IP model that incorporates correction factors to our objective function.

Relaxing the assumption of location-independent busy probabilities is difficult, as the utilization of a single ambulance depends upon how all other ambulances are deployed, resulting in nonlinear interactions. Budge et al. [21] develop an iterative procedure for the case when ambulances have already been deployed. Ingolfsson et al. [45] alternate between solving an integer program for a given set of utilizations, and using the resulting optimal solution to compute updated values, also in an iterative fashion. However, these approaches are computationally intensive.

For our computational work in Section 3.6, we study the IP model we formulated above, as the extended IP model yields qualitatively similar results; see Online Appendix B.2.3.

3.6 Computational Study of the IP

3.6.1 Experimental Setup

We base our computational experiments upon the same hypothetical EMS considered in Section 3.4. To construct our graph G, we bound the service area within a rectangular region. Using latitude and longitude information included with call records in the dataset, we find that a 26 × 19 mile region suffices. We divide this region into a 52×38 grid of 0.5 × 0.5 mile cells (demand nodes).

To compute call arrival rates associated with each node, we map each call to a cell in the grid, and take a long-run average over the two-year period for which we have data. We define the distance between two nodes as the Manhattan distance between the centers of their corresponding cells. For each node i, we define the neighborhood C_i as the set of bases from which an ambulance can be brought on scene within 9 minutes. This response interval includes the time taken by the dispatcher to assign an ambulance to a call, and by the corresponding crew to prepare for travel to the scene. Assuming this process takes two minutes, and that ambulances travel at 30 miles per hour, C_i contains all nodes lying no more than 3.5 miles away from node i. As before, we set $R_{HA} = 1$, $R_{HB} = 0.5$, $R_L = 0.6$, $C_A = 1.25$, and $C_B = 1$, and again evaluate vehicle mixes in the set $\Gamma = \{(N_A,\, N_B) : N_A \leq 70$ and $N_B = \lfloor 87.5 - 1.25 N_A \rfloor\}$.

3.6.2 Findings

For each vehicle mix in the set Γ, we solve the corresponding instance of (IP) to within 1% of optimality, and store the resulting objective function value. Although we introduce some error by not finding the optimal integer solution, the impact on our overall findings is negligible. To decrease computation times, we remove any decision variables y_{iab} for which either $a \geq 30$ or $b \geq 30$. Thus, we consider any demand node that is covered by more than 30 ALS or BLS ambulances to be covered by exactly 30 ambulances of the corresponding type instead. In doing so, we do not render infeasible any solutions that cover a node with more than 30 units, but we disregard the contributions of these excess units to the objective function. We thereby underestimate coverage, but not to a significant degree, as p_A^{30} and p_B^{30} are very small for reasonable choices of p_A and p_B. Since a and b can be as large as 70 and 87, respectively, this dramatically reduces the number of decision variables in the model.

We use the procedure from Section 3.5.2 to approximate p_A, p_B, and ϕ, but find that for each of the problem instances we consider above that $\phi = 1.0$. That is, when we solve the MDP instances corresponding to our IP instances, the optimal policy always provides ALS responses to low-priority calls when BLS ambulances are busy. These anomalous results can again be attributed to resource pooling. While a dispatcher may sometimes want to reserve ALS ambulances for future high-priority calls, this is not the case in the MDP model, as the system rarely becomes congested enough for the above policy to have detrimental effects. To identify more suitable values for ϕ, we instead solve modified MDP instances in which we accelerate arrivals using a scaling factor s. The resulting optimal policies may more closely reflect decisions made in practice, as they are derived from more heavily congested systems. Applying (3.14) to these policies, we obtain new values for ϕ, which we then use to construct modified IP instances in which all other inputs (arrival rates, p_A, and p_B) are kept fixed.

It is not at all clear how arrivals should be scaled, and so we begin our numerical study by examining the sensitivity of our results to s. Figure 3.7 below plots long-run average reward with respect to vehicle mix for values of s ranging from 1.50 to 2.50. The resulting curves are analogous to that in Figure 3.1 of Section 3.4, which we also include below.

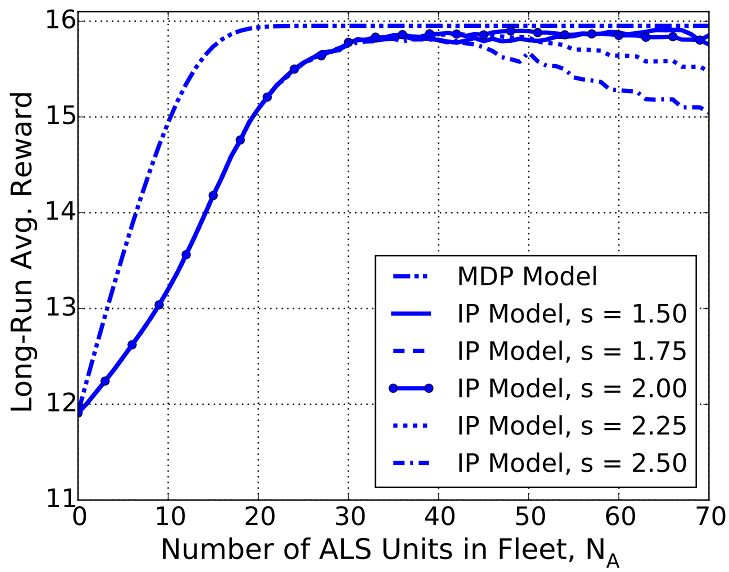

Figure 3.7: Long-run average reward attained by a near-optimal deployment policy for various arrival rate scaling factors s, overlaid with the analogous curve from the MDP model.

All of the curves in Figure 3.7 exhibit the same trend: a relatively sharp increase for small values of N_A, followed by rapidly diminishing marginal returns. The curves obtained from the IP model almost completely overlap when N_A is less than about 30, as ALS ambulances in the corresponding systems are overwhelmed by high-priority calls. Scaling factors have very little effect on the optimal policies of the corresponding MDPs, and so we obtain very similar values for ϕ. As we move towards an all-ALS fleet, we observe a decline in performance for more extreme values of s. This is because in a heavily-loaded system, the dispatcher may prefer to reserve ALS ambulances for high-priority calls (which are more likely to occur when s is large), resulting in smaller values of ϕ. However, this translates into an overly conservative dispatching policy, and thus lower performance, within the context of the IP model. Nonetheless, it is encouraging that our findings are not particularly sensitive to the dispatching policy that we employ in the IP model, as captured by the parameter ϕ. In the experiments that follow, we restrict our attention to the case when $s = 2$.

Another observation we draw from Figure 3.7 is that our models yield very different numerical results, particularly for smaller values of N_A. This may again be due to the effects of resource pooling; in the MDP model, any ambulance can respond to any call, whereas in the IP model, this is not the case. To test this hypothesis, we consider a collection of IP instances in which we artificially magnify the effects of resource pooling, to see whether we obtain quantitative results that are more consistent with those from the MDP.

In our IP model, the degree of resource pooling is captured by a single input parameter: the response time threshold T. Increasing this threshold increases the number of ambulances that can cover a given demand node. By letting T grow sufficiently large (in this case, to 45 minutes), we obtain a system with complete resource pooling, as in the MDP model. Thus, we proceed by taking the 71 problem instances generated above, and constructing modified instances in which T is increased, but all other parameters are unchanged. Figure 3.8 illustrates the curves we obtain by setting T to 18 and 45 minutes, respectively.

As we increase T, the gap between the two curves narrows, but interestingly, for larger values of T, the IP model yields objective values larger than those obtained by the MDP model, particularly when N_A is small. This can partly be attributed to our assumption that ambulances are busy independently of one another. In particular, during periods of congestion, this assumption could lead to optimistic estimates of ambulance availability. Indeed, our extended IP model,

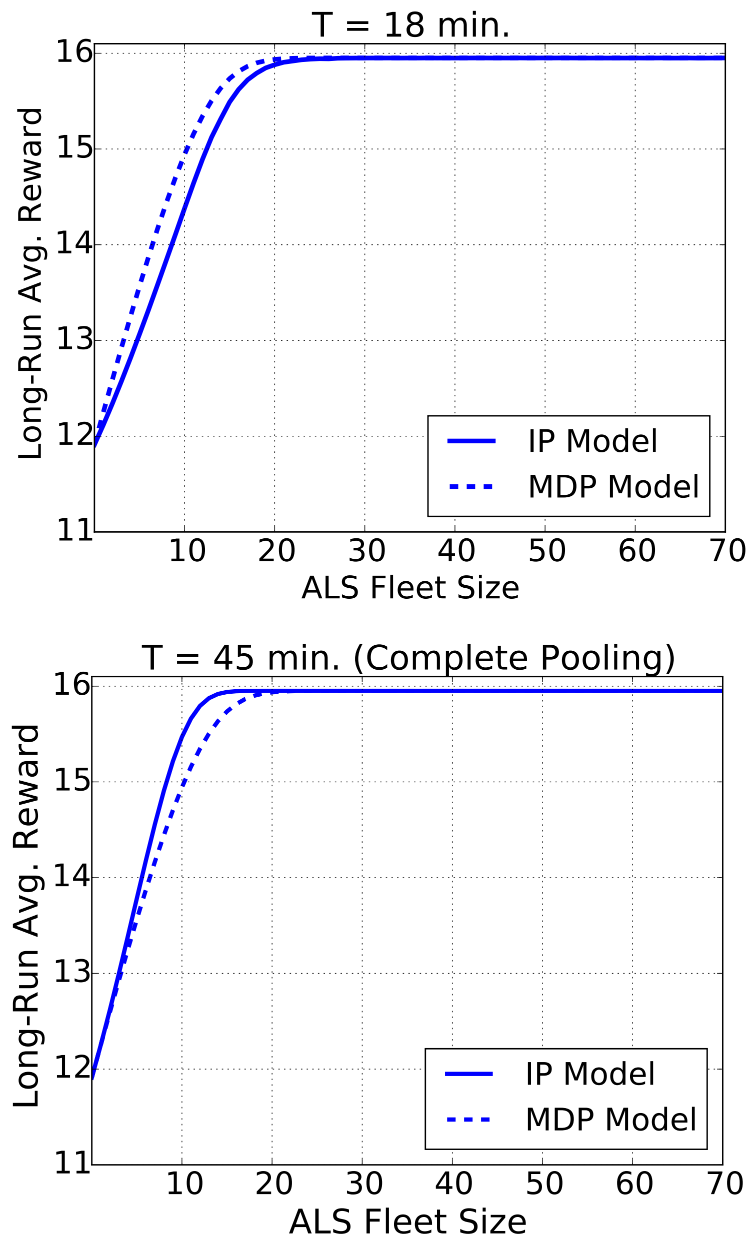

Figure 3.8: Long-run average reward attained under the IP and MDP models, for two choices of the response time threshold T.

which employs correction factors, improves the fit slightly; see Appendix B.2.3.

The above experiments suggest that our two models yield similar qualitative results, and that much of the observed discrepancy in our quantitative results can be accounted for by resource pooling. Thus, we contend that our models are generally in agreement, and that both support our claim of rapidly diminishing marginal returns associated with biasing a fleet towards all-ALS.

3.6.3 Sensitivity Analysis

We begin with a sensitivity analysis with respect to C_A that similar to that in Section 3.4; Figure 3.9 below is analogous to Figure 3.4. We again observe the

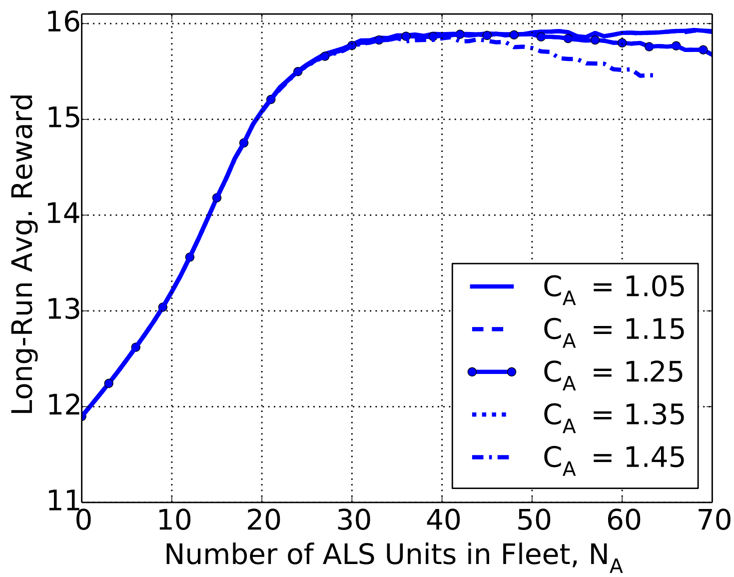

Figure 3.9: Long-run average reward as a function of vehicle mix for several choices of C_A.

same general trends. This theme recurs if we perform sensitivity analyses with respect to rewards or arrival rates, suggesting that the agreement between our two models is also robust to changes to our input parameters.

We conclude this section with a sensitivity analysis with respect to ambulance travel speeds. Because ambulances must arrive on scene within a response time threshold T, changes to these speeds can affect the distance an ambulance can travel to cover a call. While we examined this to an extent in Figure 3.8, we consider a more realistic range of values here. Figure 3.10 below illustrates the curves we obtain for speeds ranging from 21.43 to 38.57 mph. We choose these values so that the resulting ambulance coverage radii, in miles, are integer multiples of 0.5. (Recall that we discretized the service area into 0.5×0.5 mile squares.) We observe that the profile of the curves does not change dramatically

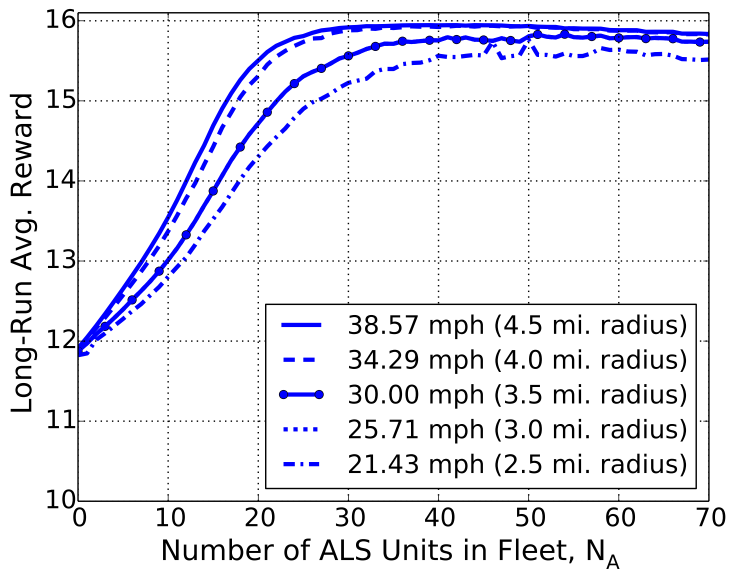

Figure 3.10: Long-run average reward as a function of vehicle mix for several choices of ambulance travel speeds.

with the speed at which ambulances travel. This is encouraging, as travel speed can depend on the time of day, as well as on geographical factors. While we assume away these effects in our IP, Figure 3.10 suggests that doing so does not substantially change our results.

3.7 Conclusion

In this chapter, we studied the effects of the vehicle mix decision on the performance of an EMS system. Inherent in this decision is a trade-off between improving the quality of service provided to high-priority calls, and increasing the size of the fleet. We analyzed this trade-off via two complementary optimization models of decision making in a tiered EMS system. Specifically, we formulated a Markov decision process that examined the operational problem of ambulance dispatching, as well as an integer program that modeled the tactical problem of deploying ambulances within a geographical region. To aid decision-making, we assigned rewards for individual responses to emergency calls, which formed the basis of a performance measure allowing quantitative comparisons between vehicle mixes to be made. Numerical experiments suggest that while ALS ambulances are essential components of EMS fleets, a wide range of mixed fleets can perform comparably to (or occasionally, outperform) all-ALS fleets. This was corroborated by both of our models, and appears to be robust to reasonable changes to the values of our input parameters. As a consequence, when constructing an ambulance fleet, secondary considerations, such as those described in the introduction, can be weighed into the decision-making process without significantly decreasing performance. Mathematically modeling these considerations, and their effects on the performance of a given vehicle mix, may be a direction for future research.

While our focus in this chapter was to construct models that can be used to quickly obtain basic insights, a natural question to ask is what additional insights can be drawn from a more sophisticated model. Another possible direction for future research would be to consider a model for dispatching in a tiered EMS system that takes the locations of ambulances into account. Such a model would have a considerably larger state space, but could be approached using Approximate Dynamic Programming (ADP), as in Maxwell et al. [59] or in Schmid [80]. This model could incorporate dynamics such as time-varying call arrival rates, multiple call priorities, and patient transport to hospitals.

CHAPTER 4

A BOUND ON THE PERFORMANCE OF OPTIMAL REDEPLOYMENT POLICIES IN LOSS SYSTEMS

4.1 Introduction

Ambulance redeployment is the practice of strategically repositioning idle ambulances (either from home bases or following service completions) to improve the system's responsiveness to future calls [88]. It has become more prevalent in recent years, as Emergency Medical Serivce (EMS) providers search for ways to improve performance in an increasingly challenging operating environment.

The problem of finding an optimal redeployment policy can be formulated as a stochastic dynamic program, and has been done, for instance, by Berman [10, 11, 12] and McLay and Mayorga [64]. However, such a model would need to include the location and status of every ambulance in the state space. This approach is impractical in large-scale systems, as it succumbs to the curse of dimensionality, and would likely yield a policy that is too unwieldy to implement. As a result, a number of methods for obtaining good policies have been proposed, and techniques for doing so include using approximate dynamic programming [59, 80], solving integer programs in real time [35, 69, 70, 101], developing heuristics [6, 48, 94, 95], and building compliance tables, which specify where ambulances should be positioned, given coarse information about the state of the system [3, 34, 90, 96].

However, an important question to ask is whether any of these heuristic policies can be shown to be near-optimal. To address this concern, methods have been proposed to construct bounds on the performance of a fairly general class of redeployment policies; see Maxwell et al. [61] and Yue et al. [100]. These bounds are typically not tight, but they may still be informative. For instance, they allow a municipality to determine if performance targets are achievable by a proposed ambulance schedule, which was the primary motivation for Maxwell et al. [61]. Alternatively, bounds provide an indication of the increase in performance that can be realized by implementing redeployment policies. Because these policies remain controversial, due to the heavier burden they place on emergency medical staff [15], bounds may help in determining whether such an operational change is worthwhile.

Computational experiments suggest that the bound in Maxwell et al. [61] is tighter than that in Yue et al. [100], but the former bound assumes that arriving calls can queue when all ambulances are busy. Many EMS systems make use of external resources (such as police and fire vehicles, or ambulances from other municipalities) during "red alert" situations. This is the case in some large-scale systems [68, 74, 83, 84], as well as many small- to medium-scale systems that pool resources with neighboring EMS providers. When this is the case, it may be more reasonable to model the EMS as a *loss system*. This is a subtle difference, but presents a significant technical challenge, as the bound in Maxwell et al. [61] is not valid for such systems.

In this paper, we propose a new upper bound that is provably valid in loss systems. Although we draw upon ideas in Maxwell et al. [61] to do so, significant new ideas are required. In contrast to Maxwell et al. [61], and like in Yue et al. [100], we obtain an upper bound via optimization along each sample path, by constructing an integer program for which the decisions made by almost any redeployment policy correspond to a feasible solution. Since we model the EMS as a loss system, we cannot directly compare our bound to that by Maxwell et al. [61], but computational experiments suggest that our upper bound is tighter than one originating from a perfect information relaxation of the problem in the style of Yue et al. [100].

The remainder of this chapter is organized as follows. In Section 4.2, we explicitly formulate our model of the ambulance redeployment problem. In Section 4.3, we review the concepts from Maxwell et al. [61] that we use in developing our upper bound, and demonstrate by counterexample that their bound is not valid for loss systems. We introduce our modification to their upper bound in Section 4.4, and prove its validity. Finally, we numerically evaluate our upper bound in Section 4.5, and conclude in Section 4.6.

4.2 Problem Formulation

Consider an EMS system operating A ambulances, whose service area is represented by a graph $G = (N, E)$, where N is a set of demand nodes, and E is a set of edges. Ambulances can be deployed to a set of base locations $B \subseteq N$. Let $t(i, j)$ denote the travel time between demand node i and base j; we assume this quantity is deterministic and time-stationary. Calls arrive according to an arbitrary stochastic process that we assume is independent of the system state, as well as of any decisions made by the EMS provider. We further assume that the locations of these calls are i.i.d. random variables; let p_i denote the probability that an arriving call originates from node $i \in N$.

Ambulances must immediately be dispatched to arriving calls whenever one is available, but we place no constraints on which ambulance must be sent. We say that a call receives a *timely response* if an ambulance reaches the call's location within a response threshold t_{resp} (typically, 9 minutes in practical applications). The response interval typically includes the time required for the dispatcher to triage and assign an ambulance to the call, and for the crew in the responding ambulance to prepare for travel to the scene, which we refer to as the *chute time*. If a call receives a response (regardless of whether or not the response is timely), the ambulance travels to the call's location, and spends a random amount of time on scene treating the patient; let F denote the distribution function of this random variable. We assume, for convenience, that calls do not require transport to a hospital following treatment on scene (and so service completes at the call's location), but this assumption can be easily relaxed, as we briefly explain later. Calls arriving when all ambulances are busy leave the system.

At any time, the EMS provider may redeploy idle ambulances, so as to increase the likelihood of timely responses to future call arrivals, and may do so in an arbitrary fashion. A redeployment move entails relocating an ambulance from a node i to a base j; we need not specify the process by which this move occurs, provided that its location at any point is a node in the graph. Specifically, our upper bound is valid if we assume that the ambulance idles at node i for $t(i,\, j)$ time units, then immediately relocates to base j, or makes a sequence of smaller "jumps" within the graph, or any variant thereof, but we do not allow for "continuous" motion within the graph. The only restriction we place on the redeployment policies we consider is that they do not anticipate— that is, they do not use information about future call arrivals.

We evaluate the performance of a policy via the expected number of calls receiving timely responses over a finite horizon $[0,\, T]$, which, by the strong law of large numbers, is roughly equivalent (for large enough T) to the objective of maximizing the long-run average proportion of timely responses to calls. For convenience, we will sometimes refer to the former quantity as the total reward collected by the system.

4.3 Preliminaries

We begin by reviewing the tools from Maxwell et al. [61] that we use in developing our bound.

4.3.1 A Bounding Reward Function

Let $v : \{0, 1, \ldots, A\} \to [0, 1]$ be a nondecreasing function for which $v(a)$ is an upper bound on the probability that a call arriving when a ambulances are available receives a timely response. In practice, this probability depends both upon the locations of the a ambulances, and upon the call's arrival time, but $v(a)$ is a valid upper bound regardless of these factors. We let $v(0) = 0$ in our model, as calls leave the system when no ambulances are available. (This is in contrast to Maxwell et al. [61], who let $v(0) = v(1)$, as a call arriving when $a = 0$ may still receive a timely response if a busy ambulance is about to become available.)

To see how this function can be used to construct an upper bound, consider an alternate EMS system in which the decision maker collects a reward $v(a)$ when a call arrives while a ambulances are available (instead of a reward of either 0 or 1, depending on whether the response is timely). Maxwell et al. [61] show that each redeployment policy π, in expectation, performs at least as well in the alternate system as in reality. More formally, let R^π and V^π be the reward collected by policy π in the original and alternate systems, respectively.

Lemma 4.3.1. *For every policy π, $E[R^\pi] \leq E[V^\pi]$.*

See Maxwell et al. [61] for a proof. They also demonstrate that for a given a, $v(a)$ can be obtained by solving an instance of the Maximal Covering Location Problem (MCLP), which is due to Church and ReVelle [25]. This entails solving an integer program which locates a ambulances on a graph so as to maximize the proportion of calls "covered" by an ambulance, where a demand node is considered to be covered if the closest ambulance can provide a timely response.

4.3.2 Bounding Service Time Distributions

Next, let $\{G_a : 0 \leq a \leq A\}$ be a collection of distribution functions, where $G_a : [0, \infty) \to [0, 1]$ is a stochastic lower bound on the service time distribution associated with a call arriving when a ambulances are available (regardless of their locations). Here, we define service time to be the sum of the chute time, the travel time to the call's location, and the on-scene treatment time. (In our case, G_0 is a Dirac delta function centered at 0, but we define $G_0 := G_1$ to be consistent with Maxwell et al. [61].) Assume the distributions are stochastically nonincreasing in a— that is, for all $x \geq 0$, we have that $G_0(x) \leq G_1(x) \leq \cdots \leq G_A(x)$. Maxwell et al. [61] observe that given x and a, $G_a(x)$ can be computed by solving a p-median problem, which locates a ambulances within a graph so as to maximize the probability that the service time associated with the next call is at most x time units. They formulate this problem as an integer linear program.

Remark: Maxwell et al. [61] show the p-median problem can be formulated so that the distributions $\{G_a : 0 \leq a \leq A\}$ are stochastic lower bounds even when patients may require transport to hospitals. Let $H \subseteq N$ be the set of hospitals in the system, q_{ih} be the probability that a call arriving to node i is transported to hospital h (we may have $\sum_{h \in H} q_{ih} < 1$ for some i), and F_h be the distribution of the patient transfer time at hospital h. Defining service time to also include any time needed to transport the patient to the hospital, as well as to complete the transfer, Maxwell et al. [61] numerically convolve F with the F_h to obtain the service time distribution associated with an arrival to node i. They use these distributions to obtain inputs for the p-median problem.

4.3.3 A Counterexample

To obtain an upper bound that is valid for any redeployment policy π, Maxwell et al. [61] construct a queueing system, which they call their *bounding system*, with A servers (ambulances) in which jobs (calls) arrive according to the same stochastic process as in the original EMS system. Rewards and service times are state-dependent, in that a call arriving when a ambulances are idle generates reward $v(a)$, and spends a random amount of time in service that is governed by the distribution function G_a. Calls arriving when all ambulances are busy are placed in an infinite buffer, and served in a first-in, first-out fashion.

They demonstrate using a coupling argument that the reward collected by the bounding system, in expectation, is at least as large as the performance attained by an optimal policy in the original EMS system. Suppose that C denotes the number of calls arriving in the interval $[0, T]$, that they arrive at times $T_1, \ldots, T_C$, and that A_c^π represents the number of ambulances available at time T_c under policy π. Let $\tilde{A}_c$ be analogous to A_c^π for the bounding system. The validity of their upper bound follows from Lemma 4.3.1, and from showing that $\tilde{A}_c \geq A_c^\pi$ holds pathwise for all π and c.

In loss systems, this relation may not hold pathwise. There may be times at which the bounding system can respond to a call, but the original EMS system may be forced to turn away the call, leading to work that only the bounding system is forced to process. Thus, there may exist sample paths ω on which $\tilde{A}_c(\omega) < A_c^{\pi}(\omega)$ for some c. This suggests there may be loss systems for which the bound is not valid, and we demonstrate this by counterexample:

Example 4.3.2. *Consider a graph with two nodes, with $p_1 = p_2 = 0.5$, connected by an edge which takes 1 minute to traverse. In this system, $A = 2$, $T = 60$, and at the start of the horizon, one ambulance is stationed at each node. The arrival process is deterministic, with calls arriving at 8, 16, 24, 29, 38, and 40 minutes after the start of the horizon. The response time threshold is set to 0 minutes, and so timely responses can only be provided from ambulances situated at the call's location. On-scene treatment times are also deterministic, and last 10 minutes.*

Suppose the EMS provider dispatches the closest available ambulance to incoming calls, and never redeploys ambulances following service completions. Then a direct calculation shows that this policy, in expectation, provides 3.25 timely responses. Computing the bound from Maxwell et al. [61], we find that $v(0) = v(1) = 0.5$ and $v(2) = 1$, that G_0 and G_1 correspond to random variables that take on the values 10 and 11 with probability 0.5, and that G_2 is a Dirac delta function centered at 10. Using simulation or analytical methods, we find their method yields an upper bound of 3. Since we have found a policy that attains a strictly higher objective value, the upper bound is not valid.

4.4 A Modified Upper Bound

4.4.1 Construction

To develop an upper bound that is valid for loss systems, we construct a random variable Z such that when it is jointly defined with V^{π} on the same probability space, Z dominates V^{π} pathwise under any policy π. We begin by specifying the sample space Ω associated with our model. Random quantities in our model include C, the call arrival times $T_1, \ldots, T_C$, the locations $L_1, \ldots, L_C$ to which these calls arrive, and the corresponding on-scene treatment times. We characterize the randomness associated with the latter by tagging the c^{th} arriving call with a Uniform$(0, 1)$ random variable U_c, from which the corresponding on-scene treatment time can be obtained by computing $F^{-1}(U_c)$. While this characterization is somewhat indirect, it facilitates our construction of Z. A sample path thus consists of realizations of C, the total number of arriving calls, and of T_c, L_c, and U_c for each call c.

As in Maxwell et al. [61], we also construct a bounding system with A ambulances, and in which calls arrive according to the same stochastic process. Rewards and service times are state-dependent in the same way as specified in Section 4.3.3. Contrary to Maxwell et al. [61], calls arriving when all ambulances are busy are lost, the full sample path is revealed to the decision-maker at the beginning of the horizon, and calls can be rejected from the system, even when ambulances are available. The latter assumption may appear counterintuitive, but it may be preferable to respond to fewer calls, and to collect the larger rewards associated with admitting these calls in "higher" system states. The bounding decision-maker seeks a "policy"— that is, a subset of calls to admit into the system— that maximizes total reward. We define $Z(\omega)$ to be the reward collected by an optimal policy in the bounding system under sample path ω.

Given such an ω, we can compute $Z(\omega)$ by solving an integer program. Let v and $\{G_a : 0 \leq a \leq A\}$ be as in Section 4.3. Suppressing ω for clarity, we let x_{ca} be a binary variable that takes on the value 1 if call c is served when a

ambulances are available, and y_c denote the number of available ambulances when the c^{th} call arrives. Finally, let Q_c be a set of pairs $(\ell,\, a)$ such that if the ℓ^{th} call is admitted when a ambulances are available, service of the call completes after time T_{c-1}, but before time T_c (again, with ω suppressed for clarity). More formally, $(\ell,\, a) \in Q_c$ iff $T_{c-1} < T_\ell + G_a^{-1}(U_c) \leq T_c$. Our formulation is as follows:

$$Z(\omega) := \max \quad \sum_{c=1}^{C} \sum_{a=1}^{A} v(a)\, x_{ca} \tag{IP(ω)}$$

$$\text{s.t.} \quad \sum_{a=1}^{A} x_{ca} \leq 1 \qquad \forall c \in \{1,\, \ldots,\, C\} \tag{4.1}$$

$$\sum_{a=1}^{A} a\, x_{ca} \leq y_c \qquad \forall c \in \{1,\, \ldots,\, C\} \tag{4.2}$$

$$y_c = y_{c-1} - \sum_{a=1}^{A} x_{c-1,a} + \sum_{(\ell,\, a) \in Q_c} x_{\ell,a} \qquad \forall c \in \{2,\, \ldots,\, C\} \tag{4.3}$$

$$y_1 = A \tag{4.4}$$

$$x_{tc} \in \{0,\, 1\}, \quad y_c \in \{0,\, 1,\, \ldots,\, A\}$$

Constraints (4.1) enforce, for each call c, that $x_{ca} = 1$ for at most one choice of a. Constraints (4.2) serve two purposes: ensuring that no calls are admitted when all ambulances are busy, and setting $x_{c,y_c} = 1$ whenever the c^{th} arriving call is admitted. The latter occurs naturally because the objective coefficients $v(a)$ are nondecreasing in a, and so we would not have $x_{ca} = 1$ for some $a < y_c$. Constraints (4.3) preserve "flow balance": the number of idle ambulances when the c^{th} call arrives matches the number of ambulances available at time T_{c-1}, minus one if an ambulance was dispatched to call $c - 1$, plus the number of ambulances that become idle between time T_{c-1} and time T_c. Finally, constraint (4.4) enforces that all ambulances are initially idle.

Remark: The integer program (IP(ω)) can be viewed as an adaptation of that by Yue et al. [100], who also consider a relaxation in which the decision-maker has perfect information. Although our model does not consider the locations of ambulances in the bounding system, knowing the number of available ambulances at a given point in time suffices to compute our upper bound.

4.4.2 Validity for Loss Systems

To verify that Z can be used to upper bound the performance of any redeployment policy in loss systems, we prove the following:

Lemma 4.4.1. *For every sample path ω and policy π, $V^\pi(\omega) \leq Z(\omega)$.*

Proof. Proof: Fix π and ω. Let $S^\pi(\omega) \subseteq \{1, 2, \ldots, C(\omega)\}$ be the set of calls to which π dispatches an ambulance on this path (in the original system). We have

$$V^\pi(\omega) = \sum_{c \in S^\pi(\omega)} v(A_c^\pi(\omega)).$$

We use $S^\pi(\omega)$ to construct a feasible (but not necessarily optimal) solution $(\bar{x}, \bar{y})$ to (IP(ω)) that attains an objective value of at least $V^\pi(\omega)$, implying that $Z(\omega)$ must be at least as large. Consider the first call ($c = 1$), and set $\bar{y}_c = A$. If $1 \in S^\pi(\omega)$, then we set $\bar{x}_{1,\bar{y}_c} = 1$, and use Constraint (4.3) to obtain $\bar{y}_2$. If this is not the case, we set $\bar{x}_{1a} = 0$ for all a, but still use Constraint (4.3) to obtain $\bar{y}_2$. We proceed in a similar fashion for all remaining calls, in order of increasing c; this results in a solution that serves the same calls as policy π.

We claim that $(\bar{x}, \bar{y})$ satisfies $\bar{y}_c \geq A_c^\pi(\omega)$ for all calls c. This has two consequences. First, we have that $(\bar{x}, \bar{y})$ is feasible, as calls are never admitted when all ambulances are busy, and the constraints of (IP(ω)) are satisfied by construction. Second, we have that the bounding system collects at least as much reward from each call as the original system, as $v(\cdot)$ is nondecreasing. Combining these two implications yields $Z(\omega) \geq V^\pi(\omega)$, as desired.

It remains to verify the claim, which we do by proving a slightly stronger claim: that the service time for each call that is treated in the original system is at least as large as the corresponding service time in the bounding system. We proceed via induction on c. Consider first the base case ($c = 1$). If $1 \notin S^{\pi}(\omega)$, then the base case holds trivially, as the call is not served. Otherwise, $G_A^{-1}(U_c(\omega))$ lower bounds the service time of the call in the original system, by construction of the dominating service time distributions G_a.

Now suppose the claim holds until the c^{th} call arrives, where $c > 1$. The induction hypothesis implies $\bar{y}_c \geq A_c^{\pi}(\omega)$. The case where $c \notin S^{\pi}(\omega)$ is again trivial. Otherwise, $G^{-1}_{A_c^{\pi}(\omega)}(U_c(\omega))$ lower bounds the service time of the call in the original system. Although it may be that $\bar{y}_c > A_c^{\pi}(\omega)$, we have that $G^{-1}_{\bar{y}_c}(U_c(\omega)) \leq G^{-1}_{A_c^{\pi}(\omega)}(U_c(\omega))$, as the distributions G_a are stochastically decreasing in a by construction. This completes the induction, and we are done. □

We have thus shown the following result:

Theorem 4.4.2. *Let π be a redeployment policy satisfying the criteria specified in Section 4.2, and let Z be as defined in Section 4.4.1. Then $E[Z]$ is an upper bound on the expected reward collected by any policy π on the horizon $[0, T]$.*

We estimate $E[Z]$ using Monte Carlo methods, by generating and averaging i.i.d. replications of Z, and solving an instance of (IP(ω)) for each replication.

Example 4.4.3. *Consider once again the system from Example 4.3.2 in Section 4.3.3. We observe that all of the randomness in this system is captured by the random variables $U_1, \ldots, U_6$. Because G_1 corresponds to a random variable taking on one of two values with probability 0.5, and G_2 is a Dirac delta function, each sample path ω can be summarized by specifying whether $U_c \leq 0.5$ or $U_c > 0.5$ for each call c. Thus, to* *compute $E[Z]$, we need only consider $2^6 = 64$ sample paths. Enumerating over each possibility, we find that $Z(\omega) = 3.5$ for each ω. This implies $E[Z] = 3.5$, which indeed upper bounds the performance of the heuristic policy we considered in Example 4.3.2.*

4.5 Computational Study

In this section, we conduct numerical experiments on a variety of EMS systems (some realistic, some not realistic) to evaluate the tightness of our upper bound.

4.5.1 Reference Bounds

We cannot directly compare our bound to that of Maxwell et al. [61], as we assume that the EMS operates as a loss system. However, we provide context via a lower bound obtained by simulating the performance of a heuristic redeployment policy, and an alternative upper bound stemming from an "information relaxation" of the original problem.

4.5.1.1 Lower Bound

Our heuristic policy is adapted from that developed by Jagtenberg et al. [48]. Their policy dispatches the closest available ambulance to arriving calls, and when an ambulance becomes idle, redeploys it to the base providing the largest marginal increase in coverage, where coverage is computed as in the objective function of Daskin's MEXCLP [27]. We do not write this objective function here, but note that it is a function of ρ, an estimate of average ambulance utilization, as well as Θ, the locations of idle ambulances at the time a redeployment decision is made; we write this quantity as $Cvg(\Theta, \rho)$.

We make two modifications to this policy to improve its performance. First, when computing coverage, we consider ambulances undergoing redeployment to be idle at their destination bases. Second, we take into account the travel time associated with redeployment moves, as a short relocation yielding a moderate improvement in coverage may be preferable to a long relocation achieving a slightly higher increase in coverage. Letting $\Delta Cvg_j(\Theta, \rho)$ denote the marginal increase in coverage associated with placing an additional ambulance at base j, i the location of the ambulance to be redeployed, and $\alpha \geq 0$, we let

$$V_{ij} := \frac{\Delta Cvg_j(\Theta, \rho)}{\max\{t(i,\, j)^{\alpha},\, 1\}}$$

denote the "value" associated with a redeployment move from node i to base j. Our heuristic policy relocates the ambulance to the base attaining the largest V_{ij} value. We use a maximum in the denominator to avoid potential divisions by zero, and to avoid excessively inflating the value of short redeployments. This is a generalization of the metric used by Jagtenberg et al. [48], which we can recover by setting $\alpha = 0$. We can generalize their policy further by treating ρ as a parameter that can be used to tune the redeployment policy, rather than as measure of system-wide utilization. This allows us to improve our lower bound via a two-dimensional grid search over ρ and α.

4.5.1.2 Information-Based Upper Bound

Our alternative upper bound stems from a perfect information relaxation of the original problem. This entails solving a problem in which the decision-maker knows the arrival times, locations, and service requirements associated with every call at the start of the horizon— quantities that are unknown before calls arrive. The optimal objective value associated with this relaxed problem, in expectation, upper bounds the performance of any policy in the original problem; this follows by formulating the problem described in Section 4.2 as a stochastic dynamic program, and using results from Brown et al. [20]; see Chapter 5.

If we assume, contrary to our problem formulation in Section 4.2, that the decision-maker can reject arriving calls when ambulances are available, then we can solve the relaxed problem for a given sample path by instead solving an equivalent linear program that is similar in spirit to the integer program appearing in Yue et al. [100]; we refer the reader to Chapter 5 for the formulation, and for the corresponding proofs. This additional assumption loosens the upper bound, but is nonetheless valid, and provides us with another way to evaluate our upper bound from Section 4.4.

4.5.2 Findings

4.5.2.1 A First Example

We begin by considering a small hypothetical EMS in which our upper bound is fairly tight. The service area, illustrated in Figure 4.1 below, is a 9 mile $\times$ 9 mile square region, which we divide into 81 square cells. We treat each cell as a demand node, and define the distance between any two nodes to be the Manhattan distance between the centroids of the corresponding cells. Calls arrive according to a homogeneous Poisson process with rate λ. To treat these calls, the EMS provider operates 4 bases and deploys 4 ambulances, each of which travel at a constant speed of 30 miles per hour. We assume that the response time threshold is 9 minutes, and that the chute time is 1 minute. Thus, an ambulance can provide an adequate response if it travels no more than 4 miles to reach a call. Service times follow a Weibull distribution with a scale parameter of 30 and a shape parameter of 3, which has a mean of 26.8 minutes and a standard deviation of 9.7 minutes.

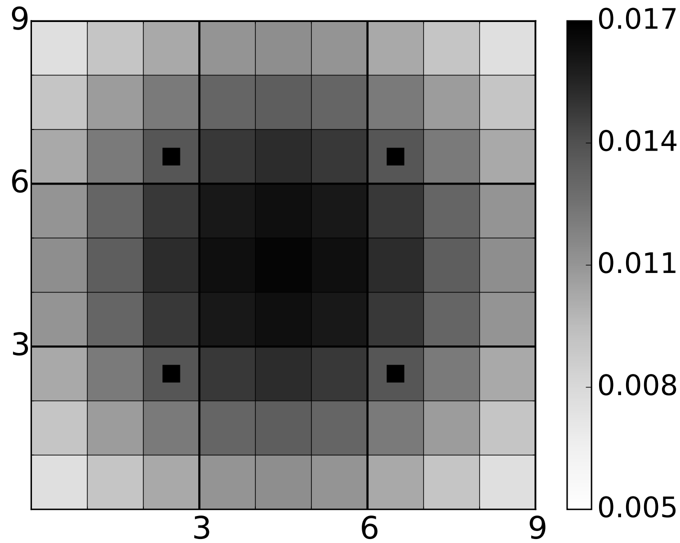

Figure 4.1: Distribution of demand and base locations (squares) in our hypothetical EMS system. Darker cells indicate areas with higher demand (larger p_i–values).

We compute the bounding function v by solving the appropriate MCLP instances to optimality. Similarly, we compute the bounding service time distributions $\{G_a : 1 \leq a \leq A\}$ by solving the corresponding p-median problems for times ranging from 1 to 90 minutes in increments of 1 minute, also to within 0.5% of optimality. We compute our lower and upper bounds by averaging 1000 i.i.d. replications, each with a length of 24 hours, for the cases when $\lambda = 2$ and 3 calls per hour (representing situations in which the system is lightly and moderately loaded). To obtain our lower bound, we simulate our heuristic policy on the same sample paths for values of ρ and α ranging from 0.05 to 0.95 in increments of 0.05, and store the objective value attained by the best policy. We summarize our results in Table 4.1.

In this example, the gap between our lower bound and our upper bound is narrow; our heuristic policy adequately responds to within 4% of the calls that can be reached under the optimal redeployment policy. By contrast, the perfect

	$\lambda = 2$	$\lambda = 3$
Lower Bound	0.807	0.686
Upper Bound	0.838	0.746
Perfect Info Upper Bound	0.963	0.909

Table 4.1: Estimated lower and upper bounds on the proportion of timely responses in a 24-hour period, for two choices of arrival rates. Half-widths of 95% confidence intervals are at most 0.001.

information upper bound is loose, which is somewhat striking, as the systems are not heavily loaded; average ambulance utilization under the heuristic policy is roughly 0.26 when $\lambda = 2$, and 0.39 when $\lambda = 3$. This suggests that there is significant value in learning call locations in advance, as is the case in the perfect information upper bound— which is intuitive, as this information allows the decision-maker to position ambulances closer to arriving calls, thus increasing the likelihood of timely responses.

4.5.2.2 Realistic Models.

Next, we study two EMS systems loosely modeled after those operated in Edmonton, Canada and Melbourne, Australia. Our models are realistic in that we use transportation networks based upon simplified versions of the road networks in both cities, and we obtain input parameters by drawing from the same datasets used by Maxwell et al. [61], which reflect real data. However, we use the term "loosely" because we make simplifying assumptions— specifically, that system dynamics (including the arrival process, edge travel times, and the size of the fleet) are time-homogeneous, and that calls do not require transport to a hospital. Thus, we do not intend for the results below to reflect how the systems in Edmonton and Melbourne actually perform. Rather, we are simply using the data available to us to gain insight into how our upper bound performs on larger, more realistic systems.

Our graph of Edmonton, which is identical to that used in Maxwell et al. [61], contains 4,725 demand nodes, which are served by 16 ambulances distributed among 11 bases. Our graph of the city of Melbourne (also from Maxwell et al. [61]) features 1,413 demand nodes, 97 ambulances, and 87 bases. As in the Section 4.5.2.1, arrivals occur according to a homogeneous Poisson process, and we use the same response time threshold, chute time, and service time distribution.

We again compute the function v by solving MCLP instances to optimality, but we compute the distributions $\{G_a : 1 \leq a \leq A\}$ by solving the linear programming relaxations of the corresponding p-median problems (due to their size). In doing so, we do not invalidate the upper bound, but we loosen it slightly. However, computational experiments suggest that the integrality gap is fairly small. To compute our lower and upper bounds, we again use 1000 replications, each with a length of 24 hours; computational experiments indicate that a heuristic policy where $\alpha = 0.50$ and $\rho = 0.70$ works well in Edmonton, and a heuristic policy where $\alpha = 0.85$ and $\rho = 0.75$ works well in Melbourne. We summarize our results in Table 4.2.

	Edmonton		Melbourne	
	$\lambda = 6$	$\lambda = 8$	$\lambda = 32$	$\lambda = 40$
Lower Bound	0.802	0.771	0.844	0.824
Upper Bound	0.853	0.849	0.888	0.887
Perfect Info Upper Bound	0.854	0.856	0.887	0.886

Table 4.2: Estimated lower and upper bounds in systems loosely modeled after Edmonton and Melbourne EMS. Half-widths of 95% confidence intervals are at most 0.001.

Our upper bounds appear to be insensitive to the level of congestion in either system. This may seem counterintuitive, as the numbers in Table 4.2 suggest that even in situations when the decision-maker has perfect information, a significant proportion of arrivals are not being adequately treated. However, these results can be explained by the fact that in both cities, a nontrivial fraction of demand originates from locations that cannot be reached by any base within the response time threshold— 0.144 in Edmonton and 0.112 in Melbourne. Thus, our upper bounds are close to the theoretical maximum possible performance that can be attained in either system.

Although the gap between our lower and upper bounds is not large, our upper bound does not improve upon the theoretical upper bound or the perfect information upper bound. To better understand the discrepancy we observe between the results in this section and those in Section 4.5.2.1, we conduct a third set of experiments on a range of EMS systems with varying characteristics.

4.5.2.3 Exploratory Models.

Figure 4.2 illustrates the 9 hypothetical EMS systems we consider. They are adapted from the "artificial cities" considered in Maxwell et al. [61]. The systems vary by the number of demand modes (1, 2, or 5), as well as by the dispersion of demand within the service area (low, medium, or high). The EMS provider in each system responds to calls within a 15 mile $\times$ 15 mile square region, which is divided into 225 square cells, and operates 25 bases. As in Section 4.5.2.1, we compute distances using the Manhattan metric, and the same service time distribution, chute time, response time threshold, and ambulance travel speed. We consider fleets operating 7 and 10 ambulances, with arrival rates of 3 and 6 per hour, respectively. Table 4.3 lists the bounds we obtain in our 9 systems, averaged over 1000 i.i.d. replications.

The perfect information upper bound is loose in each of our experiments, and appears to be insensitive to how demand is distributed within the system.

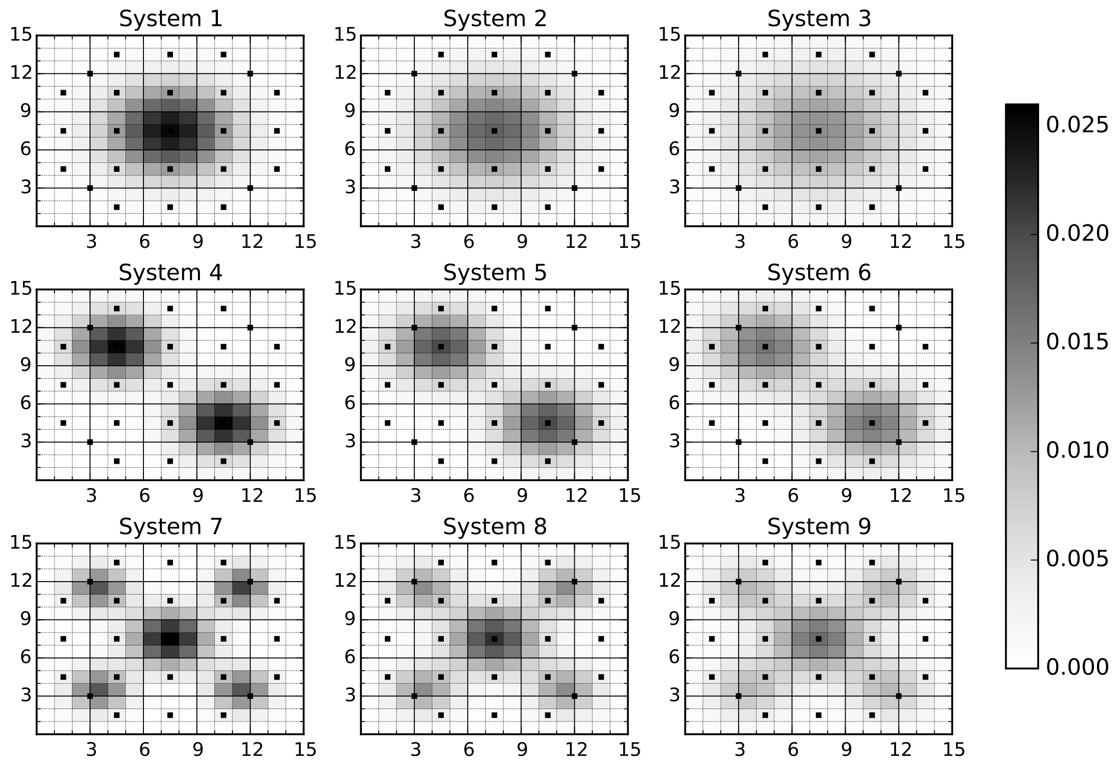

Figure 4.2: Distribution of demand and base locations in our 9 hypothetical EMS systems. Darker cells indicate areas with higher demand (larger p_i–values).

This is intuitive, as when utilization is low, variation in the locations of arriving calls have a smaller effect on performance when perfect information is available. The decision-maker can, to an extent, plan for future arrivals by positioning ambulances appropriately, regardless of where they may occur.

Our upper bound is tighter than the perfect information bound, particularly when demand is more dispersed, or when the system operates a smaller fleet. This indicates that the advantage of our upper bound may lie in its ability to take into account uncertainty in demand, and in more accurately modeling the performance of the system during periods of congestion. Nonetheless, the gap between our lower and upper bounds is significant.

7 ambulances, $\lambda = 3$	Sys. 1	Sys. 2	Sys. 3	Sys. 4	Sys. 5
Lower Bound	0.918	0.869	0.825	0.928	0.852
Upper Bound	0.972	0.949	0.926	0.979	0.967
Pf Info Upper Bound	0.996	0.993	0.990	0.996	0.991
	Sys. 6	Sys. 7	Sys. 8	Sys. 9	
Lower Bound	0.799	0.859	0.853	0.836	
Upper Bound	0.953	0.960	0.956	0.946	
Pf Info Upper Bound	0.985	0.995	0.995	0.993	
10 ambulances, $\lambda = 6$	Sys.1	Sys. 2	Sys. 3	Sys. 4	Sys. 5
Lower Bound	0.939	0.895	0.857	0.919	0.898
Upper Bound	0.985	0.973	0.961	0.988	0.982
Pf Info Upper Bound	0.998	0.997	0.995	0.996	0.995
	Sys. 6	Sys. 7	Sys. 8	Sys. 9	
Lower Bound	0.880	0.878	0.874	0.861	
Upper Bound	0.975	0.979	0.977	0.971	
Pf Info Upper Bound	0.993	0.995	0.995	0.994	

Table 4.3: Lower and upper bound estimates in 9 hypothetical EMS systems. Half-widths of 95% confidence intervals are at most 0.001.

We conjecture this may be due to the upper bound being loose, rather than to the heuristic policy being poor. This is because the bounding functions v and $\{G_a : 1 \leq a \leq A\}$ were developed by making the optimistic assumption that at any given time, ambulances are "optimally positioned"— that is, in a way that maximizes the likelihood of timely responses and short service times. When the system becomes congested, ambulances may be in a suboptimal configuration, resulting in the probability of a timely response being smaller than the function v would suggest. Furthermore, the distribution function G_a may underestimate the travel time of the responding ambulance, as dispatches may need to be made from distant bases when the system is congested. These effects may be more pronounced in large-scale systems.

This conjecture is consistent with our previous results: our bound performs well in the small-scale system in Section 4.5.2.1, but struggles with the larger-scale models in Section 4.5.2.2. Taken together, our experiments suggest that our upper bound is best suited for small-scale systems, but tends to improve upon the perfect information bound even when used outside of this setting.

4.6 Conclusion

In this chapter, we constructed an upper bound on the performance of ambulance redeployment policies in loss systems. Our work builds upon that by Maxwell et al. [61], whose bound is valid only for systems that maintain queues for calls that cannot be immediately served. The adaptation is nontrivial, and requires the introduction of new ideas. In particular, we formulate an integer program to which the set of decisions made by any nonanticipative redeployment policy (along a given sample path) correspond to a feasible solution. Computational experiments suggest that while our bound tends to be tighter than that originating from a perfect information relaxation of the ambulance redeployment problem, it is most effective in small-scale EMS systems. We conjecture that improving the gap between our lower and upper bounds hinges less upon improving the heuristic policy than upon developing refinements of the bounding functions v and $\{G_a : 1 \leq a \leq A\}$.

A natural question to ask is whether upper bounds can be obtained using other approaches. One way to do so may be to study bounds based upon relaxations of the stochastic dynamic programming formulation of the problem. This could entail, for instance, relaxing a coupling constraint, introducing Lagrange multipliers, and proceeding as in Adelman and Mersereau [1]. Alternatively, one could consider an information relaxation of the original problem. Although we consider a perfect information relaxation in our numerical study in Section

4.5, one might tighten the upper bound by applying an "information penalty" to decisions violating nonanticipativity constraints in the style of Brown et al. [20]. We attempt this in Chapter 5, but we have yet to find a penalty that meaningfully improves upon the perfect information upper bound. We also propose the construction of a more effective penalty as a direction for future research.

CHAPTER 5

AN INFORMATION PENALTY UPPER BOUND FOR AMBULANCE REDEPLOYMENT

5.1 Introduction

In this chapter, we build upon our work in Chapter 4 by developing a new upper bound on the performance of optimal ambulance redeployment policies. We do so by applying the methodology proposed by Brown et al. [20], who consider the problem of upper bounding the *value* of a stochastic dynamic program (which we define to be the expected reward obtained by an optimal policy) in situations where computing this quantity directly is computationally intractable (for instance, when the curse of dimensionality is in effect). They consider an *information relaxation* of the original dynamic program, by removing constraints requiring the decision-maker to act only upon information available at the time the action must be taken. This may entail a *perfect information relaxation*, in which the decision-maker is effectively clairvoyant, and learns the realizations of any random quantities at the start of the horizon, or more generally, any system in which more information is revealed to the decision-maker than is the case under the natural filtration. The exact means by which this is done is problem-specific, but chosen so that the relaxed problem becomes computationally tractable. Because *nonanticipative* policies remain feasible in the relaxed problem, this approach trivially results in an upper bound on the value of the dynamic program.

However, this approach often yields loose upper bounds, and so Brown et al. [20] tighten them through a function that penalizes policies that use future information to make decisions (but not nonanticipative policies). They demonstrate that there exists a choice of *penalty function* that makes the upper bound tight, but computing this penalty is no easier than solving the original problem. Nonetheless, this "ideal penalty" provides some guidance as to how effective penalties can be constructed, and so Brown et al. [20] propose a class of penalties based upon approximations to the true value function. Although they point out that "in practice, there will typically be a trade-off between the quality of the bound and the computational effort required to compute it" [20], the goal is to find a tractable penalty that closes a significant portion of the gap.

This *information penalty* framework builds upon a body of literature relating to pricing American and Bermudan options; see, for instance, Andersen and Broadie [5], Desai et al. [28], Haugh and Kogan [41], and Rogers [76]. Here, finding the exact valuation of an option also entails solving a high-dimensional stochastic dynamic program. A number of heuristic approaches (lower bounds) have been proposed, upper bounds have been devised by considering a "dual" problem in which the decision-maker is clairvoyant, but in which any benefit obtained from using this additional information to make decisions is counteracted by a "dual martingale". Brown et al. [20] view their work as an extension of this martingale duality approach to more general stochastic dynamic programs; they use it to generate effective upper bounds for problems in portfolio management [19], inventory control [20], and airline revenue management [20]. Lai et al. [52] successfully apply this methodology to the problem of storing and trading natural gas.

Our attempts at applying this framework to the ambulance redeployment problem have been less successful. We can show that solving the perfect information relaxation of the EMS provider's problem is computationally tractable, and there exists a fairly natural choice of penalty function for which solving the penalized problem is no more difficult. However, the resulting upper bound does not outperform our “loss system” bound from Chapter 4. In moderate-to-large scale systems, our information penalty bound fails to improve even upon the perfect information upper bound. We suspect that improving performance may come at the expense of ease of computation, and may entail solving a large-scale nonlinear optimization problem. We propose the development of an improved penalty, as a direction for future research.

The remainder of this chapter is organized as follows. After describing the EMS provider's problem in 5.2, we formulate it as a stochastic dynamic program in Section 5.3, from which it can very easily be seen that the curse of dimensionality is in effect. In Section 5.4, we consider the perfect information relaxation of this problem, and prove in Section 5.5 that it can be solved in a computationally tractable manner. We introduce our penalty function in Section 5.6, and demonstrate that the penalized problem is no more difficult to solve than the perfect information relaxation. Finally, in Section 5.7, we conduct an extensive numerical study to evaluate the performance of our upper bound, and to compare it to our “loss system” bound from Chapter 4.

5.2 The Model

Consider an EMS system similar to that in Chapter 4, in that its service area is represented by a graph $G = (N, E)$, and that it operates fleet of A ambulances deployed to a set of base locations $B \subseteq N$. We will sometimes abuse notation slightly and also let A denote the set of ambulances $\{1, \ldots, A\}$ in the fleet. Contrary to the previous model, time is discrete; let $\mathcal{T} = \{0, 1, 2, \ldots, T+1\}$ denote the decision horizon. At the start of time period $t \in \mathcal{T}$, a call arrives to node $i \in N$ with probability p_{ti}, where $p := \sum_{i \in V} p_{ti} \leq 1$. We assume that at most one call can arrive per time period, and that calls (e.g., their arrival times, locations, and service requirements) are independent of one another, as well as of any actions taken by the decision-maker. Thus, the arrival process in our model can be thought of as a discrete-time analogue of a non-homogeneous Poisson process. We use periods 0 and $T+1$ to initialize and terminate our model, respectively, and so we assume that calls do not arrive at these times— specifically, $\sum_{i \in V} p_{ti} = 0$ if $t = 0$ or $t = T+1$.

If a call does arrive, a dispatcher may respond to the call with a single ambulance, provided that the ambulance can reach its location within a prespecified response time threshold t_{resp}. More formally, suppose that a call arrives to node $i \in N$. Define the neighborhood

$$B_i = \{j \in B : t(i, j) \leq t_{resp}\}, \tag{5.1}$$

where $t(i, j)$ is the travel time between node i and base j. That is, B_i is the set of bases from which a timely dispatch to demand node i can be provided. We assume that dispatches to node i can only be made from bases in B_i, and that the decision-maker can ignore arriving calls— even when ambulances are available— so as to improve the system's ability to respond to future calls. This assumption is simply for convenience; the upper bounds we develop hold regardless of the dispatching policy the EMS provider employs; see Remarks 5.3.2 and 5.4.1 below. Calls that do not receive a dispatch leave the system immediately, and do not queue.

If an ambulance is dispatched, it travels to the call's location, and upon arrival, immediately begins treatment on-scene. Let τ be a random variable denoting the time spent on scene; we assume that these random variables are independent and identically distributed. This assumption can be relaxed somewhat; see Remark 5.3.1. After treatment is completed, we assume, for convenience, that the patient does not require transport to a hospital; we discuss in Remarks 5.3.1 and 5.5.4 how our methodology could be extended to incorporate hospitals. Before the ambulance can be dispatched to future calls, it must be redeployed to a base first. Once a base has been selected, the ambulance immediately begins traveling to its destination base, after which it becomes idle.

The EMS provider collects a reward of r for each timely dispatch to an incoming call; assume, without loss of generality, that $r = 1$. We seek a policy for assigning ambulances to incoming calls, as well as for for redeploying ambulances following on-scene treatment, that maximizes the expected total reward collected over the horizon $\mathcal{T}$.

5.3 Stochastic Dynamic Programming (DP) Formulation

To reduce the size of our state space, we do not keep track of ambulances' exact locations. Instead, we assume that if an ambulance is dispatched to a call originating from node i, it immediately relocates to node i, but does not become available for redeployment until the appropriate travel time and on-scene treatment time elapses. Similarly, we assume that if an ambulance is redeployed from node i to base j, that it immediately relocates to base j, but does not become available for $t(i,\, j)$ time periods.

5.3.1 State Space

At time $t \in \mathcal{T}$, the state of the system can be fully characterized by:

1. The status $s_a = (\ell_a,\, f_a,\, c_a)$ of each ambulance $a \in A$ in the system, where
   - ℓ_a is the ambulance's location within the graph
   - f_a is the time at which the ambulance is scheduled to become free, following either on-scene treatment or a redeployment move. If the ambulance is idle, then $f_a = t$.
   - c_a is the call being treated by the ambulance at time t, and ϕ otherwise
2. The location i of the arriving call in at time t, or ϕ if no arrival occurs

Let $\mathcal{S}_a = V \times \mathcal{T} \times \mathcal{T}$ denote the possible states for ambulance a; we index calls in our model by their arrival times. We can thus express the state space of our model as $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_A \times N'$, where $N' = N \cup \{\phi\}$. We sometimes write elements of $\mathcal{S}$ as ordered pairs $(s,\, i)$, where s denotes the status of the fleet, and i the location of the arriving call. Because we will sometimes be interested in the status of individual ambulances, we define $\ell_a(s)$, $f_a(s)$, and $r_a(s)$ to be the status of ambulance a implied by s.

5.3.2 Action Space

At the start of time period $t \in \mathcal{T}$, the decsion-maker may need to make up to two types of decisions:

- If a call arrives, the idle ambulance a to dispatch to the call, if any
- If one or more ambulances complete on-scene treatment, the bases to which to redeploy them

Let $D_t(s,\, i)$ denote the set of dispatching decisions that can be made at time t when the system is in state $(s,\, i)$. If $i = \phi$, only the null action ϕ can be taken. If

$i \neq \phi$, the decision-maker may dispatch an idle ambulance stationed within the neighborhood B_i, or ignore the call (and take the null action ϕ). If all ambulances are either busy or out of range, only action ϕ is available. We thus write

$$D_t(s,\, i) = \begin{cases} \{\phi\} & i = \phi \\ \{\phi\} \cup \big\{a \in A : \ell_a(s) \in B_i,\; f_a(s) = t,\; c_a(s) = \phi\big\} & i \neq \phi \end{cases}.$$

Next, let $M_t(s,\, i)$ denote the redeployment moves that can be made from state $(s,\, i)$ at time t. We represent elements of this set as tuples $m_t = (m_{t1},\, \ldots,\, m_{tA})$, where m_{ta} specifies the redeployment action performed on ambulance a at time t. If ambulance a cannot yet be redeployed (that is, if either $f_a > t$ or $c_a = \phi$), only a dummy redeployment ϕ can be made. Otherwise, we allow m_{ta} to take on values in the set of bases B. We thus write

$$M_t(s,\, i) = \Big\{(m_1, \ldots, m_A) : m_a \in B \cup \{\phi\}, \text{ with } m_a = \phi \text{ if } f_a(s) > t \text{ or } r_a(s) = \phi\Big\}$$

If $t = 0$, there are no decisions to make: we assume no calls arrive at the start of the horizon, and that ambulances are initially idle at their home bases. Thus, $D_0(s,\, i) = M_0(s,\, i) = \{\phi\}$ for states $(s,\, i)$. If $t = T+1$, calls do not arrive at the end of the horizon, but ambulances can still be redeployed, though these moves do not have any impact. We thus define the action space $\mathcal{A}$ of our model to be

$$\mathcal{A} = \bigcup_{t=1}^{T} \bigcup_{(s_t,\, i_t) \in \mathcal{S}} D_t(s_t,\, i_t) \times M_t(s_t,\, i_t).$$

5.3.3 State Transitions

Suppose that the system is in state $(s_t,\, i_t)$ at time t, and the decision-maker performs actions $d_t \in D_t(s,\, i)$ and $m_t \in M_t(s,\, i)$. Let $(S_{t+1},\, I_{t+1})$ be the state of the system at the start of the next decision epoch. We capitalize S_{t+1} and I_{t+1}, as both are random variables, given the information available to the decision-maker at time t. That is, the location of the next arriving call, if any, is random, and if a dispatch is made, the time until the ambulance becomes available is also random (as τ is random). By assumption, I_{t+1} does not depend on d_t or m_t, and

$$P(I_{t+1} = i) = \begin{cases} p_{t+1,i} & i \in N \\ 1 - \sum_{i \in N} p_{t+1,i} & i = \phi \end{cases}$$

By contrast, S_{t+1} depends upon the state of the fleet and any dispatching or redeployment decisions made at time t. Thus, we sometimes write this quantity as $S_{t+1}(s_t,\, d_t,\, m_t)$. To compute this, we consider each ambulance a individually.

Case 1: If $d_t \neq a$ and $m_{ta} = \phi$, then the ambulance is either idle, responding to a call, or en route to its destination base. We thus perform the update

$$\ell_a(S_{t+1}) = \ell_a(s_t) \tag{5.2}$$

$$f_a(S_{t+1}) = \max\{t+1,\ f_a(s_t)\} \tag{5.3}$$

$$c_a(S_{t+1}) = c_a(s_t), \tag{5.4}$$

The ambulance's location does not change, as we assume that ambulances move immediately to their destination, but remain busy. The same applies to the call to which the ambulance has been assigned, if any.

Case 2: If $d_t = a$, then ambulance a moves immediately to node i and remains busy for the time necessary to travel on scene and treat the call. We thus have

$$\ell_a(S_{t+1}) = i \tag{5.5}$$

$$f_a(S_{t+1}) = \max\{t+1,\ t + t(i,\ \ell_a(s_t)) + \tau\} \tag{5.6}$$

$$c_a(S_{t+1}) = t \tag{5.7}$$

Note that $f_a(S_{t+1})$ is a random variable, as τ is random.

Case 3: Finally, suppose $m_{ta} \neq \phi$. Ambulance a moves immediately to base m_{ta}, and remains busy for the requisite travel time, and we have

$$\ell_a(S_{t+1}) = i \tag{5.8}$$

$$f_a(S_{t+1}) = \max\{t+1,\, t + \tau(\ell_a(s_t),\, m_{ta})\} \tag{5.9}$$

$$c_a(S_{t+1}) = \phi \tag{5.10}$$

Note that we only set c_a to ϕ when the ambulance is redeployed, rather than immediately upon a service completion.

We will sometimes write $S_{t+1}(s_t,\, d_t,\, m_t)$ as $S_{t+1}(s_t,\, \pi_t)$ to emphasize the policy by which decisions are made, rather than the individual decisions themselves. When we consider a specific sample path ω, as in Section 5.4, the status of the fleet at time $t+1$ is deterministic, allowing us to write, for instance, $S_{t+1}(s_t,\, d_t,\, m_t,\, \omega)$.

5.3.4 Optimality Equations

We define a *policy* π as a collection of decision rules $(\pi_0,\, \pi_1,\, \ldots,\, \pi_{T+1})$, where each $\pi_t : \mathcal{S} \to \mathcal{A}$ is a deterministic mapping that specifies a feasible decision to take for every possible system state— that is, $\pi_t(s_t,\, i_t) \in D_t(s_t,\, i_t) \times M_t(s_t,\, i_t)$ for all states $(s_t,\, i_t)$.

By requiring π to have the above form, we restrict our attention to *nonanticipative* policies, a notion that we formally describe in Section 5.4. Let Π denote the set of all such policies; this is finite, as the state space and the number of feasible decisions in each state are finite. We assign decisions to rewards using the mapping $r : \mathcal{A} \to \{0,\, 1\}$, where

$$r(d_t,\, m_t) = \begin{cases} 1 & d_t \neq \phi \\ \\ 0 & d_t = \phi \end{cases}. \tag{5.11}$$

Although $r(d_t,\, m_t)$ depends only on the dispatching decision d_t, we include m_t as an argument for convenience. We seek a policy π that solves

$$V^* := \max_{\pi \in \Pi} \mathbb{E}\left[\sum_{t=1}^{T+1} r(\pi_t)\right], \tag{5.12}$$

where $r(\pi_t)$ denotes the (random) reward collected by policy π at time t. We refer to (5.12) as the *nonanticipative problem*, which we can solve using backward induction, via the optimality equations

$$V_t(s_t,\, i_t) = \max_{\substack{d_t \in D_t(s_t,\, i_t) \\ m_t \in M_t(s_t,\, i_t)}} \left\{ r(d_t,\, m_t) + \mathbb{E}\Big[V_{t+1}\left(S_{t+1}(s_t,\, d_t,\, m_t),\, i_{t+1}\right) \Big] \right\} \tag{DP}$$

for each $t < T + 1$, which we pair with the boundary conditions

$$V_{T+1}\left(s_{T+1},\, i_{T+1}\right) = 0 \qquad \forall (s_{T+1},\, i_{T+1}). \tag{5.13}$$

The quantity $V_t(s_t,\, i_t)$ can be interpreted as the expected number of calls that can be reached under the optimal policy from time t until the end of the horizon, if that the state of the fleet is s_t, and a call has just arrived to node i_t. Given an initial configuration of the fleet s_0, define

$$V^* = \sum_{i \in N'} p_{1i}\, V_1\left(s_0,\, i_1\right).$$

Remark 5.3.1. *Our formulation can be extended to incorporate location-dependent on-scene treatment times, as well as the possibility of patient transports to hospital. Define τ_i to be the on-scene treatment time associated with node i calls. After this completes, we assume that the patient requires transport to hospital $h \in H$ with probability ρ_{ih}, where $H \subseteq N$ is the set of hospitals in the system. We could have $\sum_{h \in H} \rho_{ih} < 1$. If* *transport is necessary, the ambulance travels to the appropriate hospital, spends a random time τ_h transferring the patient to the hospital, then becomes idle. The formulation and optimality equations would largely be unchanged: the only modification is to the state transitions associated with the ambulance a dispatched to the call. Specifically, $\ell_a(sS_{t+1})$ would represent the ambulance's location after serving a call (which could be on-scene or at a hospital), and $f_a(s_{t+1})$ must be modified to take into account the possibility of hospital transport.*

Remark 5.3.2. *If the EMS provider must adhere to certain constraints when performing dispatches, our formulation can also be modified accordingly. The only change would be to the set $D_t(s, i)$ defined in Section 5.3.3. If, for instance, the closest available ambulance must be sent (regardless of whether or not it is in range), we could write*

$$D_t(s,\, i) = \begin{cases} \{\phi\} & i = \phi \\ \big\{a \in \{1,\, 2,\, \ldots,\, A\big\} : f_a(s) = t,\, m_a(s) = \phi\big\} & i \neq \phi \end{cases}.$$

5.4 Perfect Information Relaxation

The state space of our dynamic program has $3A + 1$ dimensions, and so finding V^* via the optimality equations (DP) is intractable, even for moderately-sized systems. We proceed by considering the *perfect information* relaxation of the decision-maker's problem, in which the realizations of any random variables in our model are known at the beginning of the horizon. Uncertainty in our model is fully characterized by the sequence of call arrivals over the horizon $\mathcal{T}$: specifically their arrival times, arrival locations, and service requirements. For $t \in \mathcal{T}$, let I_t denote the location of the call arriving at time t ($I_t = \phi$ if none), and τ_t the call's service requirement (if applicable, ϕ otherwise). Because we assume that no calls arrive at time 0 and at time $T{+}1$, we have that $I_0 = \tau_0 = I_{T+1} = \tau_{T+1} = \phi$.

Let Ω denote the set of all sample paths in our model, where a sample path consists of realizations of the random variables I_t and τ_t for all time periods t. Next, let $\mathcal{F} = \{\mathcal{F}_t\}_{t \in \mathcal{T}}$ be the *natural filtration*, where $\mathcal{F}_t$ represents the decision-maker's state of knowledge at time t. More formally:

$$\mathcal{F}_t = \sigma\left(I_0,\, \tau_0, \ldots, I_{t-1},\, \tau_{t-1},\, I_t\right). \tag{5.14}$$

We omit τ_t from $\mathcal{F}_t$, as we assume that the decision-maker does not know a call's service requirement until after a dispatching decision is made. We consider a policy π to be *nonanticipative* if π_t is $\mathcal{F}_t$–measurable for all t. That is, any decision made at time t can only depend upon information revealed to the decision maker according to $\mathcal{F}_t$.

If the decision-maker has perfect information, then each π_t can $\mathcal{F}_{T+1}-$measurable. That is, decisions at any time period can be made with full knowledge of the sample path ω. Let Π^{PI} denote the collection of all such anticipative policies. The *perfect information relaxation* of (DP) entails solving

$$V^{PI} := \max_{\pi \in \Pi^{PI}} \mathbb{E}\left[\sum_{t=0}^{T+1} r(\pi_t)\right] = \mathbb{E}\left[\max_{\pi \in \Pi^{PI}} \sum_{t=0}^{T+1} r(\pi_t)\right]. \tag{5.15}$$

The second equality follows because π can be $\mathcal{F}_{T+1}$—measurable, implying that the decision-maker has the freedom to optimize along individual sample paths. Clearly, $V^* \leq V^{PI}$, as $\Pi \subseteq \Pi^{PI}$, and given a path $\omega \in \Omega$, we can solve for the optimal policy on that path via backward induction, this time using the optimality equations

$$V_t^{PI(\omega)}(s_t,\, i_t) = \max_{\substack{d_t \in D_t(s_t, i_t) \\ m_t \in M_t(s_t, i_t)}} \left\{ r(d_t) + V_{t+1}^{PI(\omega)}\left(S_{t+1}(s_t,\, d_t,\, m_t,\, \omega),\, I_{t+1}(\omega)\right)\right\}, \tag{PIDP(ω)}$$

along with a boundary condition analogous to (5.13). We parametrize our optimality equations by ω to emphasize the fact that the perfect information problem must be solved on a path-by-path basis. The right-hand side of (PIDP(ω))

no longer contains an expectation, as the system evolves deterministically once we know ω. Given an initial configuration s_0 of the fleet, it follows that

$$V^{PI} = \int_{\omega} V_0^{PI(\omega)}(s_0,\, \phi)\, d\mathbb{P}(\omega)$$

Evaluating this integral is intractable, as $|\Omega|$ will either be infinite (if τ_{svc} follows a continuous distribution), or grow exponentially with T (if τ_{svc} follows a discrete distribution). However, we can use Monte Carlo simulation to estimate V^{PI}. Given i.i.d. replications $\omega_1, \ldots, \omega_n$, we obtain the sampled estimate

$$\hat{V}^{PI} \approx \frac{1}{n}\sum_{k=1}^{n} V^{PI}(\omega_k) = \frac{1}{n}\sum_{k=1}^{n} V_0^{PI(\omega_k)}(s_0,\, \phi).$$

Remark 5.4.1. *The quantity V^{PI} remains an upper bound even if the EMS provider's dispatching policy is constrained as in Remark 5.3.2. In this case,* (PIDP(ω)) *can be viewed as a perfect information relaxation in which the decision-maker is granted additional freedom when performing dispatches.*

5.5 Solving the Perfect Information Problem

Given a sample path ω, finding $V^{PI}(\omega)$ using backward induction remains intractable, as the curse of dimensionality is still in effect. However, the perfect information relaxation problem can be formulated as an equivalent integer program. Suppose the decision-maker deploys a_j ambulances to base $j \in B$ at the start of the horizon. Let $\mathcal{C} \subseteq \mathcal{T}$ be the subset of time periods at which calls arrive. For each $c \in \mathcal{C}$, we define the neighborhood B_c, the set of bases from which an ambulance can reach call c's location within the response time threshold t_{resp}. Given a base $j \in B_c$, define the quantity let $f(j,\, c) = c + t(j,\, I_c) + \tau_c$ denote the time at which an ambulance responding from base j to call c would become free. Since we have fixed ω, both I_c and τ_c are deterministic, but we suppress ω for clarity. Next, we define the sets

$$Q_{tj} = \{(c,\, k) : c \in \mathcal{C},\, k \in B,\, f(k,\, c) + t(I_c,\, j) = t\}, \tag{5.16}$$

the set of call-base pairs $(c,\, k)$ for which a dispatch from base k to call c, followed by a redeployment to j, results in the ambulance becoming idle at time t. Finally,

$$x_{cj} = \begin{cases} 1 & \text{if call } c \text{ served by ambulance from base } j \in B_c \\ 0 & \text{otherwise} \end{cases}$$

$$y_{tj} = \text{At time } t\text{, the number of ambulances idle at base } j$$

$$z_{tcj} = \begin{cases} 1 & \text{if at time } t\text{, ambulance responding to call } c \text{ begins redeployment} \\ & \qquad \text{to base } j \\ 0 & \text{otherwise} \end{cases}$$

Our integer programming formulation is as follows:

$$\max \quad \sum_{c \in \mathcal{C}} \sum_{j \in B_c} x_{cj} \tag{IP1(ω)}$$

$$\text{s.t.} \quad \sum_{j \in B_c} x_{cj} \le 1 \qquad \forall c \in \mathcal{C} \tag{5.17}$$

$$x_{cj} \le y_{cj} \qquad \forall c \in \mathcal{C}, j \in B \tag{5.18}$$

$$\sum_{k \in B} z_{f(j,c),c,k} = x_{cj} \qquad \forall c \in \mathcal{C}, j \in B_c \tag{5.19}$$

$$y_{t+1,j} = y_{tj} - x_{tj} I(t \in \mathcal{C},\, j \in B_t) \tag{5.20}$$

$$+ \sum_{(c,\,k) \in Q_{t+1,j}} z_{f(k,c),c,j} \qquad \forall t \in \mathcal{T}, j \in B \tag{5.21}$$

$$y_{0j} = a_j \qquad \forall j \in B \tag{5.22}$$

$$x_{cj} \in \{0,\, 1\} \qquad \forall c \in \mathcal{C}, j \in B \tag{5.23}$$

$$y_{tj} \in \{0,\, 1,\, \ldots,\, A\} \qquad \forall t \in \mathcal{T}, j \in B \tag{5.24}$$

$$z_{tcj} \in \{0,\, 1\} \qquad \forall t \in \mathcal{T}, c \in \mathcal{C}, j \in B \tag{5.25}$$

Like (PIDP(ω)), the integer program (IP1(ω)) aims to maximize the number of calls receiving timely responses. Constraints (5.17) ensure that at most one ambulance that is in range can be dispatched to a call, while Constraints (5.18) only allow these dispatches to be made from bases with idle ambulances.

Constraints (5.19) require ambulances to be redeployed immediately after they complete service with a call. Specifically, if $x_{cj} = 1$, then a redeployment must occur at time $f(j, c)$, the time at which the responding ambulance completes the call. Constraints (5.21) maintain flow balance at each base over time. Specifically, the number of ambulances idle at base j at time $t + 1$ equals the number idle at time t, minus the number of ambulances dispatched from j at time t, plus the number of ambulances that finish redeployment to j between time t and $t + 1$. Finally, Constraints (5.22) initialize the system.

We can solve (IP1(ω)) in lieu of (PIDP(ω)):

Lemma 5.5.1. *The problems* (PIDP(ω)) *and* (IP1(ω)) *are equivalent.*

We prove this equivalence in Appendix C.1, which hinges upon showing that any feasible solution to one problem can be converted to a feasible solution to the other with the same objective value, and vice versa.

Although Lemma 5.5.1 allows us to circumvent the curse of dimensionality, (IP1(ω)) is not easy to solve for large-scale systems, as there are roughly $O(|\mathcal{T}|\,|\mathcal{C}|\,|B|)$ decision variables and $O(|\mathcal{T}|\,|B| + |\mathcal{C}|\,|B|)$ constraints. As an example, the EMS system in Melbourne, Australia operates 87 bases, and receives on average roughly 25 calls per hour. If we define a time period in our model to be one minute, and consider a horizon length of 24 hours (so $|\mathcal{T}|$ = 1440), a naïve implementation of (IP1(ω)) would have on the order of millions of decision variables and hundreds of thousands of constraints. Consequently, (IP1(ω)) may not be significantly easier to solve than (DP) or (PIDP(ω)).

We proceed by simplifying our integer program (IP1(ω)). To do so, we observe that we need not treat dispatch decisions separately from redeployment decisions in our model, as the decision-maker has perfect information. To see why this is the case, fix a sample path ω, and consider a feasible solution to (IP1(ω)). Suppose that in this solution, $x_{cj} = 1$ for some call c and base j. Let t be the time at which treatment of the call completes, which can easily be computed from the sample path ω. If $t \leq T + 1$, Constraint (5.19) requires that the ambulance be redeployed to another base k, and the decision-maker selects a base at this time. However, to a decision-maker with perfect information, there is nothing to gain by making dispatching and redeployment decisions at two separate points in time (specifically, at time c and time t), as the information available at both points in time is identical. Thus, if we reformulate our integer program so that redeployment decisions must be made immediately after calls arrive, we obtain a model that is equivalent to (IP1(ω)).

We proceed by removing the decision variables z_{tcj}, omitting any constraints including these decision variables, and modifying the x–variables so that they also specify where ambulances are redeployed. Specifically, define

$$x_{cjk} = \begin{cases} 1 & \text{if call } c \text{ served by ambulance from base } j \in B_c \text{ that is later} \\ & \qquad \text{redeployed to base } k \\ 0 & \text{otherwise} \end{cases}$$

Letting the variables y_{tj} be as before, we reformulate (IP1(ω)) is as follows:

$$\max \quad \sum_{c\in\mathcal{C}}\sum_{j\in B_c}\sum_{k\in B} x_{cjk} \tag{IP2(ω)}$$

$$\text{s.t.} \quad \sum_{j\in B_c}\sum_{k\in B} x_{cjk} \le 1 \qquad \forall c\in\mathcal{C} \tag{5.26}$$

$$\sum_{k\in B} x_{cjk} \le y_{cj} \qquad \forall c\in\mathcal{C}, j\in B_c \tag{5.27}$$

$$y_{t+1,j} = y_{tj} - x_{tj} I(t\in\mathcal{C},\, j\in B_t) \tag{5.28}$$

$$+ \sum_{(c,\,k)\in Q_{t+1,j}} x_{ckj} \qquad \forall t\in\mathcal{T}, j\in B \tag{5.29}$$

$$y_{0j} = a_j \qquad \forall j\in B \tag{5.30}$$

$$x_{cjk}\in\{0,\,1\} \qquad \forall c\in\mathcal{C}, j\in B_c, k\in B \tag{5.31}$$

$$y_{tj}\in\{0,\,1,\,\ldots,\,A\} \qquad \forall t\in\mathcal{T}, j\in B \tag{5.32}$$

Constraints (5.26), (5.27), and (5.29) are analogous to Constraints (5.17), (5.18), and (5.21) in (IP1(ω)), respectively.

Lemma 5.5.2. *Integer programs* (IP1(ω)) *and* (IP2(ω)) *are equivalent.*

The proof again hinges upon showing that any feasible solution to (IP1(ω)) can be converted into a feasible solution to (IP2(ω)) with the same objective value, and vice versa. Constructing solutions is straightforward, as if $x_{cjk} = 1$ for some triplet $(c,\, j,\, k)$ in (IP2(ω)), then we can set $x_{cj} = 1$ and $s_{tck} = 1$ in (IP1(ω)) (for the appropriate value of t), and vice versa. It remains to check feasibility of the resulting solutions, and we do this in Appendix C.2.

While (IP2(ω)) has significantly fewer decision variables and constraints than (IP1(ω)), solving it may still be impractical for large-scale systems; in the case of Melbourne, we would need to solve an integer program with potentially hundreds of thousands of decision variables and constraints. However, linear programs of this size are solved routinely, and so solving the linear programming relaxation of (IP2(ω)) is tractable. Doing so would loosen our upper bound if (IP2(ω)) has an integrality gap. However, this has not been the case in any of the problem instances we have encountered. This leads us to conjecture that (IP2(ω)) does not have an integrality gap:

Conjecture 5.5.3. *Integer program* (IP2(ω)) *has no integrality gap: that is, for any sample path ω, any optimal solution to the linear programming relaxation of* (IP2(ω)) *attains an objective value of $V^{PI}(\omega)$.*

We suspect that a proof would entail demonstrating that the coefficient matrix associated with (IP2(ω)) is totally unimodular. We can show this for a relaxation of (IP2(ω)) in which Constraints (5.26) are relaxed, which can be reformulated as an equivalent min-cost (or, in this case, max-reward) network flow problem, implying total unimodularity. Reintroducing Constraints (5.26) eliminates this network structure, but we suspect that total unimodularity is preserved. We elaborate upon these ideas in Appendix C.3.

If our conjecture holds, then solving the perfect information relaxation to the ambulance redeployment problem entails solving a moderately-sized linear program, which is tractable even in large-scale EMS systems.

Remark 5.5.4. *The integer programs* (IP1(ω)) *and* (IP2(ω)) *can be extended to take hospital transfers into account, in the spirit of Remark 5.3.1. This would entail modifying the constants $f(j, c)$ and sets M_{tj} to take into account the possibility of hospital transports. Since having perfect information entails knowing of any potential transfers at the beginning of the horizon, this modification does not present significant difficulties, and all of the results presented above can be adapted to this setting.*

Remark 5.5.5. *Our integer program* (IP2(ω)) *is based heavily upon that by Yue et al. [100], who consider the problem of ambulance dispatching when the decision-maker has perfect information. We adapt their model to allow for ambulances to be redeployed following on-scene treatment.*

5.6 A Penalty Function

5.6.1 Preliminaries

Although the computational issues associated with computing V^{PI} appear to have been resolved, the perfect information relaxation yields a weak upper bound, as we demonstrate through computational experiments in Section 5.7. To tighten the upper bound, we introduce a penalty function as in Brown et al. [20]. Recall that doing so entails the penalizing policies that use future information to make decisions.

We begin by revisiting Brown et al.'s definition of a penalty function, and adapt it to our setting. Let $\pi = (\pi_0, \ldots, \pi_{T+1}) \in \Pi^{PI}$, and define, for each time t, a function $z_t(s_t,\, i_t,\, \pi_t)$ denoting the penalty assessed to the decision-maker when taking action $\pi_t(s_t,\, i_t) = (d_t,\, m_t)$ when the system is in state $(s_t,\, i_t)$. To avoid penalizing nonanticipative polices, Brown et al. [20] only consider penalties $z = (z_1,\, \ldots, z_{T+1})$ that are *dual feasible*, in that they satisfy

$$\mathbb{E}\left[\sum_{t=0}^{T+1} z_t(S_t^\pi,\, I_t,\, \pi_t)\right] \leq 0 \qquad \forall \pi \in \Pi, \tag{5.33}$$

where S_t^π denotes the (random) state of the system under policy π at time t. That is, a dual feasible penalty does not penalize nonanticipative policies in expectation. Let $\mathcal{Z}$ be the collection of all such penalties. Given $z \in \mathcal{Z}$, we can incorporate it into the perfect information relaxation (PIDP(ω)) as follows:

$$
\begin{aligned}
V^{PI}(z) :&= \max_{\pi \in \Pi^{PI}} \mathbb{E}\left[\sum_{t=0}^{T+1} r(\pi_t) - z_t(S_t^{\pi},\, I_t,\, \pi_t)\right] \\
&= \mathbb{E}\left[\max_{\pi \in \Pi^{PI}} \sum_{t=0}^{T+1} r(\pi_t) - z_t(S_t^{\pi},\, I_t,\, \pi_t)\right],
\end{aligned}
\tag{5.34}
$$

Brown et al. [20] call (5.34) a *dual problem* with respect to the original stochastic DP, and we will sometimes refer to it as a *penalized problem*. Given a path ω, we can again find the optimal policy along that path via backward induction

$$
\begin{aligned}
V_t^{PI(z,\omega)}(s_t,\, i_t) = \max_{\substack{d_t \in D_t(s_t,\, i_t) \\ m_t \in M_t(s_t,\, i_t)}} \Big\{ & r(d_t) - z_t(s_t,\, i_t,\, (d_t,\, m_t)) \\
& + V_{t+1}^{PI(z,\omega)}\big(S_{t+1}(s_t,\, d_t,\, m_t,\, \omega),\, I_{t+1}(\omega)\big)\Big\},
\end{aligned}
\tag{5.35}
$$

along with the boundary conditions $V_{T+1}^{PI(z,\omega)}(s_{T+1},\, \phi) = 0$ for all s_{T+1}. If z is dual feasible, then $V^{PI}(z)$ upper bounds V^*, a result that is analogous to weak duality in linear programming.

Lemma 5.6.1. *If $z = (z_0,\, z_1,\, \ldots,\, z_{T+1})$ is dual feasible, then $V^* \leq V^{PI}(z)$.*

Proof. Following the proof in Brown et al. [20], we have that

$$
\begin{aligned}
V^* &= \max_{\pi \in \Pi} \mathbb{E}\left[\sum_{t=0}^{T+1} r(\pi_t)\right] \\
&\leq \max_{\pi \in \Pi} \mathbb{E}\left[\sum_{t=0}^{T+1} r(\pi_t) - z_t\left(S_t^{\pi},\, I_t,\, \pi_t\right)\right] \\
&\leq \max_{\pi \in \Pi^{PI}} \mathbb{E}\left[\sum_{t=0}^{T+1} r(\pi_t) - z_t\left(S_t^{\pi},\, I_t,\, \pi_t\right)\right] = V^{PI}(z),
\end{aligned}
$$

where the second line follows by definition of dual feasibility, and the third because $\Pi \subseteq \Pi^{PI}$. $\square$

We seek a penalty that makes the upper bound as tight as possible, which entails solving the optimization problem

$$\inf_{z \in \mathcal{Z}} V^{PI}(z) = \inf_{z \in \mathcal{Z}} \max_{\pi \in \Pi^{PI}} \mathbb{E}\left[\sum_{t=0}^{T+1} r(\pi_t) - z_t(S_t^{\pi},\, I_t,\, \pi_t)\right]. \tag{5.36}$$

Brown et al. [20] prove the existence of a penalty yielding a zero-variance, tight upper bound: a result that is analogous to strong duality in linear programming.

Lemma 5.6.2. *Let $(V_1^*, \ldots, V_{T+1}^*)$ be the value functions associated with an optimal nonanticipative policy, in that they solve the optimality equations* (DP). *Given $\pi \in \Pi^{PI}$ and a sample path ω, define for each $t < T+1$ the penalty*

$$\begin{aligned} z_t^*\left(s_t,\, i_t,\, \pi_t,\, \omega\right) = {} & V_{t+1}^*\left(S_{t+1}(s_t,\, \pi_t,\, \omega),\, I_{t+1}(\omega)\right) \\ & - E\left[V_{t+1}^*\left(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\right) \middle| \mathcal{F}_t\right] \end{aligned} \tag{5.37}$$

Then for every sample path ω and initial system state s_0, we have $V^{PI}(z^*) = V^*$.

Remark: The conditional expectation on the right-hand side of (5.37) is with respect to the location of the next arriving call (if any) and the time required to serve the call arriving in the current time period (provided a dispatch is made).

We defer the proof to Appendix C.4. Although Lemma 5.6.2 implies the existence of a penalty that makes the upper bound tight, finding the value functions V_t^* is no easier that solving (DP) to optimality. To find a "good" penalty— one that is computationally tractable, but that yields a looser upper bound— Brown et al. [20] consider a more general class of penalties:

Lemma 5.6.3. *Let $(\tilde{V}_1, \ldots, \tilde{V}_{T+1})$ be a collection of functions where for each t, $\tilde{V}_t : \mathcal{S} \to \mathbb{R}$. Define, for each sample path ω, policy $\pi \in \Pi^{PI}$, and $t < T+1$, the penalty*

$$\begin{aligned} z_t\left(s_t,\, i_t,\, \pi_t,\, \omega\right) = {} & \tilde{V}_{t+1}\left(S_{t+1}(s_t,\, \pi_t,\, \omega),\, I_{t+1}(\omega)\right) \\ & - E\left[\tilde{V}_{t+1}\left(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\right) \middle| \mathcal{F}_t\right] \end{aligned} \tag{5.38}$$

along with the boundary condition $z_{T+1} = 0$. Then $z = (z_0, \ldots, z_{T+1})$ is dual feasible.

We defer the proof to Appendix C.5. Note that we recover the "ideal penalty" from Lemma 5.6.2 by setting $\tilde{V}_t = V_t^*$ for each t. In this case, the right-hand side of (5.38) has a fairly intuitive interpretation. In particular, the difference (5.38) is positive if action $\pi_t(s_t,\, i_t)$ is more appealing to a decision-maker with perfect information than to one obtaining information according to the filtration $\mathcal{F}$. Thus, $z_t^*\left(s_t,\, i_t,\, \pi_t\right)$ can be viewed as the value of information associated with taking action π_t at time t when the system is in state $(s_t,\, i_t)$, and $z^* = \sum_{t=0}^{T+1} z_t^*$ as the value of information over the horizon $\mathcal{T}$— or, alternatively, the value of knowing the full sample path ω at time 0.

Together, Lemmas 5.6.2 and 5.6.3 provide guidance as to how good penalties can be constructed. Although computing V_t^* is a hopeless task, we could instead develop an approximation $\tilde{V}_t$ to the true value function V_t^*, and construct a penalty as in (5.38). Brown et al. [20] observe that for such a penalty to be effective, the approximation need not satisfy $\tilde{V}_t \approx V_t^*$ for each t, but

$$
\begin{aligned}
&E\left[\tilde{V}_{t+1}\left(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\right) \middle| \mathcal{F}_{T+1}\right] - E\left[\tilde{V}_{t+1}\left(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\right) \middle| \mathcal{F}_t\right] \\
&\qquad \approx E\left[V_{t+1}^*\left(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\right) \middle| \mathcal{F}_{T+1}\right] - E\left[V_{t+1}^*\left(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\right) \middle| \mathcal{F}_t\right]
\end{aligned}
$$

instead. In other words, a good penalty need not be based upon a $\tilde{V}_t$ that approximates the true value function well, but that approximates the value function differences in (5.38) well.

5.6.2 A Tractable Penalty

In selecting a penalty for our problem, we face a trade-off between tractability and performance. Although several value function approximations have been proposed in the literature (see, for instance Maxwell et al. [59, 60] and Schmid [80]), we encounter computational issues when we attempt to use these approximations as the basis for a penalty function. Although we could apply (5.38) to construct a penalty and modify (IP2(ω)) accordingly, we obtain an objective function that is nonlinear, which is not amenable to optimization (even in the absence of an integrality gap). Thus, we restrict attention to penalties that can be written as linear functions of the decision variables appearing in (IP2(ω)). We begin by considering a penalty based upon the value function approximation

$$V_t^{cvg}(s_t,\, i_t) = \sum_{j \in B_{i_t}} y_{tj}, \tag{5.39}$$

the number of ambulances that can cover the call arriving at time t (if any), given that the status of the fleet s_t at that time. That is, we approximate the reward-to-go from time t onward by a linear function of the coverage the system can provide to the call arriving at time t. Thus, V^{cvg} can be viewed as a one-step look-ahead approximation of the true value function V^*. This is a fairly crude approximation, but looking more than one step ahead would result in a nonlinear objective function. We observe that y_{tj} can easily be computed from s_t by counting the number of idle ambulances at base j:

$$y_{tj} = \sum_{a=1}^{A} I(\ell_a(s_t) = j,\, f_a(s_t) = t,\, c_a(s_t) = \phi).$$

We use V^{cvg} to build a penalty as in (5.38). Given ω, Equation (5.39) implies

$$V_{t+1}^{cvg}\left(S_{t+1}(s_t,\, \pi_t,\, \omega),\, I_{t+1}(\omega)\right) = \sum_{j \in B_{I_{t+1}(\omega)}} y_{t+1,j} = \sum_{j \in B} \mathbb{1}\big(j \in I_{t+1}(\omega)\big)\, y_{t+1,\,j}$$

To compute $E\left[\tilde{V}_{t+1}\left(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\right) \big| \mathcal{F}_t\right]$, we make two observations. First, the expectation need only be taken over I_{t+1}, the location of the call arriving at time $t+1$. Second, the number of idle ambulances at time $t+1$ is a deterministic function of s_t and π_t. This is because we implicitly assume in our DP formulation that the decision-maker knows the time at which ambulances become free once they have been dispatched (and by making the additional mild assumption that ambulances dispatched at time t would be busy at time $t+1$ with probability 1). It follows that

$$E\left[V_{t+1}^{cvg}\left(S_{t+1}(s_t,\,\pi_t),\,I_{t+1}\right)\middle|\mathcal{F}_t\right] = \sum_{i\in N} p_{t+1,i} \sum_{j\in B_i} y_{t+1,j}.$$

For convenience, we rewrite this conditional expectation as

$$E\left[V_{t+1}^{cvg}\left(S_{t+1}(s_t,\,\pi_t),\,I_{t+1}\right)\middle|\mathcal{F}_t\right] = \sum_{j\in B} y_{t+1,\,j} \sum_{\substack{i\in N:\\ j\in B_i}} p_{t+1,i} =: \sum_{j\in B} \rho_{t+1,j}\, y_{t+1,\,j},$$

and so the value function approximation V^{cvg} induces the penalty

$$\begin{aligned} z_t^{cvg}\left(s_t,\,i_t,\,\pi_t,\,\omega\right) &= \sum_{j\in B} \mathbb{1}\big(t+1\in\mathcal{C},\, j\in I_{t+1}(\omega)\big)\, y_{t+1,\,j} - \sum_{j\in B} \rho_{t+1,j}\, y_{t+1,\,j} \\ &= \sum_{j\in B}\Big[\mathbb{1}\big(t+1\in\mathcal{C},\, j\in I_{t+1}(\omega)\big) - \rho_{t+1,j}\Big]\, y_{t+1,\,j} \end{aligned} \tag{5.40}$$

Incorporating this penalty into the objective function of (IP2(ω)), yields the *penalized integer program*

$$\begin{aligned} \max \quad & \sum_{c\in\mathcal{C}} \sum_{j\in B_c} \sum_{k\in B} x_{cjk} \\ & \quad - \sum_{t=0}^{T} \sum_{j\in B} \Big[\mathbb{1}\big(t+1\in\mathcal{C},\, j\in I_{t+1}(\omega)\big) - \rho_{t+1,j}\Big]\, y_{t+1,\,j} \\ \text{s.t.} \quad & \text{Constraints (5.27) – (5.32)} \end{aligned} \tag{PIP(ω)}$$

We can reason as in Lemmas 5.5.1 and 5.5.2 to show that (PIP(ω)) is equivalent to the penalized DP (5.35) when we substitute in the penalty $z^{cvg} = (z_1^{cvg}, \ldots, z_T^{cvg})$. Furthermore, (PIP($\omega$)) is no more difficult to solve than (IP2(ω)), as both models have the same constraints and linear objective functions. Provided Conjecture 5.5.3 holds, we can solve (PIP(ω)) via its linear programming relaxation.

Remark 5.6.4. *Our value function approximation* (5.39) *double-counts coverage, in that it distinguishes between situations in which one ambulance can respond to a call,* *and multiple ambulances can do so. One may wonder if it is more appropriate to disregard any ambulances covering a location beyond the first, as we only allow one response to an incoming call. Although we agree that this would yield a better approximation, incorporating this penalty into* (IP2(ω)) *is challenging. This is because we require additional logical decision variables that take on the value 1 if* $\sum_{j \in B_i} y_{tj} > 0$ *for the appropriate time* t *and node* i*, and 0 otherwise. Doing so requires adding big–*M *constraints, which introduces a significant integrality gap, and renders the model intractable.*

5.6.3 Tuning Our Penalty Function

We can tighten the upper bound generated by (PIP(ω)) by considering a slight generalization of the penalty z^{cvg}. In particular, consider the penalty arising from the value function approximation

$$V_t^{cvg,\gamma}(s_t,\, i_t) = \sum_{j \in B_{i_t}} \gamma_{tj}\, y_{tj}, \tag{5.41}$$

where the γ_{tj} are coefficients that allow us to vary how ambulances contribute to coverage with location and time. Note that we recover our original value function approximation (5.39) by setting $\gamma_{tj} = 1$ for all t and j. Given a collection $\gamma = \{\gamma_{tj} : t \in \mathcal{T},\, j \in B\}$, the approximation $V^{cvg,\gamma}$ induces the penalty

$$z_t^{cvg,\gamma}\,(s_t,\, i_t,\, \pi_t,\, \omega) = \sum_{j \in B} \Big[\mathbb{1}\big(j \in I_{t+1}(\omega)\big)\, \gamma_{t+1,\,j} - \gamma_{t+1,\,j}\, \rho_{t+1,j} \Big] y_{t+1,\,j}, \tag{5.42}$$

and incorporating this modified penalty into (IP2(ω)) yields the integer program

$$\begin{aligned} \max \quad & \sum_{c \in \mathcal{C}} \sum_{j \in B_c} \sum_{k \in B} x_{cjk} && \text{(PIP}(\gamma,\,\omega)) \\ & \quad - \sum_{t=0}^{T} \sum_{j \in B} \Big[\mathbb{1}\big(j \in I_{t+1}(\omega)\big)\, \gamma_{t+1,\,j} - \gamma_{t+1,\,j}\, \rho_{t+1,j} \Big] y_{t+1,\,j} \\ \text{s.t.} \quad & \text{Constraints (5.27) – (5.32)} \end{aligned}$$

As is the case with (PIP(ω)), the integer program (PIP(γ, ω)) has the same constraints as (IP2(ω)), so we can again solve an LP relaxation, provided that Conjecture 5.5.3 holds. Let $V^{PIP}(\gamma)$ be a family of random variables, parameterized by γ, for which $V^{PIP}(\gamma, \omega)$ is the objective value attained by an optimal solution to (PIP(γ, ω)). Because $z^{cvg,\gamma}$ is a dual feasible penalty by Lemma 5.6.3, it follows that $\mathbb{E}[V^{PIP}(\gamma)] \geq V^*$ for any choice of coefficients γ. To make our upper bound as tight as possible, we would solve the optimization problem

$$\min_{\gamma} \mathbb{E}\left[V^{PIP}(\gamma)\right]. \tag{5.43}$$

Assuming that we can evaluate the objective function at a given γ, solving (5.43) is a straightforward task, as the objective function is convex in γ:

Lemma 5.6.5. *The function $\mathbb{E}\left[V^{PIP}(\gamma)\right]$ is convex in γ.*

We present the proof in Appendix C.6. Although we need only solve a convex optimization problem, we again run into technical issues because we cannot evaluate the expectation in (5.43). Instead, we consider a sample average approximation to this expectation: specifically, by solving

$$\min_{\gamma} \widehat{V}^{PIP}(\gamma) := \min_{\gamma} \frac{1}{n} \sum_{i=1}^{n} V^{PIP}(\gamma, \omega_i) \tag{5.44}$$

over i.i.d. replicates $\omega_1, \ldots, \omega_n$ instead. The sampled upper bound function $\widehat{UB}(\gamma)$ is also convex in γ, as we show in Lemma 5.6.5 that $V^{PIP}(\gamma, \omega)$ is convex in γ, and a convex combination of convex functions is itself convex.

To solve the convex optimization problem (5.44), we use gradient descent. Computing a gradient is straightforward once we have solved the corresponding integer programming instances to optimality. That is, fix γ and a path ω, and let $\big(x^*(\gamma,\,\omega),\, y^*(\gamma,\,\omega)\big)$ be an optimal solution to (PIP($\gamma,\,\omega$)). We have

$$\begin{aligned}\frac{\partial}{\partial\gamma_{tj}} V^{PIP}(\gamma,\,\omega) &= \frac{\partial}{\partial\gamma_{tj}}\Bigg(\sum_{c,\,j,\,k} x^*_{cjk}(\gamma,\,\omega) - \sum_{t=0}^{T}\sum_{j\in B}\Big[\mathbb{1}\big(j\in I_{t+1}(\omega)\big)\,\gamma_{t+1,\,j} \\ &\qquad\qquad - \gamma_{t+1,\,j}\,\rho_{t+1,j}\Big] y^*_{t+1,\,j}(\gamma,\,\omega)\Bigg) \\ &= \Big[\mathbb{1}\big(j\in I_t(\omega)\big) - \rho_{tj}\Big] y^*_{tj}(\gamma,\,\omega),\end{aligned}$$

from which it follows that

$$\frac{\partial}{\partial\gamma_{tj}}\widehat{UB}(\gamma) = \frac{1}{n}\sum_{i=1}^{n}\Big[\mathbb{1}\big(j\in I_t(\omega_i)\big) - \rho_{tj}\Big] y^*_{tj}(\gamma,\,\omega_i). \tag{5.45}$$

We initialize our gradient search procedure by setting $\gamma = 0$ (that is, by considering an unpenalized problem), computing $\widehat{UB}(0)$ by solving the corresponding instances of (IP2(ω)), and applying (5.45) to the resulting optimal solutions to compute $\nabla\widehat{UB}(0)$. We move in the direction of the gradient over a prespecified step size (provided this decreases the objective function), recompute the gradient, and repeat until either the improvement in the objective within an iteration falls below a certain threshold, or a certain number of iterations have elapsed: whichever occurs first. Let γ^* be the vector we obtain upon termination.

We cannot use $\widehat{UB}(\gamma^*)$ to evaluate the performance of our upper bound, as this is a biased estimate of $UB(\gamma)$. To remove bias, we generate an independent collection of i.i.d. replicates $\omega_{n+1},\,\ldots,\,\omega_{2n}$, and compute the estimate

$$\frac{1}{n}\sum_{i=n+1}^{2n} V^{PIP}(\gamma^*,\,\omega_i).$$

5.7 Computational Results

In this section, we study the performance of our upper bound by applying it to a wide range of EMS systems, and by comparing it to the "loss system" bound we developed in Chapter 4. In the experiments that follow, we use the term *duality gap* to describe to the relative difference between the objective attained by an optimal policy and that attained by a given upper bound.

5.7.1 A Toy Example

We begin with a toy EMS system, for which the optimal policy (and its corresponding objective value) can be computed exactly. Its service area, which we illustrate in Figure 5.1, is a 5 mile $\times$ 5 mile square region, which we subdivide into four 2.5 mile $\times$ 2.5 mile cells with demand concentrated at their centroids. As in our experiments in Section 4.5, we assume that ambulances travel at 30 mph, that the chute time is 1 minute, and that the response time threshold t_{resp} is 9 minutes. Thus, an ambulance can provide a timely response to a call if it is based at or adjacent to the call's location.

In our computational study from Section 4.5, we assumed that arrivals follow a Poisson process and that service times are Weibull distributed. We do the same here, but we must adapt these distributions to a discrete-time setting. In particular, for each time period t, we assume that a call arrives into our system with probability $p = \sum_{i \in N} p_{ti}$, and the location to which this call arrives follows the distribution illustrated in Figure 5.1. That is, our arrival process is a discrete-time analogue of a homogeneous Poisson process. Similarly, we consider a service time distribution that is a discretization of a Weibull random variable with scale parameter 20 and shape parameter 3 (which has a mean of 17.9 min. and a standard deviation of 6.5 min.), in that we round service times up to the nearest multiple of t.

We again consider a horizon length of 1 day, and divided this into 1-minute

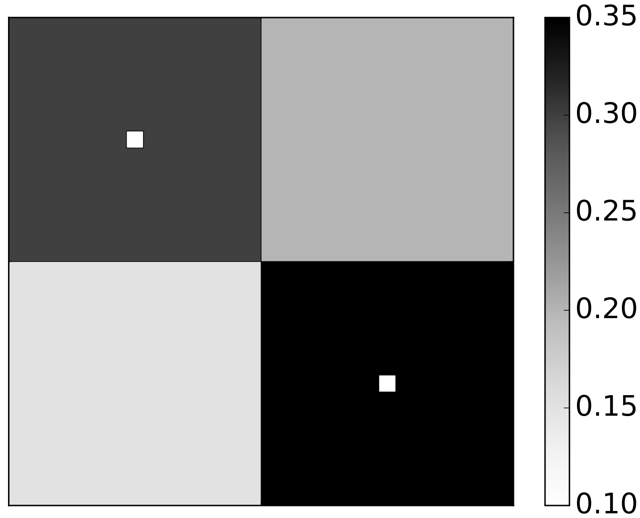

Figure 5.1: Distribution of demand and base locations (white squares) in our toy EMS system. Darker cells indicate areas with higher demand (larger p_{ti}–values).

increments; thus, $T = 1440$. We study fleets containing 1 and 2 ambulances, as well as two arrival regimes, to examine the performance of our bound in lightly-loaded and moderately-loaded systems. To compute the performance V^* attained by an optimal policy in each of the four systems, we solve the corresponding DP instances using the optimality equations (DP). We compare this to a sampled estimate of the perfect information upper bound, which we obtain by generating 5000 i.i.d. replications $\omega_1, \ldots, \omega_{5000}$, solving the corresponding instances of (IP2(ω)), and averaging the resulting optimal objective values.

To estimate our "information penalty" upper bound from Section 5.6, we formulate an instance of the convex optimization problem (5.44), using replications $\omega_1, \ldots, \omega_{5000}$. We find a near-optimal choice of coefficients γ by applying 10 iterations of gradient descent, with an initial step size of 5.0. We step in the direction of the gradient only if it improves the objective function; otherwise, we halve our step size and repeat. Once this procedure completes, we obtain an unbiased estimate of our upper bound by solving generating 5000 new i.i.d. replicates, and solving the corresponding instances of (PIP(ω)). To provide a baseline, also compare our upper bounds to the loss system bound from Chapter 4, which we compute as in Section 4.5. Table 5.1 summarizes our results.

At first glance, the system appears to perform poorly when $A = 1$, but this can be attributed to the fact that when only one ambulance is available, it can only provide timely responses to calls at most 70% of the time, and so 0.700 is a theoretical upper bound on system performance. Our loss system upper bound is almost tight, and yields duality gaps on the order of 1%. By contrast, our perfect information bound is fairly loose, which is unsurprising, given the numerical results in Chapter 4. However, our information penalty bound only marginally improves upon the perfect information upper bound.

	A = 1 ambulance	
	p = 1/300 (0.2 calls/hr)	p = 1/200 (0.3 calls/hr)
Optimal Value	0.644	0.619
Loss System Upper Bound	0.649 $\pm$ 0.002	0.627 $\pm$ 0.002
Perfect Info Upper Bound	0.849 $\pm$ 0.003	0.845 $\pm$ 0.002
Penalized Upper Bound	0.849 $\pm$ 0.003	0.832 $\pm$ 0.002
	A = 2 ambulances	
	p = 1/80 (0.75 calls/hr)	p = 1/40 (1.5 calls/hr)
Optimal Value	0.907	0.807
Loss System Upper Bound	0.907 $\pm$ 0.001	0.817 $\pm$ 0.001
Perfect Info Upper Bound	0.963 $\pm$ 0.002	0.902 $\pm$ 0.001
Penalized Upper Bound	0.958 $\pm$ 0.002	0.902 $\pm$ 0.001

Table 5.1: 95% confidence intervals of our upper bounds for the toy EMS system in Figure 5.1, averaged over 5000 i.i.d. replications.

Our bound's poor performance can be attributed to the fact that our penalty function is overparameterized. Because $T = 1440$ and $|B| = 2$, finding the optimal coefficients γ entails a search in 2880–dimensional space. This presents computational issues because the objective function to our optimization problem is a noisy (although unbiased) estimate of the true objective $\mathbb{E}\left[V^{PIP}(\gamma)\right]$. Thus, 5000 replications may be insufficient to mitigate the effects of noise when optimizing in such a high-dimensional space.

To address these concerns, we consider two slightly less general classes of penalties, by imposing constraints on the values that γ_{tj} can take on. First, we consider a family of penalties where for each time t, we require $\gamma_{tj} = \gamma_{tk}$ for all bases j and k— that is, penalties of the form

$$z_t^{cvg,\gamma}\left(s_t,\, i_t,\, \pi_t,\, \omega\right) = \gamma_{t+1} \sum_{j \in B} \Big[\mathbb{1}\big(j \in I_{t+1}(\omega)\big) -\, \rho_{t+1,j}\Big] y_{t+1,\,j}. \tag{5.46}$$

We call this our *time-dependent penalty function*. Finding the best such penalty again entails solving a convex optimization problem, which can be shown by a proof nearly identical to that used for Lemma 5.6.5. We can again find near-optimal coefficients $(\gamma_1, \ldots, \gamma_{T+1})$ via gradient search, using an expression similar to (5.45) to evaluate the gradient of the objective at a particular value of γ. Second, we consider a class of penalties where for each base j, we require $\gamma_{tj} = \gamma_{t'j}$ for all time periods t and t'— that is, penalties of the form

$$z_t^{cvg,\gamma}\left(s_t,\, i_t,\, \pi_t,\, \omega\right) = \sum_{j \in B} \gamma_j \Big[\mathbb{1}\big(j \in I_{t+1}(\omega)\big) -\, \rho_{t+1,j}\Big] y_{t+1,\,j}. \tag{5.47}$$

We call this our *base-dependent penalty function*, for which we can again find the tightest such upper bound using gradient descent. Applying our modified bounds to the previous problem instances, we obtain Table 5.2.

Our base-dependent penalty function consistently yields the tightest upper bound, and closes a significant portion of the gap between the perfect information upper bound and the optimal objective value. Although it is still outper-

	$A = 1$ ambulance	
	$p = 1/300$ (0.2 calls/hr)	$p = 1/200$ (0.3 calls/hr)
Optimal Value	0.644	0.619
Loss System Upper Bound	0.649 ± 0.002	0.627 ± 0.002
Perfect Info Upper Bound	0.849 ± 0.003	0.845 ± 0.002
Penalized Upper Bound	0.846 ± 0.003	0.839 ± 0.002
Time-Dependent Penalty Bound	0.849 ± 0.003	0.832 ± 0.002
Base-Dependent Penalty Bound	0.666 ± 0.000	0.672 ± 0.000
	$A = 2$ ambulances	
	$p = 1/80$ (0.75 calls/hr)	$p = 1/40$ (1.5 calls/hr)
Optimal Value	0.907	0.807
Loss System Upper Bound	0.907 ± 0.001	0.817 ± 0.001
Perfect Info Upper Bound	0.963 ± 0.002	0.902 ± 0.002
Penalized Upper Bound	0.963 ± 0.002	0.902 ± 0.001
Time-Dependent Penalty Bound	0.958 ± 0.001	0.884 ± 0.001
Base-Dependent Penalty Bound	0.944 ± 0.001	0.858 ± 0.000

Table 5.2: 95% confidence intervals of our alternative upper bounds for the toy EMS system in Figure 5.1, averaged over 5000 i.i.d. replications.

formed by our loss system upper bound, it is reasonably tight, and yields duality gaps on the order of 5%. It is worth mentioning that given a sufficiently large sample (to mitigate the effects of noise), our original penalty function should, in theory, outperform our modified penalties, as the optimal time-dependent and base-dependent penalties yield feasible solutions to the optimization problem (5.44). However, even with 5000 samples, optimizing in a lower-dimensional space appears to improve our bound's performance.

5.7.2 Small-Scale Systems

We now turn our attention to more reasonably-sized EMS systems, and begin with the small-scale system from Section 4.5.2.1, for which we reproduce the illustration in Figure 5.2. Recall that we considered an EMS system operating a fleet of 4 ambulances within a 9 mile $\times$ 9 mile service area, subdivided into 81 square cells. Ambulances again travel at 30 mph, and the chute time is again 1 minute. Arrivals follow a homogeneous Poisson process, that we discretize as in the previous section. Service times are follow a Weibull distribution with scale parameter 30 and shape parameter 3, rounded up to the nearest integer. This time, the curse of dimensionality prevents us from solving (DP) exactly, but we can obtain some indication of the duality gap by again using our heuristic policy from Section 4.5.1.1 as a lower bound. Estimating our bounds using 5000 i.i.d. replications, we obtain Table 5.3.

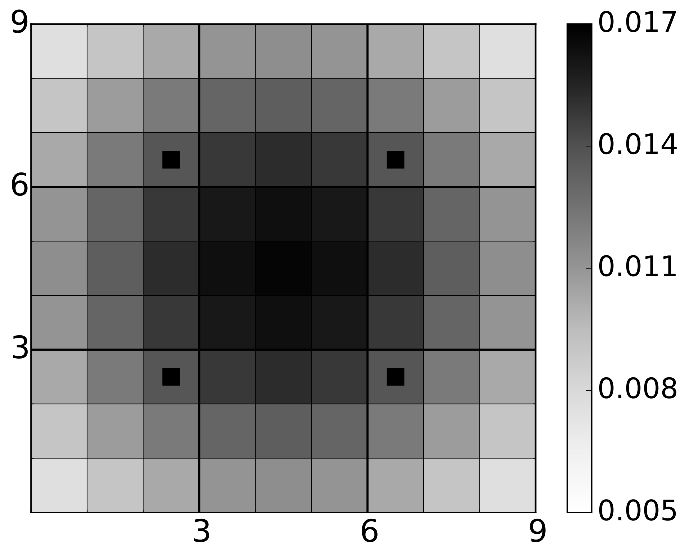

Figure 5.2: Distribution of demand and base locations (squares) in our first small-scale EMS system. Darker cells indicate areas with higher demand (larger p_i–values).

There are slight discrepancies between our numerical results in Table 5.3 and those from Section 4.5.2.1. However, they can be attributed to the fact that we modeled the system in continuous time in Chapter 4. Once again, our base-dependent penalty outperforms our other two penalty functions, but the improvement over the perfect information bound is much more modest; in both cases, the penalty closes only roughly 10% of the duality gap. Moreover, our loss

	$A = 1$ ambulance	
	$p = 1/30$ (2 calls/hr)	$p = 1/20$ (3 calls/hr)
Heuristic Lower Bound	0.810 ± 0.001	0.691 ± 0.001
Loss System Upper Bound	0.843 ± 0.001	0.752 ± 0.000
Perfect Info Upper Bound	0.965 ± 0.001	0.911 ± 0.001
Penalized Upper Bound	0.965 ± 0.001	0.890 ± 0.001
Time-Dependent Penalty Bound	0.958 ± 0.001	0.871 ± 0.000
Base-Dependent Penalty Bound	0.952 ± 0.001	0.869 ± 0.000

Table 5.3: 95% confidence intervals of our lower and upper bounds for our first small-scale EMS system, averaged over 5000 i.i.d. replications.

system bound again outclasses our information penalty bounds— this time, by a much wider margin. This drop in performance (relative to our toy EMS system) can likely be attributed to two factors: a larger service area (which increases the value of perfect information, and widens the duality gap) and a larger fleet (which increases the degree to which resources can be pooled).

To test these hypotheses, we study an (unrealistic) EMS system, illustrated in Figure 5.3, in which resources cannot be effectively pooled. Although the service area, as in our first small-scale system, is a 9 mile $\times$ 9 mile square, demand is much more dispersed, and ambulances travel at half the speed (15 mph). Thus, the system in Figure 5.3 can be viewed as a 3×3 grid of closed subsystems, each with a base at its centroid. Ambulances stationed at a base can only provide timely responses to demand nodes within the same subsystem, but can be redeployed between subsystems. Service times again follow a (discretized) Weibull distribution with scale parameter 30 and shape parameter 3.

We consider fleets containing 5 and 7 ambulances, operating under situations in which the fleet is lightly-loaded and moderately-loaded. Computing our bounds as before, we obtain Table 5.4.

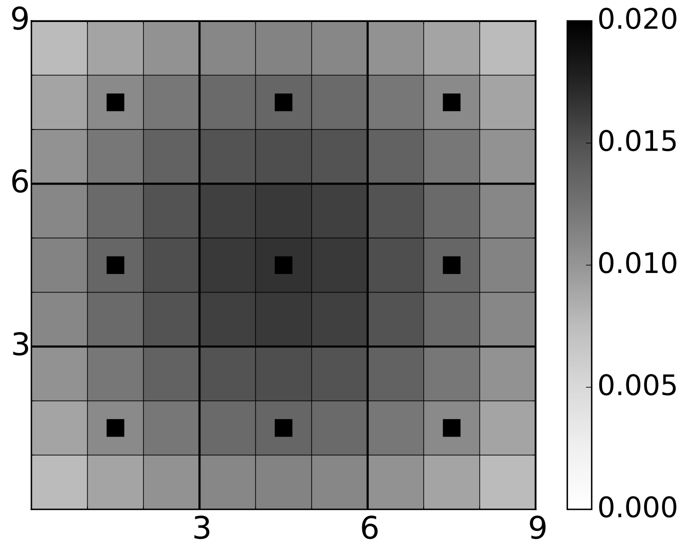

Figure 5.3: Distribution of demand and base locations (squares) in our second small-scale EMS system.

	$A = 5$ ambulances	
	$p = 1/60$ (1 call/hr)	$p = 1/30$ (2 calls/hr)
Heuristic Lower Bound	0.602 ± 0.001	0.502 ± 0.001
Loss System Upper Bound	0.650 ± 0.001	0.572 ± 0.001
Perfect Info Upper Bound	0.944 ± 0.002	0.934 ± 0.001
Base-Dependent Penalty Bound	0.708 ± 0.000	0.658 ± 0.000
	$A = 7$ ambulances	
	$p = 1/30$ (2 calls/hr)	$p = 1/20$ (3 calls/hr)
Heuristic Lower Bound	0.707 ± 0.001	0.620 ± 0.001
Loss System Upper Bound	0.758 ± 0.001	0.701 ± 0.000
Perfect Info Upper Bound	0.978 ± 0.001	0.971 ± 0.001
Base-Dependent Penalty Bound	0.864 ± 0.000	0.806 ± 0.000

Table 5.4: 95% confidence intervals of our lower and upper bounds for our second small-scale EMS system, averaged over 5000 i.i.d. replications.

Our base-dependent penalty function again outperforms our other two penalty functions, and so we omit the latter bounds from Table 5.4. Although the loss system bound is again tighter than our information-based bound, our penalty indeed appears to be more effective at closing the duality gap in systems less capable of resource pooling. This effect also appears to diminish as we grow the fleet, as our bound is weaker when $A = 7$ than when $A = 5$. Taken together, our experiments suggest that our information penalty bound is well-suited to small-scale systems, which tend to operate smaller fleets, and to spread their ambulances further geographically.

5.7.3 A Realistic System

Given that our information penalty bound appears to diminish in effectiveness with larger systems, a natural question to ask is how the bound performs with more realistically-sized problem instances. In this vein, we next consider a graph based upon a simplification of the road network in Edmonton, Canada. Our graph is due to Maxwell et al. [61], but is smaller than the one we study in Section 4.5.2.2. The system we study here contains 189 nodes, and again operates 16 ambulances distributed across 11 bases. The travel speed of ambulances in the system, chute time, arrival process, and service time distribution are as in the previous section. Computing our bounds in the usual way, and averaging over 5000 i.i.d. replications, we obtain Table 5.5.

	A = 16 ambulances		
	p = 1/10 (6 calls/hr)	p = 2/15 (8 calls/hr)	p = 1/10 (10 calls/hr)
Heuristic Lower Bound	0.901 ± 0.001	0.879 ± 0.001	0.844 ± 0.001
Loss System Upper Bound	0.927 ± 0.001	0.927 ± 0.001	0.923 ± 0.000
Perfect Info Upper Bound	0.926 ± 0.001	0.928 ± 0.001	0.928 ± 0.000
Base-Dependent Penalty Bound	0.926 ± 0.001	0.928 ± 0.001	0.928 ± 0.000

Table 5.5: 95% confidence intervals of our lower and upper bounds on a system modelled after Edmonton EMS, averaged over 5000 i.i.d. replications.

This time, our information penalty fails to improve upon the perfect information upper bound. When we attempt to solve the minimization problem (5.44), moving against the gradient does not decrease the objective, indicating that we attain our tightest upper bound when $\gamma = 0$— that is, in the absence of a penalty. Although the loss system bound also appears to struggle in this setting, Table 5.5 confirms the insights we drew from our small-scale systems. Edmonton operates a reasonably large fleet, and resources can pool easily; six ambulances are sufficient to cover 90% of demand (that is, to respond promptly to the next call arrival with probability 0.9).

This degree of resource pooling may not be anomalous, as increased call volumes make redundant coverage a necessity in urban areas, suggesting that our penalty function would be just as ineffective when applied to large-scale EMS systems. Indeed, we obtain similar results when we study the exploratory EMS systems we study in Section 4.5.2.3; our penalty function again fails to improve upon the perfect information upper bound. We omit the corresponding numerical results for brevity.

5.7.4 Discussion

The above computational experiments suggest that the value function approximation that we use in building our penalty is poor. We hypothesize that the reason for this is twofold. First, a one-step look-ahead approximation may not adequately capture the effects that decisions have on the system's ability to respond to future calls. Second, we model coverage as being a linear function of the number of idle ambulances in the system, and in practice, there are diminishing marginal returns associated with redundant coverage. One may attempt to address the former concern by a value function approximation with a wider look-ahead window. However, even a two-step look-ahead approximation is challenging to implement, as for a given time t, the state of the system at time $t+2$ is a function of the decisions at time $t+1$, which in turn is a function of the decisions made at time t; this results in nonlinear interactions.

To address the latter concern, one might consider a value function approximation that is concave in the number of idle ambulances in the system. Maxwell et al. [60] construct one such approximation based upon the Erlang loss formula. Although the resulting penalty can be written in closed-form using the decision variables in (IP2(ω)), the objective function of the corresponding penalized problem becomes nonlinear.

Implementing either approach would entail finding or developing a procedure with which the resulting nonlinear optimization problem (with integer decision variables) is computationally tractable. Alternatively, there may exist a more effective penalty that can be written as a linear function of the decision variables in (IP2(ω)). We believe this to be unlikely, as we considered several variations of the one-step look-ahead penalty without improving performance. However, we propose both as directions for future research.

5.8 Conclusion

In this chapter, we devised an upper bound on the performance of optimal ambulance redeployment policies by applying the "information penalty" framework of Brown et al. [20]. We begin by formulating the EMS provider's decision problem as a high-dimensional stochastic dynamic program, then demonstrate that a relaxation of this problem in which the decision-maker has perfect information is tractable, and can be solved via an equivalent integer program (that we conjecture to not have an integrality gap). To tighten the perfect information upper bound, we introduce a penalty function that affects policies using future information to make decisions (but not nonanticipative policies). We introduce a parametrized family of penalties (based upon a one-step look-ahead value function approximation) for which the penalized problem is no more difficult than to solve than the perfect information relaxation. Furthermore, we show that the penalty in this family yielding the tightest upper bound can be found by solving a convex optimization problem.

Numerical experiments, however, have yielded fairly discouraging results. Our information penalty upper bound appears to be fairly effective in small-scale systems, particularly in situations to which resources cannot be effectively pooled. However, it is consistently outperformed by our loss system bound from Chapter 4. Furthermore, when we consider larger, more realistic EMS systems, the penalty bound fails to improve even upon the perfect information bound. We conjecture that a tighter bound can be obtained by basing our penalty upon a more accurate value function approximation, but this presents technical challenges, as developing such a bound would likely entail solving a nonlinear integer optimization problem.

APPENDIX A

APPENDIX FOR CHAPTER 2

A.1 Proof of Lemma 2.5.4

A.1.1 Preliminaries

We make use of the following intermediate result, which states that in a finite-horizon setting, the benefit associated with being in a "better" system state increases with the length of the horizon.

Lemma A.1.1. *For all* $n \geq 1$, $\alpha \in [0, 1)$, *and applicable* i *and* j, *we have*

1. $v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i+1,\, j) \geq v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i+1,\, j)$
2. $v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j+1) \geq v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j+1)$
3. $v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i+1,\, j) \geq v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i+1,\, j)$

Proof. We verify Statement 1 with a sample path argument; the proofs of Statements 2 and 3 are similar. Start two processes on the same probability space: one in state $(i,\, j)$ and one in state $(i{+}1,\, j)$, each with n periods remaining in the horizon. Let Δ_n denote the difference in reward collected by the two processes until either coupling occurs or n time periods have elapsed. We show $\Delta_n \leq \Delta_{n+1}$ pathwise. On sample paths ω where coupling occurs in the first n periods, $\Delta_n(\omega) = \Delta_{n+1}(\omega)$. On all other paths, Process 1 has at least as many servers available as Process 2 after n periods, and collects a reward in the remaining period at least as large as that by Process 2. Thus, $\Delta_n(\omega) \leq \Delta_{n+1}(\omega)$. □

To prove Lemma 2.5.4, it suffices to show that for every n that

$$v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j+1) \le v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+2) \tag{A.1}$$

$$v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j+1) \le v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) \tag{A.2}$$

$$v_{n,\alpha}(i,\, N_B) - v_{n,\alpha}(i+1,\, N_B) \le v_{n,\alpha}(i+1,\, N_B) - v_{n,\alpha}(i+2,\, N_B) \tag{A.3}$$

$$v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) \le v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+2), \tag{A.4}$$

and we proceed via induction over the time periods. The base case ($n = 0$) is trivial, as we assume $v_{0,\alpha} = 0$. Now suppose that (A.1)–(A.4) hold over horizons of length up to n: our induction hypothesis. In the analysis that follows, we assume, for convenience, that $\alpha = 1$, allowing us to suppress α in our arguments; nearly identical reasoning can be used for the case where $\alpha < 1$.

A.1.2 Inductive Step, Inequality (A.1)

Fix $i \in \{0,\, 1,\, \ldots,\, N_A\}$ and $j \in \{0,\, 1,\, \ldots,\, N_B - 2\}$. We want to show that

$$v_{n+1,\,\alpha}(i,\, j) - v_{n+1,\,\alpha}(i,\, j+1) \le v_{n+1,\,\alpha}(i,\, j+1) - v_{n+1,\,\alpha}(i,\, j+2). \tag{A.5}$$

Using finite-horizon analogues of the optimality equations (2.3), we can write

$$\begin{aligned}
&v_{n+1,\,\alpha}(i,\, j) - v_{n+1,\,\alpha}(i,\, j+1) \qquad\qquad (\text{A.6})\\
&\qquad = \lambda_H\Big[\mathbf{1}_{\{i<N_A\}}\big(v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1)\big)\\
&\qquad\qquad + \mathbf{1}_{\{i=N_A\}}\big(v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+2)\big)\Big]\\
&\qquad\quad + \lambda_L\Big[\max\big\{R_L + v_{n,\alpha}(i,\, j+1), v_{n,\alpha}(i,\, j)\big\}\\
&\qquad\qquad - \max\big\{R_L + v_{n,\alpha}(i,\, j+2), v_{n,\alpha}(i,\, j+1)\big\}\Big]\\
&\qquad\quad + i\mu\big[v_{n,\alpha}(i-1,\, j) - v_{n,\alpha}(i-1,\, j+1)\big]\\
&\qquad\quad + j\mu\big[v_{n,\alpha}(i,\, j-1) - v_{n,\alpha}(i,\, j)\big]\\
&\qquad\quad + \mu\big[v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j)\big]\\
&\qquad\quad + (N_A + N_B - i - j - 1)\mu\big[v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j+1)\big]
\end{aligned}$$

It suffices to show that each term in brackets on the right-hand side of (A.6) is bounded above by $v_{n+1,\,\alpha}(i,\, j+1) - v_{n+1,\,\alpha}(i,\, j+2)$. Consider the first term. If $i < N_A$, we have that

$$\begin{aligned} v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) &\leq v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+2) \\ &\leq v_{n+1,\,\alpha}(i,\, j+1) - v_{n+1,\,\alpha}(i,\, j+2), \end{aligned}$$

where the first inequality follows by (A.4) and our induction hypothesis, and the second by Lemma A.1.1. Similar reasoning applies if $i = N_A$. Now consider the second term. There are three possibilities:

1. $R_L + v_{n,\alpha}(i,\, j+1) \geq v_{n,\alpha}(i,\, j)$ and $R_L + v_{n,\alpha}(i,\, j+2) < \alpha v_{n,\alpha}(i,\, j+1)$,
2. $R_L + v_{n,\alpha}(i,\, j+1) \geq v_{n,\alpha}(i,\, j)$ and $R_L + v_{n,\alpha}(i,\, j+2) \geq v_{n,\alpha}(i,\, j+1)$,
3. $R_L + v_{n,\alpha}(i,\, j+1) < v_{n,\alpha}(i,\, j)$ and $R_L + v_{n,\alpha}(i,\, j+2) < v_{n,\alpha}(i,\, j+1)$.

We cannot have $R_L + v_{n,\alpha}(i,\, j+1) < v_{n,\alpha}(i,\, j)$ and $R_L + v_{n,\alpha}(i,\, j+2) \geq v_{n,\alpha}(i,\, j+1)$, as by the induction hypothesis, $v_{n,\alpha}$ is convex in j. Consider the first case; the analysis for the latter two cases is straightforward. We have

$$\begin{aligned} &\max\left\{R_L + v_{n,\alpha}(i,\, j+1), v_{n,\alpha}(i,\, j)\right\} - \max\left\{R_L + v_{n,\alpha}(i,\, j+2), v_{n,\alpha}(i,\, j+1)\right\} \\ &\qquad = R_L + v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+1) \\ &\qquad < v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+2) \\ &\qquad \leq v_{n+1,\,\alpha}(i,\, j+1) - v_{n+1,\,\alpha}(i,\, j+2), \end{aligned}$$

where the first inequality follows by assumption, and the second by Lemma A.1.1. Now consider the third term. The induction hypothesis— specifically, inequalities (A.1) and (A.2)— and Lemma A.1.1 yield

$$\begin{aligned} v_{n,\alpha}(i-1,\, j) - v_{n,\alpha}(i-1,\, j+1) &\leq v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j+1) \\ &\leq v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+2) \\ &\leq v_{n+1,\,\alpha}(i,\, j+1) - v_{n+1,\,\alpha}(i,\, j+2). \end{aligned}$$

The remaining terms follow in a similar fashion.

A.1.3 Inductive Step, Inequality (A.2)

Fix $i \in \{0, \ldots, N_A - 1\}$ and $j \in \{0, \ldots, N_B - 1\}$. We want to show that

$$v_{n+1,\alpha}(i,\, j) - v_{n+1,\alpha}(i,\, j+1) \leq v_{n+1,\alpha}(i+1,\, j) - v_{n+1,\alpha}(i+1,\, j+1), \tag{A.7}$$

and proceed using a sample path argument. Start four processes on the same probability space, each with $n+1$ periods remaining in the horizon. Processes 1 and 4 begin in states $(i,\, j)$ and $(i+1,\, j+1)$, respectively, and follow the optimal policy π^*. Processes 2 and 3 begin in states $(i,\, j+1)$ and $(i+1,\, j)$, respectively, and use potentially suboptimal policies π_2 and π_3, respectively. These policies deviate from π^* only during the first time period; we describe them later.

Let Δ be the difference in reward collected by Processes 1 and 2 until they couple; define Δ' analogously for Processes 3 and 4. Equation (2.1) implies

$$\begin{aligned} \mathbb{E}\Delta &= v_{n+1,\alpha}(i,\, j) - v^{\pi_2}_{n+1,\alpha}(i,\, j+1) \\ \mathbb{E}\Delta' &= v^{\pi_3}_{n+1,\alpha}(i+1,\, j) - v_{n+1,\alpha}(i+1,\, j+1) \end{aligned}$$

It suffices to show $\mathbb{E}\Delta \leq \mathbb{E}\Delta'$, as

$$\begin{aligned} v_{n+1,\alpha}(i,\, j) - v_{n+1,\alpha}(i,\, j+1) &\leq v_{n+1,\alpha}(i,\, j) - v^{\pi_2}_{n+1,\alpha}(i,\, j+1) \\ &\leq v^{\pi_3}_{n+1,\alpha}(i+1,\, j) - v_{n+1,\alpha}(i+1,\, j+1) \\ &\leq v_{n+1,\alpha}(i+1,\, j) - v_{n+1,\alpha}(i+1,\, j+1). \end{aligned}$$

There is one Type B server that is busy in Processes 2 and 4, but idle in Processes 1 and 3; call it Server I. We construct our probability space so that Server I completes service in all four processes simultaneously (but triggers a dummy completion in Processes 1 and 3). Similarly, there is one Type A server that is busy in Processes 3 and 4, but idle in Processes 1 and 2; call it Server II. We probabilistically link this server across all four processes as we did with Server I. In the first time period, seven transitions are possible:

1. A Type H arrival
2. A Type L arrival
3. A Type A service completion.
4. A completion by Server I.
5. A completion by Server II.
6. A completion by any other Type B server.
7. A dummy transition due to uniformization.

For $k = 1, \ldots, 7$, let A_k be the event in which the k^{th} transition occurs; these events partition the sample space. By the Tower Property, it suffices to show that $\mathbb{E}[\Delta \mid A_k] \le \mathbb{E}[\Delta' \mid A_k]$ for each k.

Case 1 (Event A_1): If $i + 1 < N_A$, all four processes admit the Type H job with a Type A server, and transition to states $(i + 1,\, j)$, $(i + 1,\, j + 1)$, $(i + 2,\, j)$ and $(i+2,\, j+1)$, respectively. Since all four processes subsequently use π^*, we have

$$\mathbb{E}[\Delta \mid A_1] = v_{n,\alpha}(i + 1,\, j) - v_{n,\alpha}(i + 1,\, j + 1)$$
$$\mathbb{E}[\Delta' \mid A_1] = v_{n,\alpha}(i + 2,\, j) - v_{n,\alpha}(i + 2,\, j + 1).$$

By the induction hypothesis and (A.1), we have $\mathbb{E}[\Delta \mid A_1] \le \mathbb{E}[\Delta' \mid A_1]$, as desired. If $i + 1 = N_A$, but $j + 1 < N_B$, then Processes 3 and 4 must admit the job with Type B servers, and once again, by the induction hypothesis and (A.1),

$$\begin{aligned}\mathbb{E}[\Delta \mid A_1] &= v_{n,\alpha}(i + 1,\, j) - v_{n,\alpha}(i + 1,\, j + 1)\\ &\le v_{n,\alpha}(i + 1,\, j + 1) - v_{n,\alpha}(i + 1,\, j + 2) = \mathbb{E}[\Delta' \mid A_1],\end{aligned}$$

Finally, if $i+1 = N_A$ and $j+1 = N_B$, then Process 3 admits the job with a Type B server, whereas Process 4 must turn the job away. We thus have

$$\begin{aligned}\mathbb{E}[\Delta \mid A_1] &= v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) \\ &\leq R_{HB} + v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+1,\, j+1) = \mathbb{E}[\Delta' \mid A_1],\end{aligned}$$

where the inequality follows this time by Lemma 2.4.2.

Case 2 (Event A_2): Suppose event A_2 occurs. Processes 2 and 3 take (potentially suboptimal) actions based upon those taken by Processes 1 and 4. Specifically:

- If Processes 1 and 4 admit the arriving Type L job, so do Processes 2 and 3.
- If Processes 1 and 4 reject the arriving Type L job, so do Processes 2 and 3.
- If Process 1 admits the job, and Process 4 rejects, then Process 2 rejects and Process 3 admits.

We need not consider the case when Process 1 rejects the job in state $(i,\, j)$, and Process 4 admits the job in state $(i+1,\, j+1)$, as by the induction hypothesis, we can assume that π^* is a monotone switching curve policy. Suppose $j+1 < N_B$. If Processes 1 and 4 both admit the job, then

$$\begin{aligned}\mathbb{E}[\Delta \mid A_2] &= v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+2) \\ &\leq v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+1,\, j+2) = \mathbb{E}[\Delta' \mid A_2],\end{aligned}$$

by (A.2) and the induction hypothesis. If Processes 1 and 4 both reject, all four processes remain in the same state, and we can again leverage (A.2) and the induction hypothesis. Finally, if Process 1 admits and Process 4 rejects, we have

$$\begin{aligned}\mathbb{E}_2[\Delta \mid A_2] &= R_L + v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+1) \\ &= R_L + v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+1,\, j+1) = \mathbb{E}[\Delta' \mid A_2].\end{aligned}$$

The case $j + 1 = N_B$ follows via similar arguments, which we omit for brevity.

Case 3 (Event A_3): The four processes transition to states $(i-1,\, j)$, $(i-1,\, j+1)$, $(i,\, j)$, and $(i,\, j+1)$, respectively, and

$$\mathbb{E}[\Delta \,|\, A_3] = v_{n,\alpha}(i-1,\, j) - v_{n,\alpha}(i-1,\, j+1) \leq v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j+1) = \mathbb{E}[\Delta' \,|\, A_3],$$

by (A.2) and the induction hypothesis.

Case 4 (Event A_4): Processes 1 and 2 both transition to state $(i,\, j)$, and coupling occurs. Processes 3 and 4 transition to state $(i+1,\, j)$, and couple as well. Thus $\mathbb{E}[\Delta \,|\, A_4] = \mathbb{E}[\Delta' \,|\, A_4] = 0$.

Case 5 (Event A_5): Processes 1 and 3 both transition to state $(i,\, j)$, while Processes 2 and 4 both transition to state $(i+1,\, j)$. We thus have $\mathbb{E}[\Delta \,|\, A_5] = \mathbb{E}[\Delta' \,|\, A_5] = v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i+1,\, j)$.

Case 6 (Event A_6): The four processes transition to states $(i,\, j-1)$, $(i,\, j)$, $(i+1,\, j-1)$, and $(i+1,\, j)$, respectively, and

$$\mathbb{E}[\Delta \,|\, A_6] = v_{n,\alpha}(i,\, j-1) - v_{n,\alpha}(i,\, j) \leq v_{n,\alpha}(i+1,\, j-1) - v_{n,\alpha}(i+1,\, j) = \mathbb{E}[\Delta' \,|\, A_6],$$

by (A.2) and the induction hypothesis.

Case 7 (Event A_7): The four processes do not change state, and $\mathbb{E}[\Delta \,|\, A_7] \leq \mathbb{E}[\Delta' \,|\, A_7]$ by (A.2) and the induction hypothesis.

Thus, $\mathbb{E}\Delta \leq \mathbb{E}\Delta'$, as desired. $\square$

A.1.4 Inductive Step, Inequality (A.3)

Fix $i \in \{0, 1, \ldots, N_A - 2\}$. We want to show that

$$v_{n+1,\alpha}(i, N_B) - v_{n+1,\alpha}(i+1, N_B) \leq v_{n+1,\alpha}(i+1, N_B) - v_{n+1,\alpha}(i+2, N_B). \quad \text{(A.8)}$$

As in the inductive proof of (A.1), we can rewrite the right-hand side of (A.8) as

$$\begin{aligned}
&v_{n+1,\alpha}(i, N_B) - v_{n+1,\alpha}(i+1, N_B) && \text{(A.9)}\\
&\quad = \lambda_H\big[v_{n,\alpha}(i+1, N_B) - v_{n,\alpha}(i+2, N_B)\big]\\
&\qquad + \lambda_L\Big[\max\big\{R_L + v_{n,\alpha}(i+1, N_B),\, v_{n,\alpha}(i, N_B)\big\}\\
&\qquad\qquad - \max\big\{R_L + v_{n,\alpha}(i+2, N_B),\, v_{n,\alpha}(i+1, N_B)\big\}\Big]\\
&\qquad + i\mu\big[v_{n,\alpha}(i-1, N_B) - v_{n,\alpha}(i, N_B)\big]\\
&\qquad + \mu\big[v_{n,\alpha}(i, N_B) - v_{n,\alpha}(i, N_B)\big]\\
&\qquad + j\mu\big[v_{n,\alpha}(i, N_B - 1) - v_{n,\alpha}(i+1, N_B - 1)\big]\\
&\qquad + (N_A + N_B - i - j)\mu\big[v_{n,\alpha}(i, N_B) - v_{n,\alpha}(i+1, N_B)\big]
\end{aligned}$$

It again suffices to show that each term on the right-hand side of (A.9) is bounded above by $v_{n+1,\alpha}(i+1, N_B) - v_{n+1,\alpha}(i+2, N_B)$. We consider only the second term, as the analysis of the remaining terms is straightforward. There are again three possibilities:

1. $R_L + v_{n,\alpha}(i+1, N_B) \geq v_{n,\alpha}(i, N_B)$ and $R_L + v_{n,\alpha}(i+2, N_B) < v_{n,\alpha}(i+1, N_B)$,
2. $R_L + v_{n,\alpha}(i+1, N_B) \geq v_{n,\alpha}(i, N_B)$ and $R_L + v_{n,\alpha}(i+2, N_B) \geq v_{n,\alpha}(i+1, N_B)$
3. $R_L + v_{n,\alpha}(i+1, N_B) < v_{n,\alpha}(i, N_B)$ and $R_L + v_{n,\alpha}(i+2, N_B) < v_{n,\alpha}(i+1, N_B)$.

We cannot have $R_L + v_{n,\alpha}(i+1, N_B) < v_{n,\alpha}(i, N_B)$ and $R_L + v_{n,\alpha}(i+1, N_B) \geq v_{n,\alpha}(i, N_B)$, as by the induction hypothesis, $v_{n,\alpha}$ is convex in i when $j = N_B$. Consider the first case; the analysis for the latter two cases is straightforward.

We have

$$\begin{aligned}
&\max\left\{R_L + v_{n,\alpha}(i+1,\ N_B), v_{n,\alpha}(i,\ N_B)\right\} \\
&\qquad\qquad - \max\left\{R_L + v_{n,\alpha}(i+2,\ N_B), v_{n,\alpha}(i+1,\ N_B)\right\} \\
&\qquad = R_L + v_{n,\alpha}(i+1,\ N_B) - v_{n,\alpha}(i+1,\ N_B) \\
&\qquad < v_{n,\alpha}(i+1,\ N_B) - v_{n,\alpha}(i+2,\ N_B) \\
&\qquad \leq v_{n+1,\alpha}(i+1,\ N_B) - v_{n+1,\alpha}(i+2,\ N_B),
\end{aligned}$$

as desired.

A.1.5 Inductive Step, Inequality (A.4)

Fix $i \in \{0,\ \ldots,\ N_A - 1\}$ and $j \in \{0,\ \ldots,\ N_B - 2\}$. We want to show that

$$v_{n+1,\alpha}(i+1,\ j) - v_{n+1,\alpha}(i+1,\ j+1) \leq v_{n+1,\alpha}(i,\ j+1) - v_{n+1,\alpha}(i,\ j+2), \quad \text{(A.10)}$$

and again use a sample path argument. Start four processes on the same probability space, each with $n+1$ periods remaining in the horizon. Processes 1 and 4 begin in states $(i+1,\ j)$ and $(i,\ j+2)$, respectively, and follow the optimal policy π^*. Processes 2 and 3 begin in states $(i+1,\ j+1)$ and $(i,\ j+1)$, respectively, and use potentially suboptimal policies π_2 and π_3, respectively. These policies deviate from π^* only during the first time period, in a way that we specify later.

Let the random variable Θ denote the difference in reward collected by Processes 1 and 2 until coupling occurs; define Θ' analogously for Processes 3 and 4. It suffices to show that $E\Theta \leq E\Theta'$. There are i Type A and j Type B servers that are busy in all four processes, and $N_A - i - 1$ Type A and $N_B - j - 2$ Type B servers that are idle in all four processes. We probabilistically the remaining three servers (one Type A, two Type B) as in Table A.1.

	Server I	Server II	Server III
Process 1, $(i+1,\ j)$	Idle Type B	Busy Type A	Idle Type B
Process 2, $(i+1,\ j+1)$	Busy Type B	Busy Type A	Idle Type B
Process 3, $(i,\ j+1)$	Idle Type B	Busy Type B	Idle Type A
Process 4, $(i,\ j+2)$	Busy Type B	Busy Type B	Idle Type A

Table A.1: Marking scheme for servers in the sample path argument for Equation (A.4).

Note that our marking scheme does not require units to be of the same type in every process, but are linked so that completions of all servers marked as Server I (similarly, Servers II and III) occur simultaneously in all four processes. In the first time period, eight transitions are possible:

1. A Type H arrival
2. A Type L arrival
3. A completion by Server I.
4. A completion by Server II.
5. A completion by Server III.
6. A completion by an unmarked Type A server.
7. A completion by an unmarked Type B server.
8. A dummy transition due to uniformization.

For $k = 1, \ldots 8$, let B_k be the event in which transition k occurs. Again, it suffices to show that $\mathbb{E}[\Theta \mid B_k] \leq \mathbb{E}[\Theta' \mid B_k]$ for each k.

Case 1 (Event B_1): If $i + 1 < N_A$, the analysis is straightforward. If $i + 1 = N_A$, then Processes 1 and 2 assign the Type H job to a Type B server, whereas

Processes 3 and 4 route the job to a Type A server. We have that

$$\begin{aligned}
\mathbb{E}[\Theta \mid B_1] &= v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+1,\, j+2) \\
&= v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+1,\, j+2) = \mathbb{E}[\Theta' \mid B_1].
\end{aligned}$$

Case 2 (Event B_2): Processes 2 and 3 take (potentially suboptimal) actions based upon those taken by Processes 1 and 4. In particular:

- If Processes 1 and 4 admit the arriving Type L job, so do Processes 2 and 3.
- If Processes 1 and 4 reject the arriving Type L job, so do Processes 2 and 3.
- If Process 1 admits the job, and Process 4 rejects, then Process 2 rejects and Process 3 admits.

We need not consider the case when Process 1 rejects the job in state $(i+1,\, j)$, and Process 4 admits the job in state $(i,\, j+2)$, as by the induction hypothesis, we can assume that π^* is a monotone switching curve policy with a slope of at least -1. The analysis for the case where both Processes 1 and 4 reject the job is straightforward, as a dummy transition occurs.

Now suppose Processes 1 and 4 both admit the job. If $j+2 < N_B$, the analysis is again straightforward. If $j+2 = N_B$, then Process 4 assigns a Type A server to the job, whereas all other processes assign Type B servers, and

$$\begin{aligned}
\mathbb{E}[\Theta \mid B_2] &= v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+1,\, j+2) \\
&\leq v_{n,\alpha}(i,\, j+2) - v_{n,\alpha}(i+1,\, j+2) = \mathbb{E}[\Theta' \mid B_2],
\end{aligned}$$

where the inequality follows by Lemma 2.4.1. Finally, consider the case where Process 1 admits the job, and Process 4 rejects. Transitions to states $(i+1,\, j+1)$, $(i+1,\, j+1)$, $(i,\, j+2)$, and $(i,\, j+2)$ occur. Processes 1 and 2 couple, as do Processes 3 and 4, and we have that $\mathbb{E}[\Theta \mid B_2] = \mathbb{E}[\Theta' \mid B_2] = R_L$.

Case 3 (Event B_3): The four processes transition to states $(i+1,\, j)$, $(i+1,\, j)$, $(i,\, j+1)$, and $(i,\, j+1)$, respectively. Processes 1 and 2 couple, as do Processes 3 and 4, and we have that $\mathbb{E}[\Theta \,|\, B_3] = \mathbb{E}[\Theta' \,|\, B_3] = 0$.

Case 4 (Event B_4): The four processes transition to states $(i,\, j)$, $(i,\, j+1)$, $(i,\, j)$, and $(i,\, j+1)$. Processes 1 and 3 couple, as do Processes 2 and 4, and we have $\mathbb{E}[\Theta \,|\, B_4] = \mathbb{E}[\Theta' \,|\, B_4]$.

Case 5 (Event B_5): A dummy transition occurs, and $\mathbb{E}[\Theta \,|\, B_5] \leq \mathbb{E}[\Theta' \,|\, B_5]$ follows immediately from the induction hypothesis.

Case 6 (Event B_6): The four processes transition to states $(i,\, j)$, $(i,\, j+1)$, $(i-1,\, j+1)$, and $(i-1,\, j+2)$, and $\mathbb{E}[\Theta \,|\, B_6] \leq \mathbb{E}[\Theta' \,|\, B_6]$ by the induction hypothesis.

Case 7 (Event B_7): The analysis is identical to that for Case 5.

Thus, $\mathbb{E}\Theta \leq \mathbb{E}\Theta'$, as desired. □

A.2 Proof of Proposition 2.5.7

Fix $\alpha \in (0, 1)$, and suppose

$$R_{HB} \leq R_L \leq R_{HB} + \frac{\mu}{\lambda_L} R_{HB} + \frac{\mu}{\lambda_H}\left(1 + \frac{\mu}{\lambda_L} + \frac{\lambda_H}{\lambda_L}\right) R_{HA}. \tag{A.11}$$

Condition (A.11) is sufficient for the value functions v_α and h to be convex in i:

Lemma A.2.1. *If $R_L > R_{HB}$, then for every $n \geq 0$ and applicable i and j, we have that*

$$v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i+1,\, j) \leq v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+2,\, j) \tag{A.12}$$

$$v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j+1) \leq v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+2) \tag{A.13}$$

$$\begin{aligned} v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i+1,\, j) \leq{}& v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) \\ &+ R_{HA} - R_{HB} \end{aligned} \tag{A.14}$$

We defer the proof until Appendix A.2.1, as it is lengthy, and involves arguments very similar to those used in the proof of Proposition 2.5.4 in Appendix A.1. In proving Lemma A.2.1, we make use of an intermediate result:

Lemma A.2.2. *If condition* (A.11) *holds, then for each* $n \geq 0$*, we have*

$$v_{n,\alpha}(N_A - 1,\, N_B) - v_{n,\alpha}(N_A,\, N_B) \leq R_{HA}. \tag{A.15}$$

Proof. By a fairly unconventional another sample path argument. Start two processes in the same probability space. Process 1 begins in state $(N_A - 1,\, N_B)$ and follows the optimal policy π^*, whereas Process 2 begins in state $(N_A,\, N_B)$ and uses a potentially suboptimal policy π that rejects Type L jobs arriving in any states $(i,\, j)$ where $j = N_B$. Let Δ denote the difference in reward collected by the two processes until coupling occurs. We have $\mathbb{E}\Delta = v_\alpha(N_A - 1,\, N_B) - v_\alpha^\pi(N_A,\, N_B)$, and it suffices to show $\mathbb{E}\Delta \leq R_{HA}$. Both processes move in parallel until one of the following occurs:

1. A service completion from the Type A server that is idle in Process 1, but busy in Process 2.
2. A Type L arrival when $j = N_B$ that Process 1 admits (but Process 2, by assumption, redirects).
3. A Type H arrival when Process 1 is in state $(N_A - 1,\, N_B)$, and Process 2 is in state $(N_A,\, N_B)$.
4. A Type H arrival when the two processes are in states $(N_A - 1, j)$ and $(N_A,\, j)$, for some $j < N_B$.

Let Ω_1, Ω_2, Ω_3, and Ω_4 be the sets of paths on which Events 1, 2, 3, and 4 occur, respectively; these sets partition Ω. We further partition the set Ω_4. After Event

4 occurs, Processes 1 and 2 are in states $(N_A,\, j)$ and $(N_A,\, j+1)$, respectively, and Process 1 has collected $R_{HA} - R_{HB}$ more reward than Process 2. From here, Process 2 switches to a policy that imitates the decisions made by Process 1 (and that no longer rejects jobs in states $(i,\, j)$ where $j = N_B$, unless Process 1 does so). Both processes continue to move in parallel until one of the following occurs.

4.1 A Type L service completion seen by Process 2, but not by Process 1.

4.2 A Type H arrival when Process 1 is in state $(N_A,\, N_B - 1)$, and Process 2 is in state $(N_A,\, N_B)$.

4.3 A Type L arrival when Process 1 is in state $(N_A,\, N_B - 1)$, and Process 2 is in state $(N_A,\, N_B)$.

Let Ω_4^1, Ω_4^2 and Ω_4^3 be the set of paths on which events 4.1, 4.2, and 4.3 occur, respectively. Then

$$\Delta(\omega) = \begin{cases} 0 & \omega \in \Omega_1 \\ R_L & \omega \in \Omega_2 \\ R_{HA} & \omega \in \Omega_3 \\ R_{HA} - R_{HB} & \omega \in \Omega_4^1 \\ R_{HA} & \omega \in \Omega_4^2 \\ R_{HA} + R_L - R_{HB} & \omega \in \Omega_4^3 \end{cases}, \tag{A.16}$$

from which it follows that

$$\begin{aligned} \mathbb{E}\Delta = R_L\mathbb{P}(\Omega_2) + R_{HA}\mathbb{P}(\Omega_3) + (R_{HA} - R_{HB})\mathbb{P}(\Omega_4^1) \\ + R_{HA}\mathbb{P}(\Omega_4^2) + (R_{HA} + R_L - R_{HB})\mathbb{P}(\Omega_4^3) \end{aligned} \tag{A.17}$$

Note that $\Delta(\omega) > R_{HA}$ only when $\omega \in \Omega_4^3$, and so we proceed by showing that $\mathbb{P}(\Omega_4^3)$ is relatively small. Suppose, for the time being, that the following inequalities hold (which we later prove):

$$\mathbb{P}(\Omega_1) \geq \frac{\mu}{\lambda_H}\,\mathbb{P}(\Omega_4) \tag{A.18}$$

$$\mathbb{P}(\Omega_4^1) \geq \frac{\mu}{\lambda_L}\,\mathbb{P}(\Omega_4^3) \tag{A.19}$$

$$\mathbb{P}(\Omega_4^2) = \frac{\lambda_H}{\lambda_L}\,\mathbb{P}(\Omega_4^3) \tag{A.20}$$

We demonstrate that (A.11), when combined with inequalities (A.18)–(A.20), imply $\mathbb{E}\Delta \leq R_{HA}$. Indeed:

$$\begin{aligned}
&R_L \leq R_{HB} + \frac{\mu}{\lambda_L}R_{HB} + \frac{\mu}{\lambda_H}\left(1 + \frac{\mu}{\lambda_L} + \frac{\lambda_H}{\lambda_L}\right)R_{HA} \\
&\quad\iff (R_L - R_{HB})P(\Omega_4^3) \leq \left[\frac{\mu}{\lambda_L}R_{HB} + \frac{\mu}{\lambda_H}\left(1 + \frac{\mu}{\lambda_L} + \frac{\lambda_H}{\lambda_L}\right)R_{HA}\right]P(\Omega_4^3) \\
&\quad\implies (R_L - R_{HB})P(\Omega_4^3) \leq R_{HB}\mathbb{P}(\Omega_4^1) + R_{HA}\frac{\mu}{\lambda_H}\left[\mathbb{P}(\Omega_4^1) + \mathbb{P}(\Omega_4^2) + \mathbb{P}(\Omega_4^3)\right] \\
&\quad\implies (R_L - R_{HB})P(\Omega_4^3) \leq R_{HB}\mathbb{P}(\Omega_4^1) + R_{HA}\mathbb{P}(\Omega_1) \\
&\quad\implies -R_{HA}P(\Omega_1) - R_{HB}\mathbb{P}(\Omega_4^1) + (R_L - R_{HB})P(\Omega_4^3) \leq 0 \\
&\quad\implies -R_{HA}P(\Omega_1) + (R_L - R_{HA})\mathbb{P}(\Omega_2) - R_{HB}\mathbb{P}(\Omega_4^1) + (R_L - R_{HB})P(\Omega_4^3) \leq 0 \\
&\quad\iff \mathbb{E}\Delta - R_{HA} \leq 0,
\end{aligned}$$

where the second line follows by inequalities (A.19) and (A.20), the third by (A.18) and the fact that $\mathbb{P}(\Omega_4) = \mathbb{P}(\Omega_4^1) + \mathbb{P}(\Omega_4^2) + \mathbb{P}(\Omega_4^3)$, and the final line by Equation (A.17).

It remains to show that (A.18) – (A.20) hold. We prove (A.18) here; inequalities (A.19) and (A.20) follow in a similar fashion. Because our MDP is uniformizable, and we have specified the policies by which Processes 1 and 2 operate, we can model them jointly as a discrete-time Markov chain. Randomness in this Markov chain is fully characterized by a sequence of i.i.d. uniform random variables $\{U_n : n \geq 1\}$, where U_i governs the state transition occurring immediately before the i^{th} decision epoch.

However, each sample path ω can be described more succinctly. Recall that in formulating our MDP, we assumed that $\Lambda = \lambda_H + \lambda_L + N_A\mu + N_B\mu = 1$. This allows us to partition the interval $[0,\, 1]$ into one subinterval of width λ_H, one subinterval of width λ_L, and $(N_A{+}N_B)$ subintervals of width μ. We can associate each subinterval with a possible state transition. For instance, we can use the subinterval $[0,\, \lambda_H)$ for transitions resulting from a Type H arrival, and $[\lambda_H + \lambda_L,\, \lambda_H + \lambda_L + \mu)$ for service completions resulting from a specific Type A server in the system (or for a dummy transitions if the server is idle). Thus, a sample path ω can be summarized by the type of transition that occurs in each decision epoch; we describe this using a sequence of random variables $\{X_n : n \geq 1\}$. We allow the X_n to take on a value in the set $\{H,\, L,\, A_1,\, \ldots,\, A_{N_A},\, B_1,\, \ldots,\, B_{N_B}\}$.

Fix $\omega \in \Omega_4$, and let $T(\omega)$ be the time at which Event 4 occurs on this path; note $X_{T(\omega)}(\omega) = H$. Consider a transformed sample path ω' that is identical to ω, except that its $T(\omega)^{\text{th}}$ element is A_1 instead of H. (Assume, without loss of generality, that the first Type A server is the one that is initially idle in Process 1, but busy in Process 2.) Let Ω_1' be the set of all paths in Ω_4 that are transformed in this way; observe that $\Omega_1' \subseteq \Omega_1$. It suffices to show $\mathbb{P}(\Omega_1') = \frac{\mu}{\lambda_H}\mathbb{P}(\Omega_4)$. Indeed:

$$\begin{aligned}
\mathbb{P}(\Omega_1') &= \sum_{n=1}^{\infty} \mathbb{P}(\Omega_1', T = n) \\
&= \sum_{n=1}^{\infty} \sum_{(x_1,\ldots,x_{n-1}):T=n} \mathbb{P}(X_1 = x_1,\, \ldots,\, X_{n-1} = x_{n-1},\, X_n = H) \\
&= \sum_{n=1}^{\infty} \sum_{(x_1,\ldots,x_{n-1}):T=n} \mathbb{P}(X_1 = x_1,\, \ldots,\, X_{n-1} = x_{n-1})\mathbb{P}(X_n = H) \\
&= \sum_{n=1}^{\infty} \sum_{(x_1,\ldots,x_{n-1}):T=n} \mathbb{P}(X_1 = x_1,\, \ldots,\, X_{n-1} = x_{n-1}) \left[\frac{\mu}{\lambda_H}\mathbb{P}(X_n = A_1)\right] \\
&= \sum_{n=1}^{\infty} \frac{\mu}{\lambda_H}\mathbb{P}(\Omega_4,\, T = n) = \frac{\mu}{\lambda_H}\mathbb{P}(\Omega_4)
\end{aligned}$$

where the second line follows because the X_n are i.i.d. Thus, (A.18) holds. $\square$

A.3 Proof of Proposition 2.5.8

For convenience, we refer to the system with rewards $(R_{HA},\, R_{HB},\, R_L)$ as the *original system*, and the system with rewards $(R'_{HA},\, R'_{HB},\, R'_L)$ as the *modified system*. Consider the long-run average reward criterion (the proof for the discounted reward criterion is similar). Because the MDP is unichain, the long-run average reward under any policy is constant and independent of the starting state. Let π_i be the threshold policy in which Type L jobs are admitted in all states $(i',\, N_B)$ where $i' \leq i$. We use π_{-1} to denote the policy that never admits Type L jobs when all Type B servers are busy. Finally, let J_i and J'_i denote the long-run average reward obtained attained by policy π_i under the original and modified systems, respectively.

For the original system, let π_{i^*} denote the best policy int the collection $\{\pi_{-1},\, \pi_0,\, \pi_1,\, \ldots,\, \pi_{N_A}\}$; if multiple policies are optimal, let π_{i^*} denote the one with the largest threshold. To avoid trivialities, assume $i^* < N_A$. By construction, $J_{i^*} > J_i$ for all $i > i^*$. It suffices to show that $J'_{i^*} > J'_i$ for all $i > i^*$ as well. Fix $i > i^*$, and construct two stochastic processes on the same probability space, both under the original system. The process $\{X(t) : t \geq 0\}$ follows policy π_{i^*}, whereas $\{\tilde{X}(t) : t \geq 0\}$ follows π_i. Assume $X(0) = \tilde{X}(0) = (i^* + 1,\, N_B)$. Let $\tau_0 = 0$, and for $n \geq 0$, define the stopping times

$$\tau_{n+1} = \inf\left\{t > \tau_n : (X(t),\, \tilde{X}(t)) = ((i^* + 1,\, N_B),\, (i^* + 1,\, N_B))\right\},$$

the times at which both processes return to state $(i^* + 1,\, N_B)$. By memorylessness, $\{\tau_n : n \geq 1\}$ constitute a renewal process. Let $R(t)$ and $\tilde{R}(t)$ denote the reward collected by processes X and $\tilde{X}$ by time t. Similarly, let R_i and $\tilde{R}_i$ denote the reward collected by X and $\tilde{X}$ during the i^{th} renewal epoch. By the

Renewal Reward Theorem,

$$J_{i^*} = \lim_{t\to\infty} \mathbb{E}\left[\frac{R(t)}{t}\right] = \frac{\mathbb{E}R_1}{\mathbb{E}\tau_1} \qquad J_i = \lim_{t\to\infty} \mathbb{E}\left[\frac{\tilde{R}(t)}{t}\right] = \frac{\mathbb{E}\tilde{R}_1}{\mathbb{E}\tau_1}. \tag{A.21}$$

The above limits are well-defined, as $\{(X(t),\, \tilde{X}(t)) \,:\, t \geq 0\}$ is an irreducible continuous-time Markov chain with finite state space, implying $\mathbb{E}\tau_1 < \infty$. Thus

$$J_{i^*} - J_i = \frac{\mathbb{E}R_1 - \mathbb{E}\tilde{R}_1}{\mathbb{E}\tau_1} > 0. \tag{A.22}$$

Consider $\mathbb{E}R_1 - \mathbb{E}\tilde{R}_1$, the expected difference in reward collected by X and $\tilde{X}$ over a single renewal epoch. Discrepancies between the reward collected by X and $\tilde{X}$ can occur in four ways. Let

1. N_1 be the number of Type L jobs that X rejects due to policy considerations, but $\tilde{X}$ admits, and assigns to Type A servers.
2. N_2 be the number of Type H jobs that X admits, but $\tilde{X}$ rejects because all servers are busy.
3. N_3 be the number of Type H jobs that X assigns to Type A servers, but $\tilde{X}$ is forced to assign to Type B servers.
4. N_4 be the number of Type L jobs that X admits, but $\tilde{X}$ rejects, either because all servers are busy, or due to policy considerations.

We make two claims, that we later prove: that the above list of discrepancies is comprehensive, and that $N_1 \geq N_4$ pathwise. Supposing these to be true for the time being, it follows that

$$\mathbb{E}[R_1 - \tilde{R}_1] = -R_L\mathbb{E}[N_1] + R_{HA}\mathbb{E}[N_2] + (R_{HA} - R_{HB})\mathbb{E}[N_3] + R_L\mathbb{E}[N_4] > 0. \tag{A.23}$$

We can similarly construct two processes $\{X'(t) : t \geq 0\}$ and $\{\tilde{X}'(t) : t \geq 0\}$, under the modified system. Let R_1', and $\tilde{R}_1'$ be defined as in (A.21). Since changing rewards does not affect the evolution of the processes X' and $\tilde{X}'$, we have

$$\begin{aligned}\mathbb{E}[R_1' - \tilde{R}_1'] &= -R_L'\mathbb{E}[N_1] + R_{HA}'\mathbb{E}[N_2] + (R_{HA}' - R_{HB}')\mathbb{E}[N_3] + R_L'\mathbb{E}[N_4] \\ &= R_{HA}'\mathbb{E}[N_2] + (R_{HA}' - R_{HB}')\mathbb{E}[N_3] - R_L'\mathbb{E}[N_1 - N_4] \\ &\geq R_{HA}\mathbb{E}[N_2] + (R_{HA} - R_{HB})\mathbb{E}[N_3] - R_L\mathbb{E}[N_1 - N_4] \\ &> 0.\end{aligned}$$

By an equation analogous to (A.22), $J_{i^*}' > J_i'$, and it is preferable under the modified system to set the threshold to i^* than to any $i > i^*$. It remains to prove the two aforementioned claims: that the events described by the random variables $N_1, \ldots, N_4$ are the only transitions in which processes X and $\tilde{X}$ do not collect the same reward, and that $N_1 \geq N_4$ pathwise.

To show the first claim, it suffices to show that if $\tilde{X}$ admits a Type H (Type L) job with a Type A (Type B) server, X does as well. Let $X_A(t)$ and $X_B(t)$ be the number of free Type A and Type B servers in process X at time t, respectively. Define $\tilde{X}_A(t)$ and $\tilde{X}_B(t)$ analogously. Since X and $\tilde{X}$ are defined on the same probability space, every service completion seen by $\tilde{X}$ is also observed by X. It follows that $X_A(t) \geq \tilde{X}_A(t)$ and $X_B(t) \geq \tilde{X}_B(t)$ for all t on every sample path.

To show the second claim, consider $X_A(t) + X_B(t) - \big[\tilde{X}_A(t) - \tilde{X}_B(t)\big]$. This difference increases by one whenever $\tilde{X}$ admits a Type L job with a Type A server that X is forced to reject (that is, when N_1 increases by one), and decreases by one when X admits a job that $\tilde{X}$ rejects (that is, when N_3 or N_4 increases by one). At the start of any renewal epoch, this difference equals zero, and so we must have $N_1 = N_3 + N_4$. $\square$

A.4 Proof of Proposition 2.5.6

If Type H calls are subject to admission control, the optimality equations must be modified slightly. For brevity, we include only those for the long-run discounted reward criterion.

$$\begin{aligned} v_\alpha(i,j) = \lambda_H \Big[& \mathbf{1}_{\{i<N_A\}} \max\{R_{HA} + \alpha v_\alpha(i+1,\, j),\, \alpha v_\alpha(i,\, j)\} \\ & + \mathbf{1}_{\{i=N_A,\, j<N_B\}} \max\{R_{HB} + \alpha v_\alpha(i,\, j+1),\, \alpha v_\alpha(i,\, j)\} \\ & + \mathbf{1}_{\{i=N_A,\, j=N_B\}} \alpha v_\alpha(i,\, j)\Big] \\ + \lambda_L \Big[& \mathbf{1}_{\{j<N_B\}} \max\{R_L + \alpha v_\alpha(i,\, j+1), \alpha v_\alpha(i,\, j)\} \\ & + \mathbf{1}_{\{i<N_A,\, j=N_B\}} \max\{R_L + \alpha v_\alpha(i+1,\, j),\, \alpha v_\alpha(i,\, j)\} \\ & + \mathbf{1}_{\{i=N_A,\, j=N_B\}} \alpha v_\alpha(i,\, j)\Big] \\ + i\mu\, \alpha v_\alpha(i-1,\, j) & + j\mu\, \alpha v_\alpha(i,\, j-1) + (N_A + N_B - i - j)\mu\, \alpha v_\alpha(i,\, j) \end{aligned} \tag{A.24}$$

In this modified setting, the structural properties stated in Section 2.4 of the main paper all still hold; they follow via identical sample path arguments as the ones presented therein. It suffices to show that $v_{n,\alpha}$ (and consequently, v_α and h) satisfies the following structural properties:

$$v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i+1,\, j) \le v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+2,\, j) \tag{A.25}$$

$$v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j+1) \le v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) \tag{A.26}$$

$$v_{n,\alpha}(N_A,\, j) - v_{n,\alpha}(N_A,\, j+1) \le v_{n,\alpha}(N_A,\, j+1) - v_{n,\alpha}(N_A,\, j+2) \tag{A.27}$$

We proceed via induction on n. The base case ($n = 0$) is trivial, as we assume $v_{\alpha,0}(i,\, j) = 0$ for all i and j. Now suppose that (A.25)—(A.27) hold over horizons of length up to n: our induction hypothesis. In the analysis that follows, we assume, for convenience, that $\alpha = 1$, allowing us to suppress α in our arguments; nearly identical reasoning can be used for the case where $\alpha < 1$.

A.4.1 Inductive Step, Inequality (A.25)

Fix $i \in \{0, \ldots, N_A - 2\}$ and $j \in \{0, \ldots, N_B\}$. We want to show that

$$v_{n+1,\alpha}(i,\, j) - v_{n+1,\alpha}(i+1,\, j) \leq v_{n+1,\alpha}(i+1,\, j) - v_{n+1,\alpha}(i+2,\, j), \tag{A.28}$$

and proceed using a sample path argument. Start four processes on the same probability space, each with $n+1$ periods remaining in the horizon. Processes 1 and 4 begin in states $(i,\, j)$ and $(i+2,\, j)$, respectively, and follow the optimal policy π^*. Processes 2 and 3 both begin in state $(i+1,\, j)$, and use potentially suboptimal policies π_2 and π_3, respectively, that deviate from π^* only during the first time period. Specifically, Processes 2 and 3 take actions that depend on those made by Processes 1 and 4 when a job arrives:

- If Processes 1 and 4 admit the arriving job, so do Processes 2 and 3.
- If Processes 1 and 4 reject the arriving job, so do Processes 2 and 3.
- If Process 1 admits the job, and Process 4 rejects, then Process 2 rejects and Process 3 admits.

We need not consider the case where Process 1 rejects in state $(i,\, j)$ and Process 4 admits in state $(i+2,\, j)$, as by the induction hypothesis, we can assume that π^* is a monotone switching curve policy. Let Δ be the difference in reward collected by Processes 1 and 2 until coupling occurs; define Δ' analogously for Processes 3 and 4. We show that $\mathbb{E}\Delta \leq \mathbb{E}\Delta'$. There are i Type A and j Type B servers that are busy in all four processes, and $N_A - i - 2$ Type A and $N_B - j$ Type B servers that are idle in all four processes. We probabilistically link the remaining two Type A servers according to the scheme in Table A.2.

Note we probabilistically link servers linked so that completions of all servers marked as Server I (similarly, Server II) occur simultaneously in all four processes. Eight transitions are possible in the first time period:

	Server I	**Server II**
Process 1, State $(i,\ j)$	Idle	Idle
Process 2, State $(i+1,\ j)$	Busy	Idle
Process 3, State $(i+1,\ j)$	Idle	Busy
Process 4, State $(i+2,\ j)$	Busy	Busy

Table A.2: Marking scheme for units in the sample path argument for Equation (A.25).

1. A Type H arrival
2. A Type L arrival
3. A completion by Server I.
4. A completion by Server II.
5. A completion by any other Type A server.
6. A Type A service completion.
7. A dummy transition due to uniformization.

For $k = 1, \ldots 7$, let A_k be the event in which transition k occurs; it suffices to show that $\mathbb{E}[\Delta \mid A_k] \leq \mathbb{E}[\Delta' \mid A_k]$ for each k. We proceed case-by-case.

Case 1 (Event A_1): If Processes 1 and 4 both admit the job, then by construction of policies π_2 and π_3, Processes 2 and 3 do so as well, and

$$\mathbb{E}[\Delta \mid A_1] = v_{n,\alpha}(i+1,\ j) - v_{n,\alpha}(i+2,\ j) \leq v_{n,\alpha}(i+2,\ j) - v_{n,\alpha}(i+3,\ j) = \mathbb{E}[\Delta' \mid A_1].$$

by the induction hypothesis and (A.25). If Processes 1 and 4 both reject, a dummy transition occurs, and the analysis is straightforward. Finally, if Process 1 admits, and Process 4 rejects, then

$$\mathbb{E}[\Delta \mid A_1] = v_{n,\alpha}(i+1,\ j) - v_{n,\alpha}(i+1,\ j) \leq v_{n,\alpha}(i+2,\ j) - v_{n,\alpha}(i+2,\ j) = \mathbb{E}[\Delta' \mid A_1].$$

Now suppose $i+2 = N_A$. Process 4 must admit Type H jobs with a Type B server (provided $j < N_B$). If Process 1 and 4 both admit the job, then

$$
\begin{aligned}
\mathbb{E}[\Delta \,|\, A_1] &= v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+2,\, j) \\
&\leq v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+2,\, j+1) \\
&\leq v_{n,\alpha}(i+2,\, j) + R_{HA} - v_{n,\alpha}(i+2,\, j+1) - R_{HB} = \mathbb{E}[\Delta' \,|\, A_1].
\end{aligned}
$$

The inequalities follow by the induction hypothesis and (A.26), respectively.

Case 2 (Event A_2): If $j < N_B$, then all four processes admit the Type L job, and the analysis is straightforward. If $j = N_B$, then admitting in all four processes entails using a Type A server. If Processes 1 and 4 both admit the job, then

$$
\mathbb{E}[\Delta \,|\, A_2] = v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+2,\, j) \leq v_{n,\alpha}(i+2,\, j) - v_{n,\alpha}(i+3,\, j) = \mathbb{E}[\Delta' \,|\, A_2],
$$

by (A.25) and the induction hypothesis. If Processes 1 and 4 both reject, a dummy transition occurs, and the analysis is again straightforward. Finally, if Process 1 admits and Process 4 rejects, we have

$$
\begin{aligned}
\mathbb{E}[\Delta \,|\, A_2] &= R_L + v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j) \\
&= R_L + v_{n,\alpha}(i+2,\, j) - v_{n,\alpha}(i+2,\, j) = \mathbb{E}[\Delta' \,|\, A_2].
\end{aligned}
$$

Case 3 (Event A_3): Processes 1 and 2 both transition to state $(i,\, j)$, and coupling occurs. Processes 3 and 4 transition to state $(i+1,\, j)$, and couple as well. Thus $\mathbb{E}[\Delta \,|\, A_3] = \mathbb{E}[\Delta' \,|\, A_3] = 0$.

Case 4 (Event A_4): Processes 1 and 3 both transition to state $(i,\, j)$, while Processes 2 and 4 both transition to state $(i+1,\, j)$. Thus $\mathbb{E}[\Delta \,|\, A_4] = \mathbb{E}[\Delta' \,|\, A_4] = v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i+1,\, j)$.

Case 5 (Event A_5): The four processes transition to states $(i-1,\, j)$, $(i,\, j)$, $(i,\, j)$, and $(i+1,\, j)$, respectively, and

$$\mathbb{E}[\Delta \mid A_5] = v_{n,\alpha}(i-1,\, j) - v_{n,\alpha}(i,\, j) \leq v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i+1,\, j) = \mathbb{E}[\Delta' \mid A_5],$$

by (A.25) and the induction hypothesis.

Case 6 (Event A_6): The four processes transition to states $(i,\, j-1)$, $(i+1,\, j-1)$, $(i+1,\, j-1)$, and $(i+2,\, j-1)$, respectively, and

$$\begin{aligned}\mathbb{E}[\Delta \mid A_6] &= v_{n,\alpha}(i,\, j-1) - v_{n,\alpha}(i+1,\, j-1) \\ &\leq v_{n,\alpha}(i+1,\, j-1) - v_{n,\alpha}(i+2,\, j-1) = \mathbb{E}[\Delta' \mid A_6],\end{aligned}$$

by (A.25) and the induction hypothesis.

Case 7 (Event A_7): A dummy transition occurs; the analysis is straightforward.

Thus, Equation (A.28) holds, as desired. □

A.4.2 Inductive Step, Inequality (A.26)

Fix $i \in \{0,\, \ldots,\, N_A - 1\}$ and $j \in \{0,\, \ldots,\, N_B - 1\}$. We want to show that

$$v_{n+1,\,\alpha}(i,\, j) - v_{n+1,\,\alpha}(i,\, j+1) \leq v_{n+1,\,\alpha}(i+1,\, j) - v_{n+1,\,\alpha}(i+1,\, j+1). \tag{A.29}$$

The proof follows by a sample path argument nearly identical to that used in the proof of Proposition 2.5.4, to show that the value function in that setting is supermodular. Thus, we restrict attention to points in the proof that deviate from the original proof. We construct four stochastic processes and define random variables Δ and Δ' as before; it again suffices to show that $\mathbb{E}\Delta \leq \mathbb{E}\Delta'$. We mark servers as before, and operate Processes 2 and 3 according to the same suboptimal policies described in Section A.4.1. Of the seven transitions that can occur in the next time period, the analysis for the latter five cases (in which arrivals do not occur) is identical. We consider remaining two possibilities below:

Case 1: Suppose a Type H arrival occurs in the next time period. This time, there is a decision to make in all four processes with respect to the Type H arrival. Consider first the case where $i+1 < N_A$. If Processes 1 and 4 both admit the job, then the four processes transition to states $(i+1,\, j)$, $(i+1,\, j+1)$, $(i+2,\, j)$ and $(i+2,\, j+1)$, respectively, and

$$\begin{aligned}\mathbb{E}[\Delta \mid A_1] &= v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) \\ &\leq v_{n,\alpha}(i+2,\, j) - v_{n,\alpha}(i+2,\, j+1) = \mathbb{E}[\Delta' \mid A_1].\end{aligned}$$

If Processes 1 and 4 both reject the job, dummy transitions occur and

$$\mathbb{E}[\Delta \mid A_1] = v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j+1) \leq v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) = \mathbb{E}[\Delta' \mid A_1].$$

Finally, if Process 1 admits and Process 4 rejects, then

$$\begin{aligned}\mathbb{E}[\Delta \mid A_1] &= v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) \\ &= v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) = \mathbb{E}[\Delta' \mid A_1].\end{aligned}$$

Now suppose $i+1 < N_A$. If Process 1 and 4 admit both admit the job (the latter with a Type B server), so do Processes 2 and 3 (the latter, again, with a Type B server), and by (A.26), we have that

$$\begin{aligned}\mathbb{E}[\Delta \mid A_1] &= v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) \\ &\leq v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+1,\, j+2) = \mathbb{E}[\Delta' \mid A_1].\end{aligned}$$

If Processes 1 and 4 both reject, a dummy transition occurs, and the analysis is straightforward. Finally, if Process 1 admits and Process 4 rejects,

$$\mathbb{E}[\Delta \mid A_1] = v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) = v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+1,\, j+1) = \mathbb{E}[\Delta' \mid A_1].$$

Case 2: Suppose a Type L arrival occurs. Since it is optimal for Type L jobs to be admitted when Type B servers are available, the analysis when $j+1 < N_B$ is straightforward. Suppose $j+1 = N_B$. If Processes 1 and 4 both admit, then

$$\begin{aligned}\mathbb{E}[\Delta \mid A_2] &= v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+1,\, j+1)\\ &\le v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+2,\, j+1) = \mathbb{E}[\Delta' \mid A_2],\end{aligned}$$

by Lemma 2.4.1 and the induction hypothesis. If Processes 1 and 4 both reject, a dummy transition occurs. Finally, if Process 1 admits and Process 4 rejects, then

$$\begin{aligned}\mathbb{E}[\Delta \mid A_2] &= R_L + v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+1)\\ &= R_L + v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+1,\, j+1) = \mathbb{E}[\Delta' \mid A_2].\end{aligned}$$

Thus, Equation (A.26) holds, as desired. □

A.4.3 Inductive Step, Inequality (A.27)

Fix $j \in \{0,\, 1,\, \ldots,\, N_B - 2\}$. We want to show that

$$v_{n+1,\alpha}(N_A,\, j) - v_{n+1,\alpha}(N_A,\, j+1) \le v_{n+1,\alpha}(N_A,\, j+1) - v_{n+1,\alpha}(N_A,\, j+2). \tag{A.30}$$

We can rewrite the left-hand side of (A.30) as

$$\begin{aligned}&v_{n+1,\alpha}(N_A,\, j) - v_{n+1,\alpha}(N_A,\, j+1) \qquad \text{(A.31)}\\ &\quad= \lambda_H\Big[\max\big\{R_{HB} + v_{n,\alpha}(N_A,\, j+1),\, v_{n,\alpha}(N_A,\, j)\big\}\\ &\qquad\qquad - \max\big\{R_{HB} + v_{n,\alpha}(N_A,\, j+2),\, v_{n,\alpha}(N_A,\, j+1)\big\}\Big]\\ &\qquad + \lambda_L\Big[R_L + v_{n,\alpha}(N_A,\, j+1) - R_L - v_{n,\alpha}(N_A,\, j+2)\Big]\\ &\qquad + N_A\mu\big[v_{n,\alpha}(N_A-1,\, j) - v_{n,\alpha}(N_A-1,\, j+1)\big]\\ &\qquad + j\mu\big[v_{n,\alpha}(N_A,\, j-1) - v_{n,\alpha}(N_A,\, j)\big]\\ &\qquad + \mu\big[v_{n,\alpha}(N_A,\, j) - v_{n,\alpha}(N_A,\, j)\big]\\ &\qquad + (N_A + N_B - i - j)\mu\big[v_{n,\alpha}(N_A,\, j) - v_{n,\alpha}(N_A,\, j+1)\big]\end{aligned}$$

In the second term, we leverage the fact that when $R_L > R_{HB}$, it is optimal to admit Type L jobs whenever Type B servers are available. It suffices to show that each term on the right-hand side of (A.31) is bounded above by $v_{n+1,\alpha}(N_A, j+1) - v_{n+1,\alpha}(N_A, j+2)$. Consider the first term. There are three possibilities:

1. $R_{HA}+v_{n,\alpha}(N_A, j+1) \geq v_{n,\alpha}(N_A, j)$ and $R_{HA}+v_{n,\alpha}(N_A, j+2) < v_{n,\alpha}(N_A, j+1)$
2. $R_{HA}+v_{n,\alpha}(N_A, j+1) \geq v_{n,\alpha}(N_A, j)$ and $R_{HA}+v_{n,\alpha}(N_A, j+2) \geq v_{n,\alpha}(N_A, j+1)$
3. $R_{HA}+v_{n,\alpha}(N_A, j+1) < v_{n,\alpha}(N_A, j)$ and $R_{HA}+v_{n,\alpha}(N_A, j+2) < v_{n,\alpha}(N_A, j+1)$

We need not consider the remaining possibility, as by the induction hypothesis, $v_{n,\alpha}$ is convex in j when $i = N_A$. Consider the first case: We have that

$$\begin{aligned}
\max\big\{R_{HA} + v_{n,\alpha}(N_A,\, j+1),\, v_{n,\alpha}(N_A,\, j)\big\}& \\
&- \max\big\{R_{HA} + v_{n,\alpha}(N_A,\, j+2),\, v_{n,\alpha}(N_A,\, j+1)\big\} \\
&= R_{HA} \\
&< v_{n,\alpha}(N_A,\, j+1) - v_{n,\alpha}(N_A,\, j+2) \\
&\leq v_{n+1,\alpha}(N_A,\, j+1) - v_{n+1,\alpha}(N_A,\, j+2)
\end{aligned}$$

where the first inequality follows by assumption, and the second by Lemma A.1.1. Similar reasoning applies for the remaining two possibilities. Now consider the third term on the right-hand side of (A.31). The induction hypothesis—specifically, inequalities (A.25) and (A.26)— yields

$$\begin{aligned}
v_{n,\alpha}(N_A-1,\, j) - v_{n,\alpha}(N_A-1,\, j+1) &\leq v_{n,\alpha}(N_A,\, j) - v_{n,\alpha}(N_A,\, j+1) \\
&\leq v_{n,\alpha}(N_A,\, j+1) - v_{n,\alpha}(N_A,\, j+2) \\
&\leq v_{n+1,\alpha}(N_A,\, j+1) - v_{n+1,\alpha}(N_A,\, j+2).
\end{aligned}$$

The remaining terms follow in a similar fashion.

A.5 Proof of Lemma A.2.1

We begin with an intermediate result:

Lemma A.5.1. *If $R_L > R_{HB}$, then for every $n \geq 0$ and applicable i and j, we have that*

$$v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j+1) \leq \frac{\lambda_H}{\lambda_H + \lambda_L} R_{HB} + \frac{\lambda_L}{\lambda_H + \lambda_L} R_L. \tag{A.32}$$

Inequality (A.32) is slightly stronger than that in Statement 2 of Lemma 2.4.2 (which states that $v_{n,\alpha}(i,\, j) - v_{n,\alpha}(i,\, j+1) \leq R_L$), by leveraging the fact that optimal policies always admit Type L calls when Type B servers are available.

Proof. We use a sample path argument. Fix $\alpha \in [0,\, 1]$, $i \in \{0,\, \ldots,\, N_A\}$, and $j \in \{0,\, \ldots,\, N_B - 1\}$, and construct two stochastic processes on the same probability space. Process 2 begins in state $(i,\, j+1)$ and follows the optimal policy π^*, while Process 1 begins in state $(i,\, j)$ and follows the suboptimal policy that imitates the decisions made by Process 2, with one exception: Process 1 admits any job that arrives when Process 2 is in state $(N_A,\, N_B)$. Let Δ denote the difference in reward collected by the two processes until coupling occurs; it suffices to show

$$\mathbb{E}\Delta \leq \frac{\lambda_H}{\lambda_H + \lambda_L} R_{HB} + \frac{\lambda_L}{\lambda_H + \lambda_L} R_L.$$

Both processes move in parallel until they couple, which can occur in two ways:

1. Process 2 sees a Type B service completion not observed by Process 1.
2. Process 1 is in state $(N_A,\, N_B - 1)$, Process 2 is in state $(N_A,\, N_B)$, and an arrival occurs.

Let A_1 and A_2 be the events in which coupling occurs via the first and second possibilities, respectively. Conditional on A_1 occurring, we have $\Delta = 0$. Conditional on A_2 occurring, with probability $\lambda_H/(\lambda_H + \lambda_L)$, a Type H arrival occurs, and $\Delta = R_{HB}$. Similarly, with probability $\lambda_L/(\lambda_H + \lambda_L)$, a Type L arrival occurs, and $\Delta = R_L$. It follows that

$$E[\Delta \mid A_2] = \frac{\lambda_H}{\lambda_H + \lambda_L} R_{HB} + \frac{\lambda_L}{\lambda_H + \lambda_L} R_L,$$

and we are done. □

To prove Lemma A.2.1, we again use induction over the time periods. The base case ($n = 0$) is trivial, as we assume $v_{\alpha,0}(i, j) = 0$ for all i and j. Now suppose that (A.12)–(A.14) hold over horizons of length $k \leq n$: our induction hypothesis. In the analysis that follows, we assume once again (for notational convenience) that $\alpha = 1$; nearly identical reasoning can be used when $\alpha < 1$.

A.5.1 Inductive Proof, Inequality (A.12)

Fix $i \in \{0, \ldots, N_A - 2\}$ and $j \in \{0, \ldots, N_B\}$. We show that

$$v_{n+1,\alpha}(i, j) - v_{n+1,\alpha}(i+1, j) \leq v_{n+1,\alpha}(i+1, j) - v_{n+1,\alpha}(i+2, j), \tag{A.33}$$

using a sample path argument. Start four processes on the same probability space, each with $n + 1$ periods remaining in the horizon. Processes 1 and 4 begin in states (i, j) and $(i + 2, j)$, respectively, and follow the optimal policy π^*. Processes 2 and 3 begin in state $(i + 1, j)$, and use potentially suboptimal policies π_2 and π_3, respectively, that deviate from π^* during the first time period. The actions Processes 2 and 3 take when a job arrives depend on those taken by Processes 1 and 4:

- If Processes 1 and 4 admit the arriving job, so do Processes 2 and 3.
- If Processes 1 and 4 reject the arriving job, so do Processes 2 and 3.
- If Process 1 admits the job, and Process 4 rejects, then Process 2 rejects and Process 3 admits.

We can ignore the case where Process 1 rejects in state $(i,\ j)$ and Process 4 admits in state $(i+2,\ j)$, as by the induction hypothesis, we can assume that π^* is a monotone switching curve policy.

Let Δ be the difference in reward collected by Processes 1 and 2 until coupling occurs; define Δ' analogously for Processes 3 and 4; it suffices to show $\mathbb{E}\Delta \leq \mathbb{E}\Delta'$. There are i Type A and j Type B servers that are busy in all four processes, and $N_A - i - 2$ Type A and $N_B - j$ Type B servers that are idle in all four processes. We probabilistically link the remaining two Type A servers according to the scheme in Table A.3.

	Server I	**Server II**
Process 1, State $(i,\ j)$	Idle Type A	Idle Type A
Process 2, State $(i+1,\ j)$	Busy Type A	Idle Type A
Process 3, State $(i+1,\ j)$	Idle Type A	Busy Type A
Process 4, State $(i+2,\ j)$	Busy Type A	Busy Type A

Table A.3: Marking scheme for units in the sample path argument for Equation (A.12).

In the first time period, seven transitions are possible:

1. A Type H arrival
2. A Type L arrival
3. A completion by Server I.
4. A completion by Server II.
5. A completion by any other Type A server.
6. A completion by a Type B server.
7. A dummy transition due to uniformization.

For $k = 1, \ldots 7$, let A_k be the event in which transition k occurs. It suffices to show that $\mathbb{E}[\Theta \,|\, A_k] \leq \mathbb{E}[\Theta' \,|\, A_k]$ for each k, and we proceed case-by-case.

Case 1 (Event A_1): If $i + 2 < N_A$, the analysis is straightforward. If $i + 2 = N_A$, Process 4 routes the job to a Type A server, and we have

$$\begin{aligned}\mathbb{E}[\Theta \,|\, A_1] &= v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+2,\, j) \\ &\leq v_{n,\alpha}(i+2,\, j) - v_{n,\alpha}(i+2,\, j+1) + R_{HA} - R_{HB} = \mathbb{E}[\Theta' \,|\, A_1],\end{aligned}$$

by the induction hypothesis and (A.14). Finally, if $i + 2 = N_A$ and $j = N_B$, then Processes 3 and 4 find themselves in state $(i + 2,\, j)$, and by (A.12), we have

$$\mathbb{E}[\Theta \,|\, A_1] = v_{n,\alpha}(i+1,\, j) - v_{n,\alpha}(i+2,\, j) \leq R_{HA} = \mathbb{E}[\Theta' \,|\, A_1].$$

Case 2 (Event A_2): By Proposition 2.4.3, we need only consider the case where $j = N_B$, as all four processes would otherwise admit the arriving Type L job, and the analysis is straightforward. Processes 2 and 3 take (potentially suboptimal) actions based upon those taken by Processes 1 and 4. In particular:

- If Processes 1 and 4 admit the arriving Type L job, so do Processes 2 and 3.
- If Processes 1 and 4 reject the arriving Type L job, so do Processes 2 and 3.
- If Process 1 admits the job, and Process 4 rejects, then Process 2 rejects and Process 3 admits.

We need not consider the remaining possibility, as by the induction hypothesis, we can assume that π^* is a threshold-type policy. The analysis for the case where both Processes 1 and 4 reject the job is straightforward; all four processes either reject the job or admit it with a Type A server, and we leverage the induction hypothesis. For the remaining possibility, the four processes transition to states

$(i+1,\, j)$, $(i+1,\, j)$, $(i+2,\, j)$, and $(i+2,\, j)$, respectively, and Processes 1 and 3 alone collect a reward R_L. Thus, we have that $\mathbb{E}[\Theta \mid A_2] = R_L = \mathbb{E}[\Theta' \mid A_2]$.

Case 3 (Event A_3): Transition to states $(i,\, j)$, $(i,\, j)$, $(i+1,\, j)$, and $(i+1,\, j)$ occur. Processes 1 and 2 couple, as do Processes 3 and 4, and $\mathbb{E}[\Theta \mid A_3] = \mathbb{E}[\Theta' \mid A_3] = 0$.

Case 4 (Event A_4): Transitions to states $(i,\, j)$, $(i+1,\, j)$, $(i,\, j)$, and $(i+1,\, j)$ occur. Processes 1 and 3 couple, as do Processes 2 and 4, and $\mathbb{E}[\Delta \mid A_4] = \mathbb{E}[\Delta' \mid A_4]$.

Case 5 (Event A_5): Transitions to states $(i-1,\, j)$, $(i,\, j)$, $(i,\, j)$, and $(i+1,\, j)$ occur. The induction hypothesis implies $\mathbb{E}[\Delta \mid A_5] = \mathbb{E}[\Delta' \mid A_5]$.

Case 6 (Event A_6): Transitions to states $(i,\, j-1)$, $(i+1,\, j-1)$, $(i+1,\, j-1)$, and $(i+2,\, j-1)$ occur, and the induction hypothesis applies.

Case 7 (Event A_7): A dummy transition occurs, and our induction hypothesis applies.

Thus, Inequality (A.12) holds for horizons of length $n+1$.

A.5.2 Inductive Proof, Inequality (A.13)

Fix $i \in \{0,\, \ldots,\, N_A - 2\}$ and $j \in \{0,\, \ldots,\, N_B\}$. We want to show that

$$v_{n+1,\alpha}(i,\, j) - v_{n+1,\alpha}(i,\, j+1) \leq v_{n+1,\alpha}(i,\, j+1) - v_{n+1,\alpha}(i,\, j+2), \tag{A.34}$$

and proceed using a sample path argument. Start four processes on the same probability space, each with $n+1$ periods remaining in the horizon. Processes 1 and 4 begin in states $(i,\, j)$ and $(i,\, j+2)$, respectively, and follow the optimal policy π^*. Processes 2 and 3 both begin in state $(i,\, j+1)$, and use potentially suboptimal policies π_2 and π_3 that mimic those described in Section A.5.1.

Let Θ be the difference in reward collected by Processes 1 and 2 until coupling occurs; define Θ' analogously for Processes 3 and 4. It suffices to show $\mathbb{E}\Theta \leq \mathbb{E}\Theta'$. There are i Type A and j Type B servers that are busy in all four processes, and $N_A - i$ Type A and $N_B - j - 2$ Type B servers that are idle in all four processes. We probabilistically link the remaining two Type B servers according to the scheme in Table A.4.

	Server I	**Server II**
Process 1, State $(i,\ j)$	Idle Type B	Idle Type B
Process 2, State $(i,\ j+1)$	Busy Type B	Idle Type B
Process 3, State $(i,\ j+1)$	Idle Type B	Busy Type B
Process 4, State $(i,\ j+2)$	Busy Type B	Busy Type B

Table A.4: Marking scheme for units in the sample path argument for Equation (A.13).

In the first time period, seven transitions are possible:

1. A Type H arrival
2. A Type L arrival
3. A completion by a Type A server.
4. A completion by Server I.
5. A completion by Server II.
6. A completion by any other Type B server.
7. A dummy transition due to uniformization.

For $k = 1, \ldots 7$, let B_k be the event in which transition k occurs. It suffices to show that $\mathbb{E}[\Theta \mid B_k] \leq \mathbb{E}[\Theta' \mid B_k]$ for each k, and we proceed case-by-case.

Case 1 (Event B_1): For the time being, assume that either $i < N_A$ or $j + 2 < N_B$; we handle later the special case in which both inequalities do not hold. If $i < N_A$, then the four processes transition to states $(i+1,\, j)$, $(i+1,\, j+1)$, $(i+1,\, j+1)$, and $(i+1,\, j+2)$, respectively, and the induction hypothesis applies. If $i = N_A$ but $j+2 < N_B$, transitions to states $(i,\, j+1)$, $(i,\, j+2)$, $(i,\, j+2)$, and $(i,\, j+3)$ occur instead, and we again leverage the induction hypothesis.

Case 2 (Event B_2): As with Case 1, assume that either $i < N_A$ or $j+2 < N_B$ for the time being. If $j+2 < N_B$, the analysis is straightforward. Suppose $j+2 = N_B$ but $i < N_A$. If Processes 1 and 4 both reject, the analysis is straightforward. If Processes 1 and 4 both admit, Processes 2 and 3 use a Type A server, and

$$\begin{aligned}\mathbb{E}[\Theta \mid B_2] &= v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i+1,\, j+1) \\ &\le v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i+1,\, j+1) - R_{HA} + R_{HB} \\ &\le v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+1,\, j+2) \\ &= \mathbb{E}[\Theta' \mid B_2],\end{aligned}$$

where the first inequality follows by Lemma 2.4.2, and the second by the induction hypothesis on inequality (A.14). Finally, if Process 1 accepts and Process 4 rejects, then $\mathbb{E}[\Theta \mid B_2] = R_L = \mathbb{E}[\Theta' \mid B_2]$.

Case 3 (Event B_3): Transitions to states $(i-1,\, j)$, $(i-1,\, j+1)$, $(i-1,\, j+1)$, and $(i-1,\, j+2)$ occur, and the induction hypothesis applies.

Case 4 (Event B_4): Transitions to states $(i,\, j)$, $(i,\, j)$, $(i,\, j+1)$, and $(i,\, j+1)$ occur, and it follows that $\mathbb{E}[\Theta \mid B_4] = \mathbb{E}[\Theta' \mid B_4] = 0$.

Case 5 (Event B_5): Transitions to states $(i,\, j)$, $(i,\, j+1)$, $(i,\, j)$, and $(i,\, j+1)$, occur, and we have $\mathbb{E}[\Theta \mid B_5] = \mathbb{E}[\Theta' \mid B_5]$.

Case 6 (Event B_6): Transitions to states $(i,\, j-1)$, $(i,\, j)$, $(i,\, j)$, and $(i,\, j+1)$ occur, and the induction hypothesis applies.

Case 7 (Event B_7): A dummy transition occurs, and our induction hypothesis applies.

It remains to show $\mathbb{E}[\Theta \mid B_1] \leq \mathbb{E}[\Theta' \mid B_1]$ and $\mathbb{E}[\Theta \mid B_2] \leq \mathbb{E}[\Theta' \mid B_2]$ when $i = N_A$ and $j+2 = N_B$. However, $\mathbb{E}[\Theta \mid B_1] = v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+2)$ and $\mathbb{E}[\Theta \mid B_1] = R_{HB}$, and so the former inequality may not hold in general. We instead prove a slightly weaker claim, that $\mathbb{E}[\Theta \mid B_1 \cup B_2] \leq \mathbb{E}[\Theta' \mid B_1 \cup B_2]$, which suffices to show that $E[\Theta] \leq E[\Theta']$. Conditional on $B_1 \cup B_2$ occurring, a Type H job arrives with probability $\lambda_H/(\lambda_H + \lambda_L)$, and a Type L job arrives with probability $\lambda_L/(\lambda_H + \lambda_L)$. Noting that $\mathbb{E}[\Theta \mid B_2] = v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+2)$ as well, we have

$$\begin{aligned}
\mathbb{E}[\Theta \mid B_1 \cup B_2] &= \frac{\lambda_H}{\lambda_H + \lambda_L}\mathbb{E}[\Theta \mid B_1] + \frac{\lambda_L}{\lambda_H + \lambda_L}\mathbb{E}[\Theta \mid B_2] \\
&= v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i,\, j+2) \\
&\leq \frac{\lambda_H}{\lambda_H + \lambda_L} R_{HB} + \frac{\lambda_L}{\lambda_H + \lambda_L} R_L \\
&\leq \frac{\lambda_H}{\lambda_H + \lambda_L}\mathbb{E}[\Theta' \mid B_1] + \frac{\lambda_L}{\lambda_H + \lambda_L}\mathbb{E}[\Theta' \mid B_2] \\
&= \mathbb{E}[\Theta' \mid B_1 \cup B_2],
\end{aligned}$$

as desired. Thus, Inequality (A.13) holds for horizons of length $n+1$.

A.5.3 Inductive Proof, Inequality (A.14)

Fix $i \in \{0,\, \ldots,\, N_A - 2\}$ and $j \in \{0,\, \ldots,\, N_B\}$. We show that

$$\begin{aligned}
v_{n+1,\,\alpha}(i,\, j) - v_{n+1,\,\alpha}(i+1,\, j) \leq v_{n+1,\,\alpha}(i+1,\, j) - v_{n+1,\,\alpha}(i+1,\, j+1) \\
+ R_{HA} - R_{HB}
\end{aligned} \qquad \text{(A.35)}$$

using a sample path argument. Start four processes on the same probability space, each with $n+1$ periods remaining in the horizon. Processes 1 and 4 begin in states $(i,\, j)$ and $(i+1,\, j+1)$, respectively, and follow the optimal policy π^*.

Processes 2 and 3 begin in state $(i+1,\ j)$, and use potentially suboptimal policies π_2 and π_3 that mimic those described in Section A.5.1.

Let Ψ be the difference in reward collected by Processes 1 and 2 until coupling occurs; define Ψ' analogously for Processes 3 and 4. It suffices to show $\mathbb{E}\Psi \leq \mathbb{E}\Psi' + R_{HA} - R_{HB}$. There are i Type A and j Type B servers that are busy in all four processes, and $N_A - i - 1$ Type A and $N_B - j - 1$ Type B servers that are idle in all four processes. We probabilistically link the remaining Type A server and Type B server according to the scheme in Table A.5.

	Server I	**Server II**
Process 1, State $(i,\ j)$	Idle Type A	Idle Type B
Process 2, State $(i+1,\ j)$	Busy Type A	Idle Type B
Process 3, State $(i+1,\ j)$	Idle Type B	Busy Type A
Process 4, State $(i+1,\ j+1)$	Busy Type B	Busy Type A

Table A.5: Marking scheme for units in the sample path argument for Equation (A.14).

In the first time period, seven transitions are possible:

1. A Type H arrival
2. A Type L arrival
3. A completion by Server I.
4. A completion by Server II.
5. A completion by an unmarked Type A server.
6. A completion by an unmarked Type B server.
7. A dummy transition due to uniformization.

For $k = 1, \ldots 7$, let C_k be the event in which transition k occurs. It suffices to show for each k that $\mathbb{E}[\Psi \mid C_k] \leq \mathbb{E}[\Psi' \mid C_k] + R_{HA} - R_{HB}$ for each k.

Case 1 (Event C_1): Assume, for the time being, that either $i+1 < N_A$ or that $j+1 < N_B$; we handle later the special case in which both inequalities are violated. If $i+1 < N_A$, the analysis is straightforward. If $i+1 = N_A$ but $j+1 < N_B$, then Process 1 admits the call with a Type A server, while all other processes use Type B servers instead, and

$$\begin{aligned}\mathbb{E}[\Psi \,|\, C_1] &= v_{n,\alpha}(i+1,\, j) + R_{HA} - v_{n,\alpha}(i+1,\, j+1) - R_{HB} \\ &\le v_{n,\alpha}(i+1,\, j+1) + R_{HA} - v_{n,\alpha}(i+1,\, j+2) - R_{HB} \\ &= \mathbb{E}[\Psi' \,|\, C_1] + R_{HA} - R_{HB},\end{aligned}$$

where the inequality follows by the induction hypothesis on inequality (A.13).

Case 2 (Event C_2): As in Case 1, assume temporarily that either $i+1 < N_A$ or $j+1 < N_B$. If $j+1 < N_B$, the analysis is straightforward. If $j+2 = N_B$ but $i+1 < N_A$, suppose that Processes 2 and 3 always admit the incoming Type L job (with Type B servers), regardless of the actions taken by the other processes. If Process 4 admits (with a Type A server), then

$$\begin{aligned}\mathbb{E}[\Psi \,|\, C_2] &= v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i+1,\, j+1) \\ &\le v_{n,\alpha}(i+1,\, j+1) - v_{n,\alpha}(i+2,\, j+1) = \mathbb{E}[\Psi' \,|\, C_2]\end{aligned}$$

by the induction hypothesis on (A.12). If Process 4 rejects instead, then

$$\begin{aligned}\mathbb{E}[\Psi \,|\, C_2] &= v_{n,\alpha}(i,\, j+1) - v_{n,\alpha}(i+1,\, j+1) \\ &\le v_{n,\alpha}(i+1,\, j+1) + R_L - v_{n,\alpha}(i+1,\, j+1) + R_{HA} - R_{HB} \\ &= \mathbb{E}[\Psi' \,|\, C_2] + R_{HA} - R_{HB},\end{aligned}$$

where the inequality follows by combining both statements of Lemma 2.4.2.

Case 3 (Event C_3): Transitions to states $(i,\, j)$, $(i,\, j)$, $(i+1,\, j)$, and $(i+1,\, j)$ occur. Coupling results, and $\mathbb{E}[\Theta \,|\, C_3] = \mathbb{E}[\Theta' \,|\, C_3] = 0$.

Case 4 (Event C_4): Transitions to states $(i,\ j)$, $(i+1,\ j)$, $(i,\ j)$, and $(i+1,\ j)$ occur. Coupling results, and we have $\mathbb{E}[\Delta \mid C_4] = \mathbb{E}[\Delta' \mid C_4]$.

Case 5 (Event C_5): Transitions to states $(i-1,\ j)$, $(i,\ j)$, $(i,\ j)$, and $(i,\ j+1)$ occur, and the induction hypothesis applies.

Case 6 (Event C_6): Transitions to states $(i,\ j-1)$, $(i+1,\ j-1)$, $(i+1,\ j-1)$, and $(i+1,\ j)$ occur, and the induction hypothesis applies.

Case 7 (Event C_7): Our induction hypothesis applies immediately.

It remains to show that $\mathbb{E}[\Psi \mid B_1] \leq \mathbb{E}[\Psi' \mid B_1]$ and $\mathbb{E}[\Psi \mid B_2] \leq \mathbb{E}[\Psi' \mid B_2]$ when $i = N_A$ and $j + 2 = N_B$. Once again, these inequalities may not hold in general, and so we again show that $\mathbb{E}[\Psi \mid B_1 \cup B_2] \leq \mathbb{E}[\Psi' \mid B_1 \cup B_2]$. Reasoning similar to that used in the inductive proof of inequality (A.13) yields

$$
\begin{aligned}
\mathbb{E}[\Psi \mid B_1 \cup B_2] &= \frac{\lambda_H}{\lambda_H + \lambda_L}\Big[v_{n,\alpha}(i+1,\ j) - v_{n,\alpha}(i+1,\ j+1) + R_{HA} - R_{HB}\Big] \\
&\quad + \frac{\lambda_L}{\lambda_H + \lambda_L}\Big[v_{n,\alpha}(i,\ j+1) - v_{n,\alpha}(i+1,\ j+1)\Big] \\
&\leq \frac{\lambda_H}{\lambda_H + \lambda_L}\Big[v_{n,\alpha}(i+1,\ j) - v_{n,\alpha}(i+1,\ j+1) + R_{HA} - R_{HB}\Big] \\
&\quad + \frac{\lambda_L}{\lambda_H + \lambda_L}\Big[v_{n,\alpha}(i+1,\ j) - v_{n,\alpha}(i+1,\ j+1) + R_{HA} - R_{HB}\Big] \\
&\leq \frac{\lambda_H}{\lambda_H + \lambda_L}\Big[R_{HB} + R_{HA} - R_{HB}\Big] + \frac{\lambda_L}{\lambda_H + \lambda_L}\Big[R_L + R_{HA} - R_{HB}\Big] \\
&= \frac{\lambda_H}{\lambda_H + \lambda_L}\Big[v_{n,\alpha}(i+1,\ j+1) + R_{HB} - v_{n,\alpha}(i+1,\ j+1)\Big] \\
&\quad + \frac{\lambda_L}{\lambda_H + \lambda_L}\Big[v_{n,\alpha}(i+1,\ j+1) + R_L - v_{n,\alpha}(i+1,\ j+1)\Big] \\
&\quad + R_{HA} - R_{HB} \\
&= \mathbb{E}[\Psi' \mid B_1 \cup B_2] + R_{HA} - R_{HB},
\end{aligned}
$$

where the first inequality follows by Lemma 2.4.2, and the second by Lemma A.5.1.

APPENDIX B

APPENDIX FOR CHAPTER 3

B.1 An Extended MDP Model

We consider an extension of our MDP from Section 3.3 that allows low-priority calls to queue, and ALS units to assist BLS responses to high-priority calls. As before, we assume that high-priority and low-priority calls arrive according to independent Poisson processes with rates λ_H and λ_L, respectively, and that service times are exponentially distributed with rate μ. If a low-priority call arrives when all BLS units are busy, and does not immediately receive an ALS response, it enters a queue with capacity C, to be served when the system becomes less congested. If a high-priority call arrives when only BLS units are available, we assume that with probability p, the call cannot adequately be treated on-scene. In this case, the BLS unit remains on scene in a "limbo" state, during which it cannot respond to other calls. When an ALS unit becomes available, it is immediately brought on scene, freeing the BLS unit, and allowing the high-priority patient to leave the system after an exponentially distributed service time (also with rate μ). We still assume that high-priority calls do not queue, as such a queue would only be utilized when all ambulances are busy, in which case, external resources would likely be brought in.

B.1.1 State Space

The state of the system can be fully characterized by four quantities: the number of busy ALS ambulances, the number of busy BLS ambulances, the number of low-priority calls in queue, and the number of BLS units in limbo. However, but by leveraging our assumption that service times are identically distributed, a two-dimensional state space suffices. In particular, define

$$\mathcal{S}_0 = \{0, 1, \ldots, N_A, N_A + 1, \ldots, N_A + N_B\} \times \{0, 1, \ldots, N_B, N_B + 1, \ldots, N_B + C\},$$

and suppose $(i, j) \in \mathcal{S}_0$. If $i \leq N_A$ and $j \leq N_B$, we can interpret (i, j) as before. If $i > N_A$, then BLS units are in limbo, and if $j > N_B$, then the queue is nonempty. Specifically, if the system is in state (i, j), where $i = N_A + i'$ and $j = N_B + j'$ for some $i', j' > 0$, then all ALS units are busy, i' BLS units are in limbo, $N_B - i'$ BLS units are busy serving low-priority calls, and j' low priority calls are in queue. This construction is valid because we assumed that BLS units will only be dispatched to high-priority calls when all ALS units are busy, and low-priority calls queue only when all BLS units are busy.

Not all states in $\mathcal{S}_0$ are reachable: for instance, the state $(N_A + 1, 0)$. More generally, if the system is in state (i, j), where $i > N_A$, then $j \geq i - N_A$ must hold. Thus, we redefine our state space to be $\mathcal{S} = \{(i, j) \in \mathcal{S}_0 : j \geq (i - N_A)^+\}$, where for a real number x, we define $x^+ = \max\{x, 0\}$.

B.1.2 Action Space

As before, we assume that whenever possible, the system provides ALS responses to high-priority calls, and BLS responses to low-priority calls. If a BLS unit completes service while the queue is nonempty, the unit immediately begins service on a queued call. If an ALS unit becomes free while a BLS unit is in limbo and a low-priority call is in queue, assisting the BLS unit takes priority.

Thus, there are only decisions to be made when ALS units are available, all BLS units are busy with low-priority calls, and the queue is nonempty— that is, in states where $i < N_A$ and $j > N_B$. In this case, the decision-maker must choose between dispatching an ALS unit to a queued low-priority call (Action 1), or reserving system resources for future high-priority calls (Action 0). We assign all other states a dummy action (Action 0). Thus, we define $\mathcal{A} = \cup_{(i,j)\in\mathcal{S}} A(i,j)$, where $A(i,\, j) = \{0,1\}$ if $i < N_A$ and $j > N_B$, and $\{0\}$ otherwise.

We make two remarks about our construction of $\mathcal{A}$. First, if $i < N_A$ and $j = N_B$, an arriving low-priority call enters the queue. However, Action 1 can be taken immediately, in which case, the call spends no time in queue. Second, we need not consider situations where multiple ALS units are simultaneously dispatched to queued calls, as a dispatch would have been performed when the previous low-priority call entered the queue.

B.1.3 Uniformization

Our extended MDP is also uniformizable; without loss of generality, assume $\Lambda = 1$. If the system begins a uniformized time period in state $(i,j) \in \mathcal{S}$, then the next event is

- With probability λ_H, the arrival of a high-priority call,
- With probability λ_L, the arrival of a low-priority call,
- With probability $\min\{i, N_A\}\mu$, an ALS service completion,
- With probability $[\min\{j, N_B\} - (i - N_A)^+]\,\mu$, a BLS service completion
- With probability $(N_A - i)^+ + [(N_B - j)^+ + (i - N_A)^+]\,\mu$, a dummy transition due to uniformization.

The fourth probability follows because if $i > N_A$, then $i - N_A$ BLS units are in limbo, implying that only $\min\{j, N_B\} - (i - N_A)$ BLS units can complete service.

B.1.4 Rewards

Define R_{HA}, R_{HB}, and R_L as before. We assume that the system collects a reward R_{HB} from each BLS response to a high-priority call, regardless of whether or not an ALS unit provides assistance. We also assume that the system collects reward when serving queued low-priority calls, but also incurs a holding cost h per unit time for each such call in queue. Assuming that actions take effect at the start of the period, and events at the end, we obtain the one-stage rewards

$$R((i,j),a) = \begin{cases} \lambda_H R_{HA} + \lambda_L R_L & \text{if } i < N_A,\ j < N_B,\ a = 0, \\ \lambda_H R_{HB} + \lambda_L R_L & \text{if } i \geq N_A,\ j < N_B,\ a = 0, \\ \mathbf{I}(i < N_A)\lambda_H R_{HA} - h(j - N_B) & \text{if } i \leq N_A,\ j \geq N_B,\ a = 0, \\ \quad +\mathbf{I}(j > N_B) N_B \mu R_L & \\ \mathbf{I}(i + 1 < N_A)\lambda_H R_{HA} + R_L & \text{if } i < N_A,\ j > N_B,\ a = 1, \\ \qquad + \mathbf{I}(j - N_B > 1) N_B \mu R_L & \\ \qquad -h(j - N_B - 1) & \\ N_A \mu R_L + (N_B - (i - N_A))\mu R_L & \text{if } i \geq N_A,\ j > N_B,\ a = 0. \\ \qquad -h(j - N_B) & \end{cases}$$

The first two terms are straightforward. The third term follows because the system collects reward from incoming high-priority calls only if an ALS unit is available, and incurs holding cost from queued low-priority calls. If a BLS service completion is the next event to occur, a dispatch is immediately made to a call in queue, and the system collects an additional reward R_L.

In the fourth term, the system collects a reward R_L from taking Action 1, and the queue shrinks by one. The system may collect an additional reward R_L if there is another call in queue, and a BLS service completion occurs at the end of the period. If Action 1 results in all ALS units becoming busy, the system does not collect reward if a high-priority call arrives. In the fifth term, all ambulances are busy, the queue is nonempty, and BLS units are in limbo. The system pays holding costs, but may collect a reward R_L if a BLS unit becomes free.

B.1.5 Transition Probabilities

Because the probabilities $P\big((i',\, j')\,|(i,\, j),\, a\big)$ are fairly cumbersome, we do not fully specify them here, and consider only the two most interesting cases.

Case 1: If $i < N_A$, $j > N_B$, and $a = 1$, an immediate transition to state $(i+1, j-1)$ occurs, and

$$P\big((i',\, j')\,|(i,\, j),\, 1\big) = \begin{cases} \mathbf{I}(i+1 < N_A)\lambda_H & (i', j') = (i+2, j-1) \\ \lambda_L & (i', j') = (i+1, j) \\ (i+1)\mu & (i', j') = (i, j) \\ N_B\mu & (i', j') = (i+1, j-2) \\ \mathbf{I}(i+1 = N_A)\lambda_H + (N_A - i - 1)\mu & (i', j') = (i+1,\, j-1) \end{cases}$$

We use an indicator because if $i + 1 = N_A$, all ALS units become busy after Action 1 is performed, and subsequent high-priority calls are redirected.

Case 2: If $i > N_A$ and $j < N_B$, only Action 0 is available, and

$$P\big((i',\, j')\,|(i,\, j),\, 0\big) = \begin{cases} p\lambda_H & \text{if } (i',\, j') = (i+1, j+1), \\ (1-p)\lambda_H + \lambda_L & \text{if } (i',\, j') = (i, j+1), \\ N_A\mu & \text{if } (i',\, j') = (i-1, j-1), \\ [j - (i - N_A)]\mu & \text{if } (i',\, j') = (i,\, j-1), \\ (i - N_A)\mu & \text{if } (i',\, j') = (i,\, j). \end{cases}$$

The $(i - N_A)$ BLS units in limbo trigger dummy transitions. All arriving calls receive a BLS response, and with probability $p\lambda_H$, the responding unit is brought into limbo. If an ALS unit becomes idle, a BLS unit is immediately freed from limbo and becomes idle.

To maximize long-run average reward, we can again restrict our attention to policies that are stationary, deterministic mappings $\pi : \mathcal{S} \to \{0, 1\}$. Since our extended MDP is also irreducible, it can be solved via by setting up optimality equations analogous to those in 2.4, and applying a policy iteration algorithm.

B.1.6 Computational Study

We base our study on the same hypothetical EMS considered in Section 3.6, and again set $R_{HA} = 1$, $R_{HB} = 0.5$, $R_L = 0.6$, $\mu = 0.75$, $C_A = 1.25$, and $C_B = 1$ as our base values. Again, to compensate for the effects of resource pooling, we introduce congestion in our system by shrinking the operating budget B to 43.75. The input parameters p, h, and C cannot be readily estimated from our dataset, so we select $p = 0.5$, $h = 0.6$, and $C = 10$. We choose h so that our system does not collect any reward from a low-priority call if it spends more than one hour in queue.

Evaluating long-run average reward in both MDP models for each vehicle mix in the set $\Gamma = \{(N_A, N_B) : N_A \leq 70 \text{ and } N_B = \lfloor 43.75 - 1.25N_A \rfloor\}$, we obtain Figure 3.4.2 in Chapter 3. As previously noted, both models yield similar quantitative results, but the extended MDP model more heavily penalize vehicle mixes operating too few ALS ambulances. Sensitivity analyses with respect to our input parameters yield similar results. Figure B.1 below illustrates the curves we obtain as we vary C_A, and is analogous to Figure 3.4 in Chapter 3, which we reproduce here for reference. As in Figure 3.4.2, the extended MDP

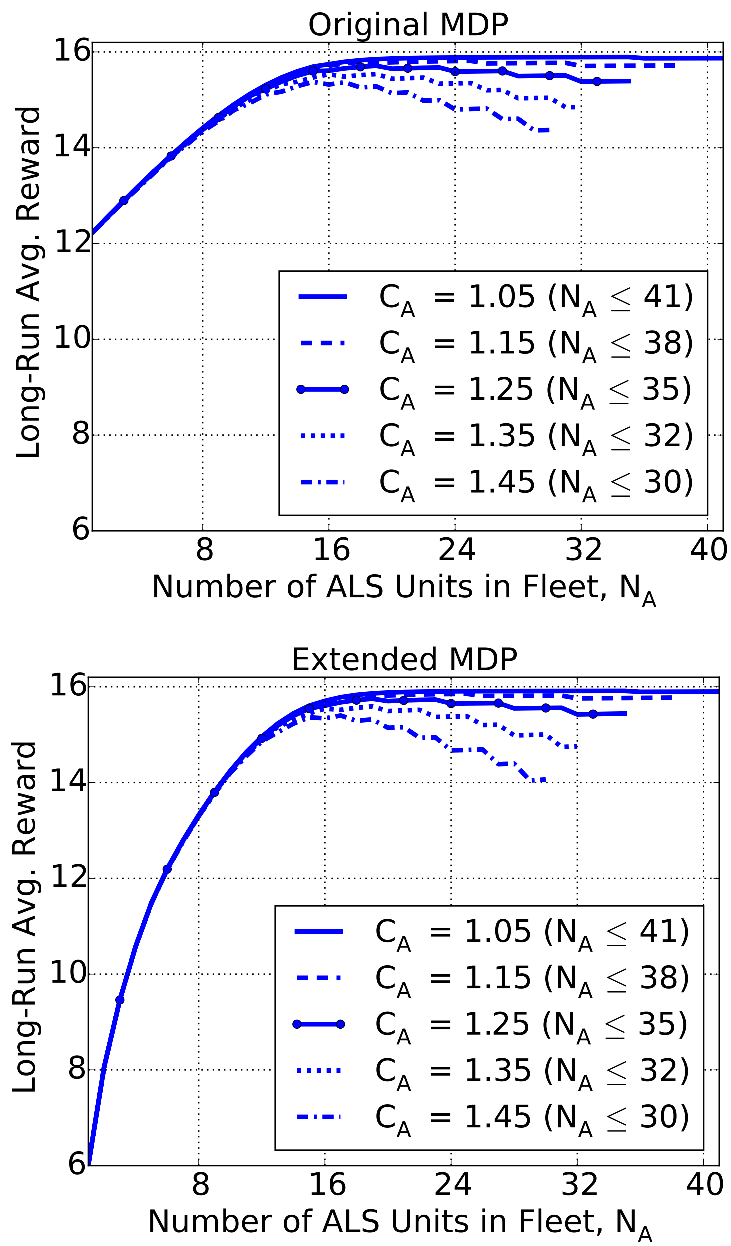

Figure B.1: Long-run average reward vs. vehicle mix for several choices of C_A, for both MDP models.

predicts a lower performance for fleets operating too few ALS ambulances. The model also penalizes fleets expending too much of the budget on ALS ambulances when C_A is large. If we restrict our attention to more carefully chosen fleets, we again observe the same general trends. We obtain similar results when performing other sensitivity analyses, which we omit for brevity. Taken together, our experiments suggest that our qualitative conclusions may not be particularly sensitive to our modeling assumptions.

B.2 An Extended Integer Program

In formulating our IP model in Section 3.5, we assumed that ambulances are busy independently of one another. However, knowing that a particular ambulance is busy may indicate that the system is congested, thus increasing the probability that other ambulances may be busy. Indeed, some studies have suggested that this assumption may lead to optimistic estimates of coverage; see, for instance Baron et al. [8] and Borrás and Pastor [16]. In this section, we consider an IP model in which we relax the independence assumption using correction factors, in a manner similar to that of Larson [54].

B.2.1 Formulation

Suppose demand node i is covered by a ALS and b BLS ambulances. The system obtains reward from this node at a rate $\lambda_i^H R_H(a,b) + \lambda_i^L R_L(a,b)$ per unit time. Under our original IP model:

$$R_H(a,b) = R_{HA}\,(1-(p_A)^a) + R_{HB}\,(p_A)^a(1-(p_B)^b) \tag{B.1}$$

$$R_L(a,b) = R_L[1-(p_B)^b + \phi\,(p_B)^b\,(1-(p_A)^a)], \tag{B.2}$$

Let $p_{HA}(a,b)$ be the long-run proportion of high-priority calls receiving an ALS response, provided the corresponding demand node is covered by a ALS and b BLS ambulances. We define $p_{HB}(a,b)$, $p_{LA}(a,b)$, and $p_{LB}(a,b)$ in a similar fashion. This allows us to rewrite (B.1) and (B.2) as follows:

$$R_H(a,b) = R_{HA}\, p_{HA}(a,b) + R_{HB}\, p_{HB}(a,b) \tag{B.3}$$

$$R_L(a,b) = R_L\left[\, p_{LA}(a,b) + p_{LB}(a,b)\,\right]. \tag{B.4}$$

Consider first the probability $p_{HA}(a,b)$. If we assume independence, then

$$p_{HA}(a,b) = 1 - (p_A)^a = \sum_{j=0}^{a-1} (p_A)^j (1 - p_A). \tag{B.5}$$

The j^{th} term in the above sum denotes the probability that the first j ambulances the dispatcher contacts are busy, but the next ambulance considered is available. Following Larson [54], we multiply each term in the right-hand side of (B.5) by a *correction factor*. To do so, consider an $M/M/a/a$ queueing system with arrival rate λ and service rate μ. Letting P_0 and P_a denote the long-run probability that all servers are idle and busy, respectively, it is straightforward to show

$$P_0 = \left(\sum_{i=0}^{a} \frac{(a\rho)^i}{i!} \right)^{-1} \qquad\qquad P_a = \frac{(a\rho)^a}{a!} P_0,$$

where $\rho = \lambda / a\mu$ is the offered load of the system. If servers are sampled without replacement while the system is in steady state, the probability that j busy servers are selected before an idle one is found is $Q(a,\, \rho,\, j)\, p^j (1-p)$, where $p = \rho(1 - P_a)/a$ denotes average server utilization, and

$$Q(a,\, \rho,\, j) = P_0 \sum_{k=j}^{a} \frac{(a-j-1)!\,(a-k)\,a^k\,\rho^{k-j}}{(k-j)!\,a!\,(1-\rho)}. \tag{B.6}$$

Here, $Q(a,\, \rho,\, j)$ can be viewed as a multiplicative constant that corrects the probability we would have obtained, had we incorrectly assumed that servers are busy independently of one another. These factors have been applied to IP models of ambulance deployment, to relax the independence assumption; see, for instance, McLay [63]. In our setting, we could replace (B.5) with

$$p_{HA}(a,b) = \sum_{j=0}^{a-1} Q(a,\, \rho_A,\, j)\, (p_A)^j (1 - p_A). \tag{B.7}$$

To compute the above probability, we need to determine ρ_A, the load offered to ALS ambulances in the system. This is nontrivial, as ALS ambulances respond to calls of both priorities, and the workload they receive depends on the dispatching policy. We approximate ρ_A using a procedure that we describe in Appendix B.2.2 below. Reasoning similar to the above yields the approximation

$$p_{LB}(a,b) = \sum_{j=0}^{b-1} Q(b,\, \rho_B,\, j)\,(p_B)^j(1-p_B), \tag{B.8}$$

where ρ_B denotes the offered load associated with BLS ambulances in the system; again, see Appendix B.2.2. We next consider the proportion of low-priority calls receiving ALS responses. With probability $1-p_{LB}(a,b)$, all BLS ambulances are busy when a low-priority call arrives. Conditional on this occurring, we use $p_{HA}(a,b)$ to approximate the probability that at least one ALS ambulances is available. This is a simplification, but it is milder than our original independence assumption. Since an ALS ambulance is dispatched to a proportion ϕ of these calls, we obtain the approximation

$$p_{LA}(a,b) = \phi\, p_{HA}(a,b)\,[1 - p_{LB}(a,b)]. \tag{B.9}$$

Reasoning in a similar fashion yields

$$p_{HB}(a,b) = \; p_{LB}(a,b)\,[1 - p_{HA}(a,b)]. \tag{B.10}$$

We formulate the objective function of our extended IP model by combining (B.3), (B.4), and (B.7) – (B.10). The constraints do not change, as we did not use the independence assumption to formulate them.

B.2.2 Approximating Offered Loads

To approximate the systemwide offered loads ρ_A and ρ_B, we again use the inputs of our IP model to build an instance of the dispatching MDP from Section 3.3. Let π^* denote the optimal policy, and ν the stationary distribution of the Markov chain induced by this policy. Let f_{HA} denote the long-run proportion of high-priority calls receiving an ALS response under π^*. Since this occurs when at least one ALS ambulance is available, the PASTA property implies

$$f_{HA} = \sum_{i=0}^{N_A-1} \sum_{j=0}^{N_B} \nu(i,\, j).$$

Defining f_{LB}, f_{LA}, and f_{HB} analogously, we obtain

$$\begin{aligned}
f_{LB} &= \sum_{i=0}^{N_A} \sum_{j=0}^{N_B-1} \nu(i,\, j) \\
f_{LA} &= \sum_{i=0}^{N_A} \sum_{j=0}^{N_B} \nu(i,\, j)\, I(\pi^*(i,\, j) = 1) \\
f_{HB} &= \sum_{j=0}^{N_B-1} \nu(N_A,\, j)
\end{aligned}$$

Approximating the fraction of high-priority and low-priority calls routed to ALS ambulances using $f_{HA}/(f_{HA}+f_{HB})$ and $f_{LA}/(f_{LA}+f_{LB})$, respectively, yields

$$\begin{aligned}
\rho_A &\approx \frac{1}{\mu}\left[\frac{f_{HA}}{f_{HA}+f_{HB}}\lambda_H + \frac{f_{LA}}{f_{LA}+f_{LB}}\lambda_L\right] \\
\rho_B &\approx \frac{1}{\mu}\left[\frac{f_{HB}}{f_{HA}+f_{HB}}\lambda_H + \frac{f_{LB}}{f_{LA}+f_{LB}}\lambda_L\right].
\end{aligned}$$

B.2.3 Computational Study

We base our numerical work upon the same hypothetical EMS from Section 3.6. As in previous "base case" experiments, we use $R_{HA} = 1$, $R_{HB} = 0.6$, $R_L = 0.5$,

$C_A = 1.25$, $C_B = 1$, and $\mu = 0.75$, and evaluate the set of vehicle mixes in the set $\Gamma = \{(N_A, N_B) : N_A \leq 70$ and $N_B = \lfloor 87.5 - 1.25N_A \rfloor\}$.

We approximate p_A, p_B, and ϕ using the procedure from Section 3.5.2. Recall that we approximated ϕ using the output of a modified MDP in which arrivals were scaled up to more accurately model how dispatching decisions would be made in a congested system. We use the same MDP to compute f_{HA}, f_{HB}, f_{LA}, and f_{LB}. Solving the resulting IP instances, and plotting the curves alongside those from Figure 3.7 in Section 3.6, we obtain Figure B.2. While the extended

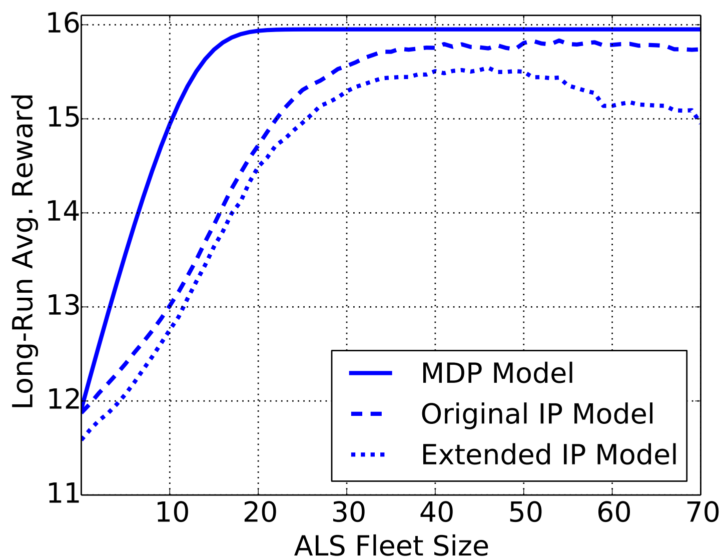

Figure B.2: Long-run average reward vs. vehicle mix, under the MDP model and both IP models.

IP model indeed yields less optimistic results than the original IP model, the curves in Figure B.2 have the same basic structure. When we increase the degree of resource pooling in the system (via the response time threshold T), we see numerical results that more closely match those produced by the MDP; see Figure B.3. In the case of complete resource pooling, the extended IP is still optimistic relative to the MDP, suggesting that the correction factors do not completely account for server dependence. Nonetheless, the fit is improved.

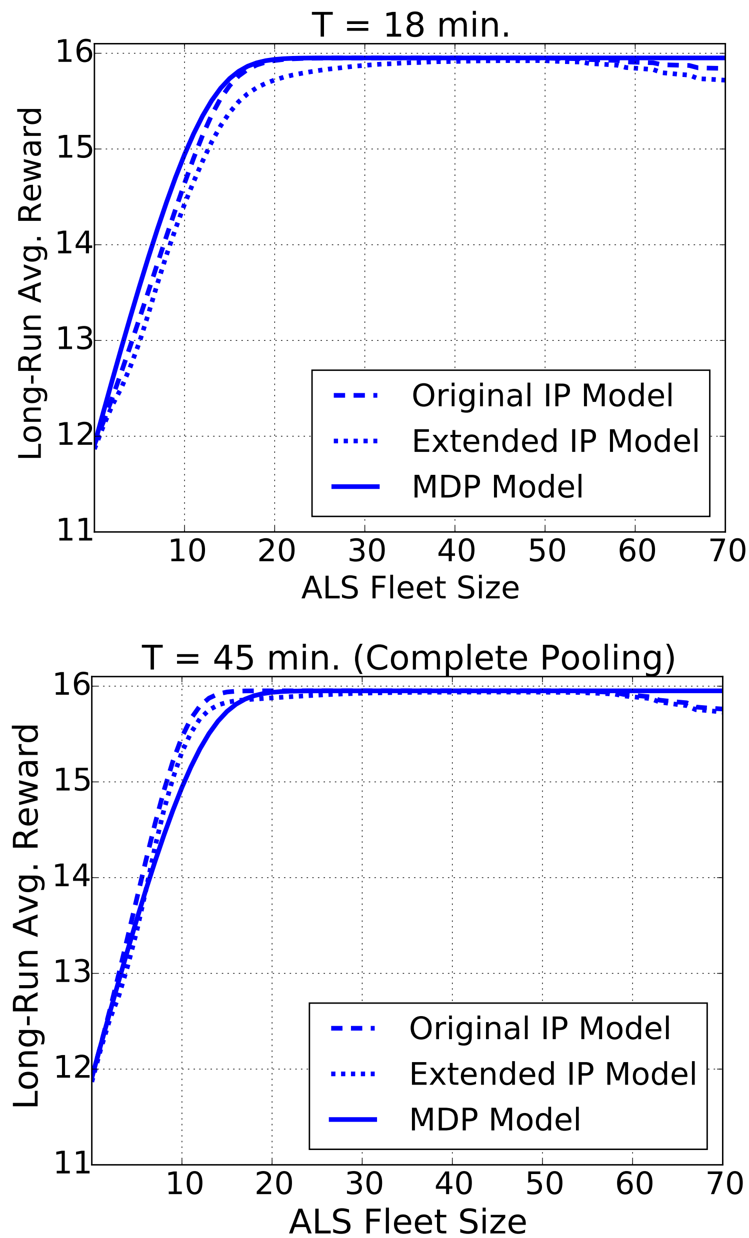

Figure B.3: Long-run average reward under the MDP and both IP models, for two response time thresholds T.

We conclude with a sensitivity analysis with respect to C_A identical to that used to generate Figure 3.9, and obtain Figure B.4. For larger values of C_A, the extended IP model predicts a more aggressive drop in performance as we move towards an all-ALS fleet, but we again see the same qualitative trends. This pattern repeats if we perform sensitivity analyses on ambulance travel speeds and rewards, but we omit plots in the interest of brevity. Taken together, our

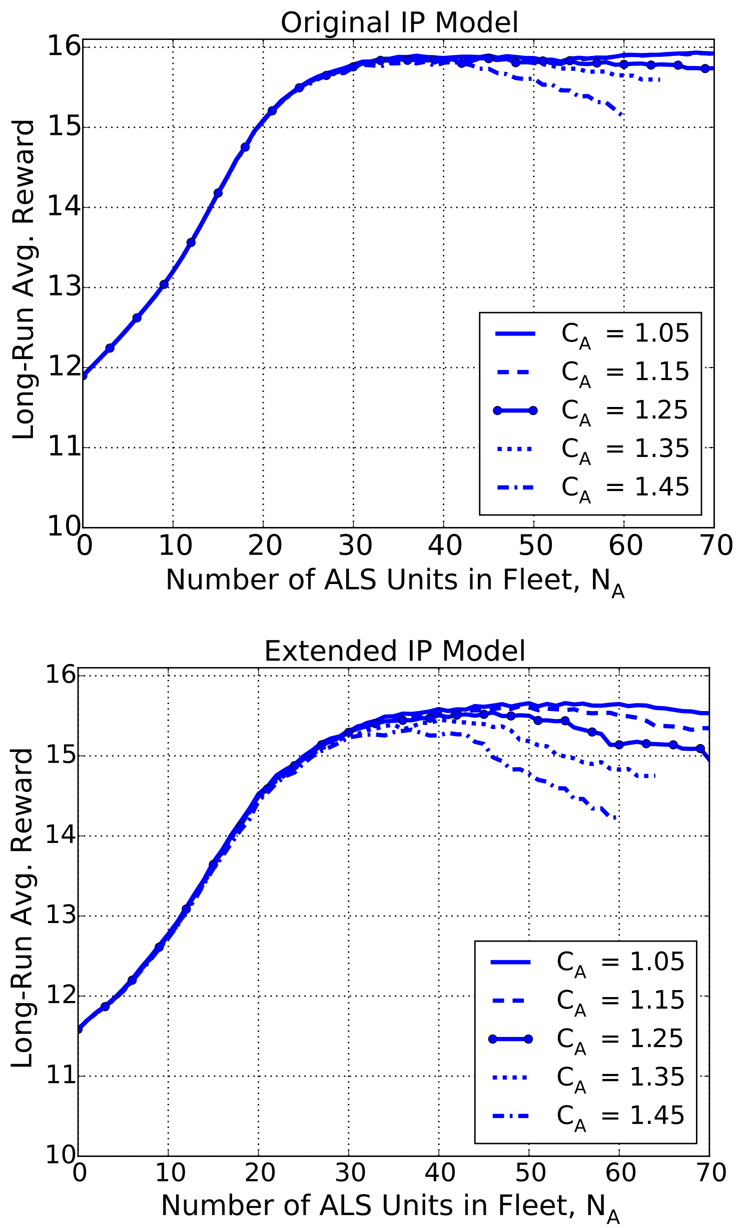

Figure B.4: Long-run average reward under the MDP and both IP models, for two response time thresholds T.

experiments suggest that both IP models yield the same qualitative insight: that of rapidly diminishing marginal returns associated with increasing N_A.

APPENDIX C

APPENDIX FOR CHAPTER 5

C.1 Proof of Lemma 5.5.1

Fix a sample path ω and an initial state s_0. We prove the result by showing that a feasible solution to (PIDP(ω)) can be converted into one for (IP1(ω)) that attains the same objective value, then by showing the converse.

C.1.1 From the DP to the IP

A feasible solution to (PIDP(ω)) can be characterized by the initial system state s_0, as well as the dispatching decisions $d = (d_0, \ldots, d_{T+1})$ and redeployment moves $m = (m_0, \ldots, m_{T+1})$ taken in each time period. Let $s = (s_0, \ldots, s_{T+1})$ denote the status of the fleet over the horizon $\mathcal{T}$, as implied by the decisions d and m. We construct a feasible solution to (IP1(ω)) as follows. Given a call c, base $j \in B_c$, and another base $k \in B$, let

$$x_{cj} = \begin{cases} 1 & \text{if } d_c = a \text{ and } \ell_a(s_c) = j \text{ for some ambulance } a \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} y_{tj} &= \sum_{a=1}^{A} I(\text{Ambulance } a \text{ is idle at base } j \text{ at time } t) \\ &= \sum_{a=1}^{A} I(\ell_a(s_t) = j,\ f_a(s_t) = t,\ c_a(s_t) = \phi) \end{aligned}$$

$$z_{tck} = \begin{cases} 1 & \text{if } d_c = a,\ f_a(s_c) = t, \text{ and } m_{ta} = k \text{ for some ambulance } a \\ 0 & \text{otherwise} \end{cases}$$

That is, $x_{cj} = 1$ if there is an ambulance a at base j that responded to call c, and $z_{tck} = 1$ if that ambulance becomes free at time t, and is subsequently redeployed to base k. We set y_{0j} to be the number of ambulances initially stationed at base j, as implied by s_0.

Let $X^{DP} = (d, m)$ denote our original feasible solution to (PIDP(ω)), and $X^{IP} = (x, y, z)$ the solution to (IP1(ω)) constructed above. To show that the solutions attain identical objective values, we observe that the reward collected by X^{DP} equals the number of calls c for which $d_c \neq \phi$, and the reward collected by X^{IP} is simply the number of x–variables taking on nonzero values. If $d_c \neq \phi$ for some call c, then by construction, there exists exactly one base j for which $x_{cj} = 1$. Conversely, if $x_{cj} = 1$ for some call c and base j, then we must have $d_c \neq \phi$. It follows that the two solutions attain identical objective values.

It remains to show that X^{IP} is feasible. Constraints (5.22) are satisfied immediately, by construction of X^{IP}. Constraint (5.17) also follows easily, as

$$\begin{aligned}\sum_{j \in B_c} x_{cj} &= \sum_{j \in B_c} \sum_{a=1}^{A} I(d_c = a \text{ and } \ell_a(s_c) = j) \\ &= \sum_{a=1}^{A} I(d_c = a) \sum_{j \in B_c} I(\ell_a(s_c) = j) \\ &\leq \sum_{a=1}^{A} I(d_c = a) \leq 1,\end{aligned}$$

where the first equality follows by construction, and the second inequality because a feasible solution to (PIDP(ω)) can dispatch only one ambulance to a call.

To show Constraints (5.18) hold, fix a call c and a base j, and suppose $x_{cj} = 1$. Ambulance a responded to the call from base j, and because X^{DP} is feasible, this occurs only if $f_a(s_c) = t$ and $m_a(s_c) = \phi$. By construction, we must have $y_{cj} \geq 1$. If $x_{cj} = 0$, then (5.18) immediately follows from nonnegativity of the y_{cj}. Next, we consider Constraint (5.19). Fix a call c and base $j \in B_c$. We have

$$\begin{aligned}\sum_{k \in B} z_{f(j,c),c,k} &= \sum_{k \in B} \sum_{a=1}^{A} I(d_c = a,\, \ell_a(s_c) = j,\, m_{f(j,c),a} = k) \\ &= \sum_{a=1}^{A} \sum_{k \in B} I(d_c = a,\, \ell_a(s_c) = j,\, m_{f(j,c),a} = k) \\ &= \sum_{a=1}^{A} I(d_c = a,\, \ell_a(s_c) = j) = x_{cj}.\end{aligned}$$

The first equality follows by construction of the z–variables. The third equality follows by feasibility of X^{DP}: if an ambulance is dispatched from base $j \in F_{tc}$ to call c, it must immediately be redeployed to a base at time $f(j, c)$. The final equality follows by construction of the x–variables.

It remains to show that Constraints (5.21) hold. Fix a time t and a base j, and assume that $t \in \mathcal{C}$, $j \in B_t$, and $x_{tj} = 1$; the analysis is simpler if any of these conditions are violated. Because $x_{tj} = 1$, an idle ambulance a responded to the call arriving at time t, and its status in the next time period is such that $\ell_a(s_{t+1}) = t \neq \phi$. Thus, ambulance a contributes to the sum used to compute y_{tj}, but not to the corresponding sum for $y_{t+1,j}$. Now suppose $z_{f(k,c),c,j} = 1$ for some $(k,\, c) \in Q_{t+1,j}$; again, the analysis is simpler if this is not the case. This corresponds to an ambulance a' that was dispatched to call c from base k, then redeployed to base j at time $f(k, c)$. By construction, ambulance a' becomes free at time $t+1$, and we would have $\ell_{a'}(s_{t+1}) = j$, $f_{a'}(s_{t+1}) = t+1$, and $c_{a'}(s_{t+1}) = \phi$. Thus, ambulance a' contributes to the sum used to compute $y_{t+1,j}$, but not to the corresponding sum for y_{tj}. Thus, to compute $y_{t+1,j}$, we start with y_{tj}, decrement if an ambulance was dispatched from that base at time t, and increment for each ambulance completing redeployment to the base at time $t + 1$.

It follows that X^{IP} is feasible for (IP1(ω)). $\square$

C.1.2 From the IP to the DP

One again, fix a sample path ω and an initial allocation $\{a_j : j \in B\}$, and let $X^{IP} = (x, y, z)$ be a feasible solution to (IP1(ω)). We construct a feasible solution to X^{DP} iteratively, and begin by setting $f_a(s_0) = 0$ and $c_a(s_0) = \phi$ for each ambulance a, reflecting the fact that each ambulance is idle at the start of the horizon. We next allocate ambulances to bases arbitrarily, provided that base j is assigned exactly a_j ambulances, and set $\ell_a(s_0)$ accordingly.

When advancing from time t to $t+1$, we use the values of the decision variables in X^{DP} to determine the values of d_{t+1}, m_{t+1}, and s_{t+1}. If an ambulance is dispatched (that is, $t \in \mathcal{C}$ and $x_{tj} = 1$ for some base $j \in B_c$), we find an ambulance a for which $\ell_a(s_t) = j$, $f_a(s_t) = t$, and $c_a(s_t) = \phi$ (breaking ties arbitrarily), set $d_t = a$, and update ambulance a's status as in (5.5)–(5.7). If an ambulance is redeployed— that is, $z_{tck} = 1$ for some base k— we find the ambulance a situated at call c's location for which $f_a(s_t) = t$ and $c_a(s_t) = c$, set $m_{ta} = k$, and update its status as in equations (5.8)–(5.10). For all remaining ambulances, we set $m_{ta} = \phi$ and update their statuses as in (5.2)–(5.4).

Let $X^{DP} = (d, m, s)$ be the resulting solution for (PIDP(ω)). Clearly, X^{IP} and X^{DP} have identical objective values, as for each call c, $d_c \neq \phi$ if and only if $x_{cj} = 1$ for some base $j \in B_c$. It remains to show feasibility, for which it suffices to prove two claims. First, we show if $x_{tj} = 1$, the status s_t of the fleet is such that there is an idle ambulance at base j. Second, we show that $\sum_{k \in B} z_{tck} = 1$ if and only if s_t is such that an ambulance has just completed service with call c. To verify the first claim, reasoning similar to that from Section C.1.1 gives

$$y_{tj} = \sum_{a=1}^{A} I(\ell_a(s_t) = j,\ f_a(s_t) = t,\ c_a(s_t) = \phi),$$

and so feasibility of dispatching decisions follows from feasibility of X^{IP} and

Constraints (5.18). To verify the second claim, we observe that if $\sum_{k \in B} z_{tck} = 1$, then $x_{cj} = 1$ for some base j, and Constraints (5.19) imply that $f(j, c) = t$. In constructing X^{DP}, we would have set $d_c = a$ for some ambulance a located at base j, implying that $f_a(s_t) = f(j, c)$ and $c_a(s_t) = c$. Thus, there would be an ambulance at call c's location that is ready to redeploy at time t. Conversely, if an ambulance a is ready to redeploy at time t, then there must exist a base j and call c for which $f(j,\, c) = t$ and $d_c = a$. By construction, this occurs only if $x_{cj} = 1$, and $\sum_{k \in B} z_{tck} = 1$ again follows by Constraints (5.19).

It follows that X^{DP} is feasible for (PIDP(ω)), and we are done. □

C.2 Proof of Lemma 5.5.2

Once again, we prove that (IP1(ω)) and (IP2(ω)) are equivalent by showing that feasible solutions to one integer program can be converted into feasible solutions of the other with the same objective values.

C.2.1 From (IP1(ω)) to (IP2(ω))

Let $X^1 = (x^1,\, y^1,\, z^1)$ be a feasible solution to (IP1(ω)). We construct a solution $X^2 = (x^2,\, y^2)$ as follows. First, we set $y^2 = y^1$: that is, we let $y^2_{tj} = y^1_{tj}$ for each time period t and base j. To construct x^2 from x^1 and z^1, consider a call c and base $j \in B_c$. There are two possibilities:

Case 1: If $f(j, c) \in \mathcal{T}$ (that is, if dispatching an ambulance from base j to call c would result in a service completion before the end of the horizon), then for each base $k \in B$, set $x^2_{cjk} = x^1_{cj}\, z_{f(j,c),c,k}$.

Case 2: If $f(j, c) \notin \mathcal{T}$, service completes after time $T + 1$, then the decision variable $z_{f(j,c),c,k}$ does not appear in (IP1(ω)) for any base k. However, (IP2(ω)) requires us to select a base to redeploy the responding ambulance each time a dispatch is made. Since this decision would have no impact on the objective function, we arbitrarily select base 1. Thus, $x^2_{cjk} = x^1_{cj}\, I(k = 1)$ for each $k \in B$. Both solutions attain identical objective values, as

$$\begin{aligned}
\sum_{c\in\mathcal{C}}\sum_{j\in B_c}\sum_{k\in B} x^2_{cjk} &= \sum_{\substack{c\in\mathcal{C}\\ f(j,c)\in\mathcal{T}}}\sum_{j\in B_c}\sum_{k\in B} x^2_{cjk} + \sum_{\substack{c\in\mathcal{C}\\ f(j,c)\notin\mathcal{T}}}\sum_{j\in B_c}\sum_{k\in B} x^2_{cjk} \\
&= \sum_{\substack{c\in\mathcal{C}\\ f(j,c)\in\mathcal{T}}}\sum_{j\in B_c}\sum_{k\in B} x^1_{cj}\, z^1_{f(j,c),c,k} + \sum_{\substack{c\in\mathcal{C}\\ f(j,c)\notin\mathcal{T}}}\sum_{j\in B_c}\sum_{k\in B} x^1_{cj}\, I(k=1) \\
&= \sum_{\substack{c\in\mathcal{C}\\ f(j,c)\in\mathcal{T}}}\sum_{j\in B_c} x^1_{cj}\sum_{k\in B} z^1_{f(j,c),c,k} \; + \sum_{\substack{c\in\mathcal{C}\\ f(j,c)\notin\mathcal{T}}}\sum_{j\in B_c} x^1_{cj} \\
&= \sum_{\substack{c\in\mathcal{C}\\ f(j,c)\in\mathcal{T}}}\sum_{j\in B_c} x^1_{cj} + \sum_{\substack{c\in\mathcal{C}\\ f(j,c)\notin\mathcal{T}}}\sum_{j\in B_c} x^1_{cj} = \sum_{c\in\mathcal{C}}\sum_{j\in B_c} x^1_{cj}.
\end{aligned}$$

The second equality follows by construction of x^2, and the fourth by feasibility of $(x^1,\, y^1,\, z^1)$, as well as constraint (5.19). It remains to verify that X^2 is feasible for (IP2(ω)). Constraint (5.26) holds because for each call c, we have

$$\begin{aligned}
\sum_{j\in B_c}\sum_{k\in B} x^2_{cjk} &= \sum_{\substack{j\in B_c\\ f(j,c)\in\mathcal{T}}}\sum_{k\in B} x^2_{cjk} + \sum_{\substack{j\in B_c\\ f(j,c)\notin\mathcal{T}}}\sum_{k\in B} x^2_{cjk} \\
&= \sum_{\substack{j\in B_c\\ f(j,c)\in\mathcal{T}}}\sum_{k\in B} x^1_{cj}\, z^1_{f(j,c),c,k} + \sum_{\substack{j\in B_c\\ f(j,c)\notin\mathcal{T}}}\sum_{k\in B} x^1_{cj}\, I(k=1) \\
&= \sum_{\substack{j\in B_c\\ f(j,c)\in\mathcal{T}}}\sum_{k\in B} x^1_{cj}\, z^1_{f(j,c),c,k} + \sum_{\substack{j\in B_c\\ f(j,c)\notin\mathcal{T}}} x^1_{cj} \\
&\le \sum_{\substack{j\in B_c\\ f(j,c)\in\mathcal{T}}} x^1_{cj} + \sum_{\substack{j\in B_c\\ f(j,c)\notin\mathcal{T}}} x^1_{cj} \le 1,
\end{aligned}$$

where the first inequality follows by Constraint (5.19), and the second by (5.17). To verify Constraints (5.27) hold, fix c and $j \in B_c$, and assume $f(j,c) \in \mathcal{T}$; the analysis for the case where $f(j,c) \notin T$ is similar. We have that

$$\sum_{k \in B} x^2_{cjk} = \sum_{k \in B} x^1_{cj}\, z^1_{f(j,c),c,k} \leq x^1_{cj} \leq y_{cj},$$

where the inequality follows by Constraint (5.19), and the second by feasibility of $(x^1,\, y^1,\, z^1)$. To show that Constraints (5.29) hold, fix t and j, and suppose that $t \in \mathcal{C}$ and $j \in B_t$; the analysis is similar if either condition is violated. Working with the right-hand side of the constraint, we find

$$\begin{aligned}
&y^2_{tj} - \sum_{k \in B} x^2_{tjk} + \sum_{(c,\,k) \in Q_{t+1,j}} x^2_{ckj} \\
&\quad = y^1_{tj} - \sum_{\substack{k \in B \\ f(j,t) \in \mathcal{T}}} x^1_{tj}\, z^1_{f(j,t),t,k} - \sum_{\substack{k \in B \\ f(j,t) \notin \mathcal{T}}} x^1_{tj}\, I(k=1) + \sum_{(c,\,k) \in Q_{t+1,j}} x^1_{ck}\, z^1_{f(k,c),c,j} \\
&\quad = y^1_{tj} - x^1_{tj}\, I(f(j,t) \in \mathcal{T}) - x^1_{tj}\, I(f(j,t) \notin \mathcal{T}) + \sum_{(c,\,k) \in Q_{t+1,j}} x^1_{ck}\, z^1_{f(k,c),c,j} \\
&\quad = y^1_{tj} - x^1_{tj} + \sum_{(c,\,k) \in Q_{t+1,j}} x^1_{cj} \\
&\quad = y^1_{t+1,j} = y^2_{t+1,j},
\end{aligned}$$

where the second equality follows by (5.19), and the third because $z^1_{f(k,c),c,j} = 1$ implies $x_{ck} = 1$. Finally, Constraints (5.30) follow immediately by construction. Thus, X^2 is feasible for (IP2(ω)). □

C.2.2 From (IP2(ω)) to (IP1(ω))

Conversely, let $X^2 = (x^2,\, y^2)$ be a feasible solution to (IP1(ω)). We construct a solution $X^1 = (x^1,\, y^1,\, z^1)$ as follows. As before, we set $y^1 = y^2$. To build x^1 from x^2 and z^2, fix c, $j \in B_c$, and $k \in B$. There are again two possibilities:

Case 1: If $f(j,c) \in \mathcal{T}$, then we let $x^1_{cj} = \sum_{k \in B} x^2_{cjk}$ and $z^1_{f(j,c),c,k} = x^2_{cjk}$.

Case 2: If $f(j,c) \notin \mathcal{T}$, then the decision variable $z_{f(j,c),c,k}$ does not exist in (IP1(ω)) for any k, and we simply let $x^1_{cj} = \sum_{k \in B} x_{cjk}$.

That the two solutions have identical objective values follows immediately, as

$$\sum_{c \in \mathcal{C}} \sum_{j \in B_c} x^1_{cj} = \sum_{c \in \mathcal{C}} \sum_{j \in B_c} \sum_{k \in B} x^2_{cjk}.$$

When examining feasibility of X^1, we note that Constraints (5.17), (5.18), and (5.22) can all be shown to hold in a similarly straightforward fashion. Turning to Constraints (5.19), and picking c and $j \in B_c$ for which $f(j,c) \in \mathcal{T}$, we find

$$\sum_{k \in B} z^1_{f(j,c),c,k} = \sum_{k \in B} x^2_{cjk} = x^1_{cj}.$$

Finally, we examine Constraints (5.21). Fix t and j, and assume $t \in \mathcal{C}$ and $j \in B_t$; the analysis for the case where this does not hold is similar. Starting with the right-hand side of the corresponding constraint, we find

$$y^1_{tj} - x^1_{tj} + \sum_{(c,k) \in M_{t+1,j}} z^1_{f(k,c),c,j} = y^2_{tj} - \sum_{k \in B} x^2_{tjk} + \sum_{(c,k) \in Q_{t+1,j}} x^2_{cjk} = y^2_{t+1,j} = y^1_{t+1,j},$$

as desired. Thus, X^1 is feasible for (IP1(ω)).

It follows that (IP1(ω)) and (IP2(ω)) are equivalent. $\square$

C.3 Discussion of Conjecture 5.5.3

To prove Conjecture 5.5.3 holds, it suffices to show that the coefficient matrix associated with the integer program (IP2(ω)) is totally unimodular, as this would imply that every basic feasible solution to (IP2(ω)) is integer-valued, from which the existence of an optimal integer solution trivially follows. We suspect that the following result would be pivotal in such a proof:

Lemma C.3.1. *Consider a relaxation (IP2R(ω))of the integer program* (IP2(ω)) *in which Constraints 5.26 are relaxed (which dictate that at most one ambulance can respond to any incoming call). Then the coefficient matrix associated with (IP2R(ω))is totally unimodular.*

The proof, which we state below, hinges upon showing that (IP2R(ω))can be reformulated as an equivalent min-cost (or in our case, max-reward) network flow problem (NF(ω)), as it is well-known that coefficient matrices with "network structure" have this property. When we reintroduce Constraints 5.26, we destroy the problem's network structure.

However, we conjecture that the proof can be completed by demonstrating that the extreme points of the feasible polytope associated with (IP2(ω)) are integer-valued. This is certainly the case for the relaxed model (IP2R(ω)), and we can view Constraints (5.26) as hyperplanes that cut into this polytope. We may be able to reason as follows: pick any call c, and consider the intersection of the feasible region of (IP2(ω)) with the hyperplane

$$\sum_{j \in B_c} \sum_{k \in B} x_{cjk} \leq 1. \tag{C.1}$$

The resulting intersection is a polytope, which we claim is the convex hull of the extreme points of (IP2R(ω)) satisfying (C.1). From here, it suffices to show that a similar equivalence holds as we introduce the remaining hyperplanes.

To prove this claim, we could show the hyperplane (C.1) does not create any fractional extreme points. We believe that any such fractional solution can be expressed as a convex combination of feasible solutions that satisfy the constraints of (IP2(ω)) as well as (C.1), resulting in a contradiction. One observation that may be critical in such a proof is that there is a one-to-one correspondence between extreme points of (IP2R(ω)), and feasible integer-valued solutions to the network flow problem (NF(ω)).

Remark C.3.2. *Our conjecture holds in the special case where $A = 1$: that is, when the fleet contains a single ambulance. Here, any fractional feasible solution x to (NF(ω)) can be decomposed into a convex combination of integer-valued flows via a fairly straightforward procedure. We begin by finding a path u_1 from source to sink consisting of edges with positive flow. Next, we "extract" flow from our feasible solution by identifying the arc on u_1 with the smallest flow value (call the corresponding value λ_1), then decrementing the flow sent along each arc in u_1 by λ_1. We repeat this process until no flow remains in the feasible solution, and obtain paths $u_1, \ldots, u_P$ for some finite P, that correspond to values $\lambda_1, \ldots, \lambda_P$. For $p = 1, \ldots, P$, let x_p denote the feasible solution to (NF(ω)) in which one unit of flow is sent along path u_p. It follows that*

$$x = \sum_{p=1}^{P} \lambda_p \, x_p.$$

If $A > 1$, we suspect that a generalization of the above procedure can be used to decompose fractional solutions to (NF(ω)) in a similar fashion. One difficulty that arises is that flows (from source to sink) must now be extracted in groups of A, and likely must be done in a structured way.

C.3.1 Proof of Lemma C.3.1

As mentioned above, it suffices to show that (IP2R(ω)) can be reformulated as an equivalent min-cost network flow problem, which we call (NF(ω)). We begin by constructing our instance of (NF(ω)), then demonstrating that any feasible solution to one problem can be converted into a feasible solution to the other with the same objective value.

To construct our graph, we begin by associating a node $(t,\, j)$ with each base $j \in B$ and time period $t \in \mathcal{T}$, and assigning each such node the supply value $b_{tj} = 0$. Next, we introduce a source node σ and a sink node τ, and set $b_\sigma = A$ and $b_\tau = -A$, where A is the number of ambulances in the fleet. For each node j, we connect an arc from σ to the node $(j,\, 0)$, and set its capacity to be a_j, the number of ambulances stationed at node j at the start of the horizon; this "initializes" our system. Similarly, for each j, we connect an uncapacitated arc from node $(j,\, T+1)$ to the sink τ; these edges help to "terminate" the problem at the end of the horizon. Next, for each base j and time $t < T+1$, we connect an uncapacitated arc from node $(t,\, j)$ to node $(t+1,\, j)$; sending one unit of flow along this arc corresponds to idling ambulance at base j from time t to $t+1$.

Finally, consider any time c at which a call arrives, and let i_c denote its location. For each $j \in B_c$ and $k \in B$, we connect an arc from node $(c,\, j)$ to node $(k,\, t(j,c,k))$, where $t(j,c,k) = \min\{f(j,c) + t(i_c,k),\, T+1\}$, and associate with this arc a reward of 1. Sending one unit of flow along this arc corresponds to dispatching an ambulance from base j to call c, and subsequently redeploying the ambulance to base k; note that $t(j,c,k)$ is the time at which such an ambulance would become idle at base k. Flow sent along this arc cannot "generate" more reward prior to time $t(j,c,k)$, reflecting the fact that a responding ambulance would be busy until that time. We define $t(j,c,k)$ using a minimum, as we can assume, without loss of generality, that ambulances that would not become idle until after the end of the horizon instead complete service at time $T+1$. To maintain the equivalence with (IP2R(ω)), we also assume, again without loss of generality that the sets $Q_{T+1,j}$ are modified such that calls that would they include calls that would become free after time $T+1$.

Let w_e denote the flow sent along arc e. A feasible flow must satisfy

$$\sum_{j \in B} w_{\sigma,\,(0,\,j)} = \sum_{j \in B} w_{(T+1,\,j),\,\tau} = A \tag{C.2}$$

$$\sum_{v \in I(t,j)} w_{v,\,(t,\,j)} = \sum_{v \in O(t,j)} w_{(t,\,j),\,v} \qquad \forall t \in \mathcal{T},\; j \in B \tag{C.3}$$

$$w_{(c,\,j),\,(t(j,c,k),\,k)} \leq 1 \qquad \forall c \in \mathcal{C}, j \in B_c,\; k \in B, \tag{C.4}$$

where $I(t,j)$ and $O(t,j)$ denote the set of nodes with arcs pointing into and out of node $(t,\,j)$, respectively. We assign feasible flows the objective function value

$$\sum_{c \in \mathcal{C}} \sum_{j \in B_c} \sum_{k \in B} w_{(c,\,j),\,(t(j,c,k),\,k)}.$$

It remains to prove that (IP2R(ω))and (NF(ω))are equivalent.

First, let $(x,\,y)$ be a feasible solution to (IP2R(ω)). To construct a feasible flow, we begin by setting $w_{u,\,(0,\,j)} = a_j$ for each base j. Next, for each triplet $(c,\,j,\,k)$ where $x_{cjk} = 1$, we set $w_{(c,\,j),\,(t(j,c,k),\,k)} = 1$. Finally, for each time period $t < T+1$ and base j, we send $y_{tj} - \sum_{k \in B} x_{cjk}\, I(t \in \mathcal{C},\; j \in B_c)$ units of flow from node $(t,\,j)$ to node $(t+1,\,j)$. The resulting flow w clearly has the same objective value as $(x,\,y)$, and so it remains to check feasibility.

Constraints (C.2) and (C.4) follow immediately by construction. Now consider the flow balance constraints (C.3). Suppose, for convenience, that $t \in \mathcal{C}$ and $j \in B_c$ (the analysis for the other cases is similar), and consider the right-hand side of (C.3). By construction, the set $I(t,j)$ consists of node $(t-1,\,j)$, as well as any call-base pairs $(c,\,k)$ for which $t(c,k,j) = t$— that is, any pairs in the set Q_{tj}. It follows that

$$\begin{aligned}\sum_{v \in I(t,j)} w_{v,\,(t,\,j)} &= w_{(t-1,\,j),\,(t,\,j)} + \sum_{(c,k) \in Q_{tj}} w_{(c,\,k),\,(t,\,j)} \\ &= y_{t-1,\,j} + \sum_{(c,k) \in Q_{tj}} x_{ckj}.\end{aligned}$$

Now consider the left-hand side of (C.3). The set $O(t,\, j)$ consists of $(t+1,\, j)$, and, for each base k, the node $(t(c, j, k),\, k)$. It follows that

$$\begin{aligned}\sum_{v \in O(t,j)} w_{(t,\, j),\, v} &= w_{(t,\, j),\, (t+1,\, j)} - \sum_{k \in B} w_{(t,\, j),\, (t(c,j,k),\, k)} \\ &= y_{tj} - \sum_{k \in B} x_{tjk}.\end{aligned}$$

Thus, Constraints (C.3) hold, as feasible solutions to (IP2R(ω)) satisfy (5.29), and so the flow w is indeed feasible.

Conversely, let w be a feasible flow. For each triplet (c, j, k) for which $c \in \mathcal{C}$, $j \in B_c$, and $k \in B$, we set x_{cjk} to be the flow transmitted along the arc connecting nodes $(c,\, j)$ and $(t(j, c, k),\, k)$, provided the reward associated with this arc is 1 (set $x_{cjk} = 0$ otherwise). Now consider the y–variables. Next, we set y_{tj} to be cumulative flow transmitted into node $(t,\, j)$ if $t > 1$, and $y_{tj} = a_j$ if $t = 0$. The resulting solution $(x,\, y)$ clearly has the same objective as the flow w, and so it remains to check for feasibility.

Constraints (5.30) follow immediately. Constraints (5.27) hold because feasible flows satisfy Constraints (C.3). That is, for each call c and base j, we have

$$\sum_{j \in B_c} \sum_{k \in B} x_{cjk} = \sum_{j \in B_c} \sum_{k \in B} w_{(c,\, j),\, (t(j,c,k),\, k)} \leq \sum_{v \in O(t,j)} w_{(c,\, j),\, v} = \sum_{v \in I(c,j)} w_{v,\, (c,\, j)} = y_{cj}$$

Finally, to verify that Constraints (5.29) hold, we observe for any time t and base j, $y_{t+1,j}$ is defined as the cumulative flow into node $(t+1,\, j)$, and by construction of w, this consists of flow originating from node $(t,\, j)$, as well as prior "dispatches" that send flow into node $(t+1,\, j)$, which is precisely the right-hand side of (5.29). Thus, $(x,\, y)$ is feasible for (IP2R(ω)), from which it follows that (IP2R(ω)) and (NF(ω)) are equivalent. $\square$

C.4 Proof of Lemma 5.6.2

As in Brown et al. [20], we show for each sample path ω that

$$V_t^{PI(z^*,\omega)}(s_t,\, i_t) = V_t^*(s_t,\, i_t) \tag{C.5}$$

and proceed using backward induction. The base case, when $t = T+1$, is trivial. Now suppose that (C.5) holds from time $t+1$ up until $T+1$. By (5.34), we have

$$\begin{aligned}
V_t^{PI(z^*,\omega)}(s_t,\, i_t) &= \max_{d_t,\, m_t} \Big\{ r(d_t) - z_t^*(s_t,\, (d_t,\, m_t)) \\
&\qquad\qquad + V_{t+1}^{PI(z^*,\omega)}\big(S_{t+1}(s_t,\, d_t,\, m_t,\, \omega),\, I_{t+1}(\omega)\big) \Big\} \\
&= \max_{d_t,\, m_t} \Big\{ r(d_t) - V_{t+1}^*\left(S_{t+1}(s_t,\, d_t,\, m_t,\, \omega),\, I_{t+1}(\omega)\right) \\
&\qquad\qquad + E\left[V_{t+1}^*\left(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\right) \middle| \mathcal{F}_t \right] \\
&\qquad\qquad + V_{t+1}^{PI(z^*,\omega)}\big(S_{t+1}(s_t,\, d_t,\, m_t,\, \omega),\, I_{t+1}(\omega)\big) \Big\} \\
&= \max_{d_t,\, m_t} \Big\{ r(d_t) + E\left[V_{t+1}^*\left(S_{t+1}(s_t,\, d_t,\, m_t),\, I_{t+1}\right) \middle| \mathcal{F}_t \right] \Big\} \\
&= V_t^*(s_t,\, i_t)
\end{aligned}$$

where the second line follows by construction of z^*, the third by the induction hypothesis, and the fourth because V^* solves the optimality equations (DP). $\square$

C.5 Proof of Lemma 5.6.3

We adapt the proof from Brown et al. [20]. Let $\pi \in \Pi$ be a nonanticipative policy, and fix a time period t and state $(s_t,\, i_t)$, and consider the first term on the right-hand side of (5.38). Given $H \in \mathcal{F}_{T+1}$, the perfect information filtration, we have

$$\begin{aligned}
&\int_H \tilde{V}_{t+1}\big(S_{t+1}(s_t,\, \pi_t,\, \omega),\, I_{t+1}(\omega)\big)\, d\mathbb{P}(\omega) \\
&\qquad = \int_H E\Big[\tilde{V}_{t+1}\big(S_{t+1}(s_t,\, \pi_t,\, \omega),\, I_{t+1}(\omega)\big) \Big| \mathcal{F}_{T+1} \Big]\, d\mathbb{P}(\omega),
\end{aligned}$$

where the equality holds because $\tilde{V}_{t+1}\left(S_{t+1}(s_t,\, \pi_t,\, \omega),\, I_{t+1}(\omega)\right)$ is clearly $\mathcal{F}_{T+1}$–measurable. The definition of conditional expectation implies

$$\tilde{V}_{t+1}\big(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\big) = E\Big[\tilde{V}_{t+1}\big(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\big)\Big|\mathcal{F}_{T+1}\Big].$$

Since the second term on the right-hand side of (5.38) does not depend on ω,

$$z_t\left(s_t,\, i_t,\, \pi_t\right) = E\Big[\tilde{V}_{t+1}\big(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\big)\Big|\mathcal{F}_{T+1}\Big] - E\left[\tilde{V}_{t+1}\left(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\right)\big|\mathcal{F}_t\right].$$

To prove the claim, it suffices to show $E\left[z_t(s_t,\, i_t,\, \pi_t)\,|\,\mathcal{F}_t\right] = 0$, as the law of iterated expectations would imply $E\left[z_t(s_t,\, i_t,\, \pi_t)\right] = 0$ for all t, from which inequality (5.33) follows easily. Indeed:

$$\begin{aligned}
E\big[z_t\left(s_t,\, i_t,\, \pi_t\right)|\mathcal{F}_t\big] &= E\left[E\left[\tilde{V}_{t+1}\big(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\big)\Big|\mathcal{F}_{T+1}\right]\Big|\mathcal{F}_t\right] \\
&\qquad - E\left[E\left[\tilde{V}_{t+1}\left(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\right)\big|\mathcal{F}_t\right]\Big|\mathcal{F}_t\right] \\
&= E\left[\tilde{V}_{t+1}\big(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\big)\Big|\mathcal{F}_t\right] \\
&\qquad - E\left[\tilde{V}_{t+1}\left(S_{t+1}(s_t,\, \pi_t),\, I_{t+1}\right)\big|\mathcal{F}_t\right] = 0,
\end{aligned}$$

where the first equality follows because π_t is $\mathcal{F}_t$–measurable (because π is nonanticipative), implying that $S_{t+1}(s_t,\, \pi_t)$ is $\mathcal{F}_t$–measurable as well. It follows that $\tilde{V}_{t+1}\big(S_{t+1}(s_t,\, \pi_t)\big)$ is also $\mathcal{F}_t$–measurable (as well as $\mathcal{F}_{T+1}$–measurable), and can be pulled outside the inner conditional expectation in both terms. $\square$

C.6 Proof of Lemma 5.6.5

Consider any sample path ω. It suffices to show that $V^{PIP}(\gamma,\, \omega)$ is convex in γ, as convexity is preserved through integration. We do this by finding a subgradient for the function $V^{PIP}(\gamma,\, \omega)$. Fix γ, and let $\big(x^*(\gamma),\, y^*(\gamma)\big)$ be the optimal solution to (PIP($\gamma,\, \omega$)). Similarly, fix another γ', and let $\big(x^*(\gamma'),\, y^*(\gamma')\big)$ be the corresponding optimal solution to (PIP(γ', ω)). We have that

$$
\begin{aligned}
&V^{PIP}(\gamma,\, \omega) - V^{PIP}(\gamma',\, \omega) \\
&\quad = \sum_{c,\, j,\, k} x^*_{cjk}(\gamma) - \sum_{t=0}^{T} \sum_{j \in B} \Big[\mathbb{1}\big(j \in I_{t+1}(\omega)\big)\, \gamma_{t+1,\, j} - \gamma_{t+1,\, j}\, \rho_{t+1,j} \Big] y^*_{t+1,\, j}(\gamma) \\
&\qquad - \left[\sum_{c,\, j,\, k} x^*_{cjk}(\gamma') - \sum_{t=0}^{T} \sum_{j \in B} \Big[\mathbb{1}\big(j \in I_{t+1}(\omega)\big)\, \gamma'_{t+1,\, j} - \gamma'_{t+1,\, j}\, \rho_{t+1,j} \Big] y^*_{t+1,\, j}(\gamma') \right] \\
&\quad \le \sum_{c,\, j,\, k} x^*_{cjk}(\gamma) - \sum_{t=0}^{T} \sum_{j \in B} \Big[\mathbb{1}\big(j \in I_{t+1}(\omega)\big)\, \gamma_{t+1,\, j} - \gamma_{t+1,\, j}\, \rho_{t+1,j} \Big] y^*_{t+1,\, j}(\gamma) \\
&\qquad - \left[\sum_{c,\, j,\, k} x^*_{cjk}(\gamma) - \sum_{t=0}^{T} \sum_{j \in B} \Big[\mathbb{1}\big(j \in I_{t+1}(\omega)\big)\, \gamma'_{t+1,\, j} - \gamma'_{t+1,\, j}\, \rho_{t+1,j} \Big] y^*_{t+1,\, j}(\gamma) \right] \\
&\quad = - \sum_{t=0}^{T} \sum_{j \in B} \Big[\mathbb{1}\big(j \in I_{t+1}(\omega)\big) - \rho_{t+1,j} \Big] y^*_{t+1,\, j}(\gamma)\, \gamma_{t+1,\, j} \\
&\qquad - \left[- \sum_{t=0}^{T} \sum_{j \in B} \Big[\mathbb{1}\big(j \in I_{t+1}(\omega)\big) - \rho_{t+1,\, j} \Big] y^*_{t+1,\, j}(\gamma)\, \gamma'_{t+1,\, j} \right] \\
&\quad = \sum_{t=0}^{T} \sum_{j \in B} \Big[\mathbb{1}\big(j \in I_{t+1}(\omega)\big) - \rho_{t+1,j} \Big] y^*_{t+1,\, j}\, (\gamma'_{t+1,\, j} - \gamma_{t+1,\, j}) \\
&\quad = \sum_{t=0}^{T} \sum_{j \in B} \nabla_{t+1,\, j}\, (\gamma'_{t+1,\, j} - \gamma_{t+1,\, j}),
\end{aligned}
$$

where the inequality follows because $\big(x^*(\gamma),\, y^*(\gamma)\big)$ is a feasible for (PIP(γ', ω)). Thus, ∇ is a subgradient of the function $V^{PIP}(\gamma,\, \omega)$, and we are done. $\square$

BIBLIOGRAPHY

[1] Daniel Adelman and Adam J. Mersereau. Relaxations of weakly coupled stochastic dynamic programs. *Operations Research*, 56(3):712–727, 2008.

[2] O. Zeynep Akşin and Fikri Karaesmen. Characterizing the performance of process flexibility structures. *Operations Research Letters*, 35(4):477–484, 2007.

[3] Ramon Alanis, Armann Ingolfsson, and Bora Kolfal. A Markov chain model for an ems system with repositioning. *Prod Oper Manag*, 22(1):216–231, 2012.

[4] E. Altman, T. Jimenez, and G. Koole. On optimal call admission control in resource-sharing system. *IEEE Transactions on Communications*, 49(9): 1659–1668, 2001.

[5] Leif Andersen and Mark Broadie. Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50(9): 1222–1234, 2004.

[6] T. Andersson and P. Vaerband. Decision support tools for ambulance dispatch and relocation. *Journal of the Operational Research Society*, 58:195–201, 2015.

[7] G. Bakalos, M. Mamali, C. Komninos, E. Koukou, A. Tsantilas, S. Tzima, and T. Rosenberg. Advanced life support versus basic life support in the pre-hospital setting: a meta-analysis. *Resuscitation*, 82(9):1130–1137, 2011.

[8] O. Baron, O. Berman, S. Kim, and D. Krass. Ensuring feasibility in location problems with stochastic demands and congestion. *IIE Transactions*, 41(5): 467–481, 2009.

[9] S. L. Bell and R. J. Williams. Dynamic scheduling of a system with two parallel servers in heavy traffic with resource pooling: asymptotic optimality of a threshold policy. *The Annals of Applied Probability*, 11(3):608–649, 2001.

[10] O. Berman. Dynamic repositioning of indistinguishable service units on transportation networks. *Transportation Science*, 15(2):115–136, 1981.

[11] O. Berman. Repositioning of distinguishable urban service units on networks. *Computers and Operations Research*, 8(2):105–118, 1981.

[12] O. Berman. Repositioning of two distinguishable service vehicles on networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 11(3):187–193, 1981.

[13] S Bhulai and G. Koole. A queueing model for call blending in call centers. *IEEE Transactions on Automatic Control*, 48(8):1434–1438, 2000.

[14] J.P.C. Blanc, P.R. de Wall, P. Nain, and D. Towsley. Optimal control of admission to a multi-server queue with two arrival streams. *IEEE Transactions on Automatic Control*, 37(6):785–797, 1992.

[15] B.E. Bledsoe. EMS myth #7: System Status Management (SSM) lowers response times and enhances patient care. *Emergency Medical Services*, 32 (9):158–159, 2003.

[16] Fernando Borrás and Jess T. Pastor. The ex-post evaluation of the minimum local reliability level: An enhanced probabilistic location set covering model. *Annals of Operations Research*, 111(1/4):51–74, 2002.

[17] O. Braun, R. McCallion, and J. Fazackerley. Characteristics of midsized urban EMS systems. *Annals of Emergency Medicine*, 19(5):536–546, 1990.

[18] Luce Brotcorne, Gilbert Laporte, and Frdric Semet. Ambulance location and relocation models. *European Journal of Operational Research*, 147(3): 451–463, 2003.

[19] David B. Brown and James E. Smith. Information relaxations, duality, and convex stochastic dynamic programs. *Operations Research*, 62(6):1394–1415, 2014.

[20] David B. Brown, James E. Smith, and Peng Sun. Information relaxations and duality in stochastic dynamic programs. *Operations Research*, 58(4): 785–801, 2010.

[21] Susan Budge, Armann Ingolfsson, and Erhan Erkut. Technical Note–Approximating vehicle dispatch probabilities for emergency service systems with location-specific service times and multiple units per location. *Operations Research*, 57(1):251–255, 2009.

[22] E. Carrizosa, E. Conde, and M. Mũnoz Márquez. Admission policies in loss queueing models with heterogeneous arrivals. *Management Science*, 44(3):311–320, 1998.

[23] A. Charnes and J. Storbeck. A goal programming model for the siting of multilevel EMS systems. *Socio-Economic Planning Science*, 14(4):155–161, 1980.

[24] Kenneth C. Chong, Shane G. Henderson, and Mark E. Lewis. The vehicle mix decision in emergency medical service systems. *Manufacturing & Service Operations Management*, 2015.

[25] R. Church and C. ReVelle. The maximal covering location problem. *Papers of the Regional Science Assocation*, 32(1):101–118, 1974.

[26] J.J. Clawson. Emergency medical dispatching. In In W.R. Roush, editor, *Principles of EMS Systems: A Comprehensive Text for Physicians*. American College of Emergency Physicians, Dallas, 1989.

[27] M. S. Daskin. A maximum expected covering location model: formulation, properties and heuristic solution. *Management Science*, 17(1):48–70, 1983.

[28] Vijay V. Desai, Vivek F. Farias, and Ciamac C. Moallemi. Pathwise optimization for optimal stopping problems. *Management Science*, 58(12): 2292–2308, 2012.

[29] Douglas G. Down and Mark E. Lewis. The N–network model with upgrades. *Prob. Eng. Inf. Sci.*, 24(02):171–200, 2010.

[30] Erhan Erkut, Armann Ingolfsson, and Güneş Erdoğan. Ambulance location for maximum survival. *Naval Research Logistics*, 55(1):42–58, 2008.

[31] Eugene A. Feinberg and Martin I. Reiman. Optimality of randomized trunk reservation. *Prob. Eng. Inf. Sci.*, 8(04):463–489, 1994.

[32] Noah Gans and Sergei Savin. Pricing and capacity rationing for rentals with uncertain durations. *Management Science*, 53(3):390–407, 2007.

[33] Noah Gans and Yong-Pin Zhou. A call-routing problem with service-level constraints. *Operations Research*, 51(2):255–271, 2003.

[34] M Gendreau, G Laporte, and F Semet. The maximal expected coverage relocation problem for emergency vehicles. *Journal of the Operational Research Society*, 57(1):22–28, 2005.

[35] Michel Gendreau, Gilbert Laporte, and Frdric Semet. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. *Parallel Computing*, 27(12):1641–1653, 2001.

[36] C.R. Gold. Prehospital advanced life support vs 'scoop and run' in trauma management. *Annals of Emergency Medicine*, 16(7):797–801, 1987.

[37] J.B. Goldberg. Operations research models for the deployment of emergency medical services vehicles. *EMS Management Journal*, 1(1):20–39, 2004.

[38] Linda V. Green and Peter J. Kolesar. Anniversary article: Improving emergency responsiveness with management science. *Management Science*, 50 (8):1001–1014, 2004.

[39] Suri Gurumurthi and Saif Benjaafar. Modeling and analysis of flexible queueing systems. *Naval Research Logistics*, 51(5):755–782, 2004.

[40] J. Michael Harrison. Heavy traffic analysis of a system with parallel servers: asymptotic optimality of discrete-review policies. *The Annals of Applied Probability*, 8(3):822–848, 1998.

[41] Martin B. Haugh and Leonid Kogan. Pricing American options: A duality approach. *Operations Research*, 52(2):258–270, 2004.

[42] S.G. Henderson and A.J. Mason. Ambulance service planning: simulation and data visualisation. In In M. L. Brandeau, F. Sainfort, and W. P. Pierskalla, editors, *Operations Research and Health Care: A Handbook of Methods and Applications*, pages 77–102. Kluwer Academic, Boston, 2004.

[43] Shane G. Henderson. Operations research tools for addressing current

challenges in emergency medical services. *Wiley Encyclopedia of Operations Research and Management Science*, 2011.

[44] A. Ingolfsson. EMS planning and management. In In G.S. Zaric, editor, *Operations Research and Health Care Policy*, pages 105–127. Springer Science and Business Media, New York, 2013.

[45] Armann Ingolfsson, Susan Budge, and Erhan Erkut. Optimal ambulance location with random delays and travel times. *Health Care Management Science*, 11(3):262–274, 2008.

[46] D.L. Isenberg and R. Bissell. Does advanced life support provide benefits to patients? A literature review. *Prehospital and Disaster Medicine*, 20(4): 266–270, 2005.

[47] L.M. Jacobs, A. Sinclair, A. Beiser, and R. B. D'agostino. Prehospital advanced life support: benefits in trauma. *The Journal of Trauma*, 24(1):8–12, 1984.

[48] C.J. Jagtenberg, S. Bhulai, and R.D. van der Mei. An efficient heuristic for real-time ambulance redeployment. *Operations Research for Health Care*, 4: 27–35, 2015.

[49] J. P. Jarvis. *Optimization in Stochastic Service Systems with Distinguishable Servers*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1975.

[50] J.P. Jarvis. Approximating the equilibrium behavior of multi-server loss systems. *Management Science*, 31(2):235–239, 1985.

[51] William C. Jordan and Stephen C. Graves. Principles on the benefits

of manufacturing process flexibility. *Management Science*, 41(4):577–594, 1995.

[52] Guoming Lai, François Margot, and Nicola Secomandi. An approximate dynamic programming approach to benchmark practice-based heuristics for natural gas storage valuation. *Operations Research*, 58(3):564–582, 2010.

[53] R.C. Larson. A hypercube queueing model for facility location and redistricting in urban emergency services. *Computers and Operations Research*, 1 (1):67–95, 1974.

[54] R.C. Larson. Approximating the performance of urban emergency service systems. *Operations Research*, 23(5):845–868, 1975.

[55] Mark E. Lewis, Hayriye Ayhan, and Robert D. Foley. Bias optimality in a queue with admission control. *Proability in the Engineering and Informational Sciences*, 13(3):309–327, 1999.

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

[57] Marvin B. Mandell. Covering models for two-tiered emergency medical services systems. *Location Science*, 6(1-4):355–368, 1998.

[58] A.J. Mason. Simulation and real-time optimised relocation for improving ambulance operations. In In B.T. Denton, editor, *Handbook of Healthcare Operations Management: Methods and Applications*, pages 289–317. Springer Science and Business Media, New York, 2013.

[59] Matthew S. Maxwell, Mateo Restrepo, Shane G. Henderson, and Huseyin Topaloglu. Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing*, 22(2):266–281, 2010.

[60] Matthew S. Maxwell, Shane G. Henderson, and Huseyin Topaloglu. Tuning approximate dynamic programming policies for ambulance redeployment via direct search. *Stochastic Systems*, 3(2):322–361, 2013.

[61] Matthew S. Maxwell, Eric Cao Ni, Chaoxu Tong, Shane G. Henderson, Huseyin Topaloglu, and Susan R. Hunter. A bound on the performance of an optimal ambulance redeployment policy. *Operations Research*, 62(5): 1014–1027, 2014.

[62] Maria E. Mayorga, Damitha Bandara, and Laura A. McLay. Districting and dispatching policies for emergency medical service systems to improve patient survival. *IIE Transactions on Healthcare Systems Engineering*, 3(1):39–56, 2013.

[63] Laura A. McLay. A maximum expected covering location model with two types of servers. *IIE Transactions*, 41(8):730–741, 2009.

[64] Laura A. McLay and Maria E. Mayorga. A dispatching model for server-to-customer systems that balances efficiency and equity. *Manufacturing & Service Operations Management*, 15(2):205–220, 2013.

[65] Laura A. McLay and Maria E. Mayorga. A model for optimally dispatching ambulances to emergency calls with classification errors in patient priorities. *IIE Transactions*, 45(1):1–24, 2013.

[66] W.F. McManus, D.D. Tresch, and J.C. Darin. An effective prehospital emergency system. *The Journal of Trauma*, 17(4):304–310, 1977.

[67] Bruce L. Miller. A queueing reward system with several customer classes. *Management Science*, 16(3):234–245, 1969.

[68] myFoxDetroit. No ambulance available for shooting victims, 2011. http://www.myfoxdetroit.com/story/18473311. Accessed July 6, 2015.

[69] Rahul Nair and Elise Miller-Hooks. Evaluation of relocation strategies for emergency medical service vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, 2137:63–73, 2009.

[70] Joe Naoum-Sawaya and Samir Elhedhli. A stochastic optimization model for real-time ambulance redeployment. *Computers & Operations Research*, 40(8):1972–1978, 2013.

[71] E. Lerzan Örmeci and Jan van der Wal. Admission policies for a two class loss system with general interarrival times. *Stochastic Models*, 22(1):37–53, 2006.

[72] J.P. Ornato, E.M. Racht, J.J. Fitch, and J.F. Berry. The need for ALS in urban and suburban EMS systems. *Annals of Emergency Medicine*, 19(12): 1469–1470, 1990.

[73] Felix Papier and Ulrich W. Thonemann. Capacity rationing in stochastic rental systems with advance demand information. *Operations Research*, 58 (2):274–288, 2010.

[74] Laura Pedersen. Too few paramedics to answer call: Union official. Toronto Sun, 2012. http://www.torontosun.com/2012/05/13/too-few-paramedics-to-answer-call-union-official. Accessed July 6, 2015.

[75] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley and Sons, 1994.

[76] L. C. G. Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12(3):271–286, 2002.

[77] O. Ryynänen, T. Iirola, J. Reitala, H. Pälve, and A. Malmivaara. Is advanced life support better than basic life support? A systematic review. *Scandinavian Journal of Trauma, Resuscitation, and Emergency Medicine*, 18 (62), 2010.

[78] E. S. Savas. Simulation and cost-effectiveness analysis of New York's emergency ambulance service. *Management Science*, 15(12):608–627, 1969.

[79] Sergei V. Savin, Morris A. Cohen, Noah Gans, and Ziv Katalan. Capacity management in rental businesses with two customer bases. *Operations Research*, 53(4):617–631, 2005.

[80] Verena Schmid. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. *European Journal of Operational Research*, 219(3):611–621, 2012.

[81] Richard F. Serfozo. Technical Note – An equivalence between continuous and discrete time Markov decision processes. *Operations Research*, 27(3): 616–620, 1979.

[82] David Simchi-Levi and Yehua Wei. Understanding the performance of the long chain and sparse designs in process flexibility. *Operations Research*, 60 (5):1125–1141, 2012.

[83] Jodie Sinnema. ER waits keep paramedics away in calls. Edmonton Journal, 2010. http://www.canada.com/story.html?id=c265934a-267e-40a4-a5e3-2b4872f9fd65. Accessed July 6, 2015.

[84] Jodie Sinnema. 'Busy' ambulance system causes concern of paramedics. Edmonton Journal, 2012. http://www2.canada.com/edmontonjournal/

news/archives/story.html?id=0a795570-199f-4ed0-91a5-c1c562743faa. Accessed July 6, 2015.

[85] C.M. Slovis, T.B. Carruth, W.J. Seitz, C.M. Thomas, and W.R. Elsea. A priority dispatch system for emergency medical services. *Annals of Emergency Systems*, 14(11):1055–1060, 1985.

[86] S. Stidham. Optimal control of admission to a queueing system. *IEEE Transactions on Automatic Control*, 30(8):705–713, 1985.

[87] Shaler Stidham and Richard Weber. A survey of Markov decision models for control of networks of queues. *Queueing Systems*, 13:291–314, 1993.

[88] J. Stout. System status management: The fact is, it's everywhere. *JEMS: A Journal of Emergency Medical Services*, 14(4):65–71, 1989.

[89] J. Stout, P.E. Pepe, and V.N. Mosesso Jr. All-advanced life support vs tiered-response ambulance systems. *Prehospital Emergency Care*, 4(1):1–6, 2000.

[90] Kanchala Sudtachat, Maria E. Mayorga, and Laura A. Mclay. A nested-compliance table policy for emergency medical service systems under relocation. *Omega*, 58:154–168, 2016.

[91] A.J. Swersey. The deployment of police, fire, and emergency medical units. In In S.M. Pollock, M.H. Rothkopf, and A. Barnett, editors, *Operations Research and the Public Sector*, pages 151–200. Elsevier, North Holland, Amsterdam, 1994.

[92] Constantine Toregas, Ralph Swain, Charles ReVelle, and Lawrence Bergman. The location of emergency service facilities. *Operations Research*, 19(6):1363–1373, 1971.

[93] John N. Tsitsiklis and Kuang Xu. On the power of (even a little) resource pooling. *Stochastic Systems*, 2(1):1–66, 2012.

[94] T. C. van Barneveld, S. Bhulai, and R. D. van der Mei. A dynamic ambulance management model for rural areas. *Health Care Management Science*, 2015.

[95] T.C. van Barneveld, S. Bhulai, and R.D. van der Mei. The effect of ambulance relocations on the performance of ambulance service providers. *European Journal of Operational Research*, 252(1):257–269, 2016.

[96] Thije van Barneveld. The minimum expected penalty relocation problem for the computation of compliance tables for ambulance vehicles. *INFORMS Journal on Computing*, 28(2):370–384, 2016.

[97] Rodney B. Wallace and Ward Whitt. A staffing algorithm for call centers with skill-based routing. *Manufacturing & Service Operations Management*, 7(4):276–294, 2005.

[98] Ward Whitt. Understanding the efficiency of multi-server service systems. *Management Science*, 38(5):708–723, 1992.

[99] B. Wilson, M.C. Gratton, J. Overton, and W.A. Watson. Unexpected ALS procedures on non-emergency ambulance calls: the value of a single-tier system. *Prehospital and Disaster Medicine*, 7(4):380–382, 1992.

[100] Y. Yue, L. Marla, and R. Krishnan. An efficient simulation-based approach to ambulance fleet allocation and dynamic redeployment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2012.

[101] L. Zhang. *Simulation Optimisation and Markov Models for Dynamic Ambulance Redeployment*. PhD thesis, The University of Auckland, New Zealand, 2012.