



Virginia Commonwealth University  
**VCU Scholars Compass**

---

Theses and Dissertations

Graduate School

---

2014

## Dispatch, Delivery, and Location Logistics for the Aeromedical Evacuation of Time-Sensitive Military Casualties Under Uncertainty

Benjamin Grannan  
*Virginia Commonwealth University*

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>



Part of the [Systems Engineering Commons](#)

© The Author

---

Downloaded from

<https://scholarscompass.vcu.edu/etd/3536>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact [libcompass@vcu.edu](mailto:libcompass@vcu.edu).

Dispatch, Delivery, and Location Logistics for the Aeromedical Evacuation of  
Time-Sensitive Military Casualties Under Uncertainty

A Dissertation submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy at Virginia Commonwealth University.

by

Benjamin C. Grannan  
Doctor of Philosophy

Director: Dr. Laura A. McLay, Associate Professor  
Department of Industrial & Systems Engineering, University of Wisconsin - Madison

Virginia Commonwealth University  
Richmond, Virginia  
August 2014

## Acknowledgment

A Ph.D. is not an achievement by one person. Many people have made this dissertation possible. I want to especially thank

- my loving wife Whitney for encouraging me to apply to the Ph.D. program, her patience during the long journey, and being my stability during everyday of our life.
- my adviser Dr. Laura McLay for giving me the opportunity to be her Ph.D. student, showing me what it takes to be a successful researcher, and sharing so much of her time over the past three years.
- Dr. Jason Merrick for his vision and effort in creating the Systems Modeling & Analysis Ph.D. program at VCU and investing in me with a doctoral assistantship.
- my dissertation committee members Dr. Paul Brooks, Dr. D'Arcy Mays, and Dr. Jose Dulá for providing feedback on my proposal and serving on my committee.
- Mom, Chris, Nick, Mom, Eddie, Grandma, and Uncle Dave for their continuous love and support. Their sacrifices have allowed me to follow my dream.
- my in-laws Dawn and George for everything. They have been an unwavering source of support.
- Bobby Crace for sixteen years and counting of friendship that includes reading my writing and helping improve the quality.

## Contents

Abstract	vii
1 Introduction and Motivation	1
1.1 History of Military Aeromedical Evacuation . . . . .	3
1.2 Literature Review . . . . .	7
1.2.1 Military Aeromedical Evacuation . . . . .	7
1.2.2 Dispatching Servers to Patients . . . . .	8
1.2.3 Facility Location Problems . . . . .	9
1.3 Dissertation Overview . . . . .	12
2 Dispatching Military Aeromedical Evacuation Assets and Transporting Casualties	14
2.1 Introduction . . . . .	14
2.2 Markov Decision Process Model . . . . .	17
2.3 Computational Example . . . . .	24
2.3.1 Problem Setup . . . . .	24
2.3.2 Policy Comparison . . . . .	28
2.3.3 Dispatching and Transporting Sensitivity . . . . .	30
2.4 Value Iteration Convergence Algorithm . . . . .	35
2.5 Theoretical Results . . . . .	36
2.6 Conclusions . . . . .	40
3 Locating and Dispatching Two Classes of Military Aeromedical Evacuation Assets	41
3.1 Introduction . . . . .	41
3.2 Literature Review . . . . .	43
3.3 Queuing Model . . . . .	47
3.4 Binary Linear Programming Model . . . . .	54
3.5 Computational Example . . . . .	57
3.6 Concluding Remarks . . . . .	65
4 Co-locating Military Aeromedical Evacuation Assets and Security Escorts	67
4.1 Introduction . . . . .	67
4.2 Literature Review . . . . .	68
4.3 Model . . . . .	70

4.4	Computational Example . . . . .	76
4.4.1	Base Case Example . . . . .	79
4.4.2	Sensitivity Analysis . . . . .	80
4.5	Conclusions . . . . .	83
5	Summary . . . . .	85
5.1	Limitations and Drawbacks . . . . .	85
5.2	Extensions . . . . .	86
5.3	Other Problem Domains . . . . .	87
5.4	Conclusion . . . . .	88
	Bibliography . . . . .	89
	Vita . . . . .	99

## List of Tables

2.1	Average utility when transporting a true high-priority casualty to the role II medical treatment facility and the role III medical treatment facility . . . . .	27
2.2	Average service times and average transport times (in hours) . . . . .	27
2.3	Input parameters and ranges considered for sensitivity analysis . . . . .	28
2.4	Proportion of calls for service at each location that are responded to by the closest aeromedical evacuation asset . . . . .	33
2.5	Proportion of true high-priority casualties transported to role III medical treatment facility based on the responding aeromedical evacuation asset . . . . .	34
3.1	Location and Busy Probabilities for Type A and Type B Air Assets in the Base Case Example . . . . .	61
3.2	Independence Correction Factors for Type A and Type B Air Assets in the Base Case Example . . . . .	61
3.3	Optimal Solution Versus Closest Server Heuristic Solution with Additional Type A Air Assets . . . . .	64
3.4	Optimal Solution Versus Closest Server Heuristic Solution with Varying Call Arrival Rate $\lambda$ . . . . .	65
4.1	Number of covered casualty locations by evacuation assets . . . . .	78
4.2	Distribution of each type of asset for the base case example solution . . . . .	79
4.3	Number of required escort assets to achieve $\alpha$ reliability . . . . .	81
4.4	Distribution of each type of asset with respect to varying upper bounds . . . . .	82

## List of Figures

1.1	9 line request for Army aeromedical evacuation . . . . .	2
1.2	The Sikorsky VS-300 helicopter nicknamed the Hoverfly. Used in service circa 1943. Photo from National Museum of the USAF [85]. . . . .	4
1.3	The Bell UH-1 Iroquois helicopter commonly known as the Huey. Used in service circa 1979. Photo from The Aviation History Online Museum [70].	5
1.4	The Sikorsky UH-60Q Black Hawk helicopter. Used in service present day. Photo from Federation of American Scientists [1] . . . . .	6
2.1	Time line during military aeromedical evacuation . . . . .	18
2.2	Geography of military aeromedical evacuation system with four casualty locations and four aeromedical evacuation assets . . . . .	25
2.3	Simulated policy comparison . . . . .	29
2.4	Sensitivity analysis on the proportion of true high-priority casualties $H'$ transported to the role 3 medical treatment facility, with respect to distance ratio $\theta$ . . . . .	31
3.1	25 Potential MTF sites in NATO Regional Command-South (represented by red crosses) . . . . .	58
3.2	Sensitivity Analysis of the Objective Function for the Number of Type A Air Assets and Call Arrival Rate $\lambda$ . . . . .	62
3.3	Contiguous 1st priority dispatching assignments for type A and type B air assets in the base case example. Casualty locations that comprise an air asset's 1st priority district are the same color. . . . .	63
4.1	25 Potential MTF sites in NATO Regional Command-South (represented by red crosses) . . . . .	77
4.2	Expected number of high and low-priority casualty location coverings . . .	80
4.3	Sensitivity of the objective function value to the number of type A evacuation assets and the number of escort assets . . . . .	81

## Abstract

### DISPATCH, DELIVERY, AND LOCATION LOGISTICS FOR THE AEROMEDICAL EVACUATION OF TIME-SENSITIVE MILITARY CASUALTIES UNDER UNCERTAINTY

By Benjamin C. Grannan, Doctor of Philosophy.

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2014.

Director: Dr. Laura A. McLay, Associate Professor, Department of Industrial & Systems Engineering, University of Wisconsin - Madison.

Effective aeromedical evacuation of casualties is one of the most important problems in military medical systems because high-priority casualties will not survive without timely medical care. The decision making process for aeromedical evacuation consists of the following components: (1) identifying which aeromedical evacuation asset (see figure 1) to dispatch to the casualty, (2) locating aeromedical evacuation assets strategically in anticipation of incoming demand, and (3) deciding which medical treatment facility to transport the casualty. These decisions are further complicated because prioritization of casualties is based on severity of injury while aeromedical evacuation assets and medical treatment facilities operate with varying capabilities. In this dissertation, discrete optimization models are developed to examine dispatch, delivery, and location logistics for the effective aeromedical evacuation of casualties in military medical systems.



## Introduction and Motivation

*“The standard for medical evacuation in Iraq was an hour...But Afghanistan is a lot tougher terrain. And so it came to my attention that they had settled on two hours. And I said: ‘Bulls–t. It’s going to be the same in Afghanistan as in Iraq’. And the medical guys pushed back on me and said: ‘No, no, it really doesn’t matter.’ And I said: ‘Well if I’m a soldier and I’m going out on patrol it matters to me.’”* – former U.S. Secretary of Defense Robert Gates

Effective medical evacuation (MEDEVAC) of wounded soldiers (casualties) in military operations is heavily reliant on reducing the time between injury and treatment which in turn increases the survivability of the combat soldier [88]. Transporting casualties to a hospital in a timely manner prevents the deteriorating health and potential death of casualties. The effective MEDEVAC of casualties also contributes to the potential psychological advantage for those participating in combat, who understand that medical assistance will come quickly if requested [8]. The U.S. Army uses a 9 line request form as shown in figure 1.1 to relay pertinent information when an evacuation is requested [38]. Military decision makers must interpret the 9 line request and make a quick decision regarding which medical evacuation (ground or air) asset to dispatch to the casualty event. Critical information found in the 9 line request include the geographical location and number of casualties to be evacuated. The medical evacuation of casualties is further complicated by triage according to precedence (urgent, priority, and routine) and the security of the pickup site. Military medical systems must make hard resource allocation decisions to reduce the time until treatment in order to maximize the survivability of casualties.

While ground evacuation is sometimes used, aeromedical evacuation assets such as helicopters are most commonly dispatched to evacuate casualties in military medical systems

REQUEST ARMY AIR MEDEVAC		
NOTE: Send secure or encrypt all items.		
LINE	ITEM	✓
1	Location of pick-up site	
2	Pick-up site radio frequency, call sign, and suffix	
3	# of patients by precedence (urgent, priority, routine)	
4	Special equipment required	
5	# of patients by type (litter, ambulatory)	
6	Security of pick-up site	
7	Method of marking pick-up site	
8	Patient nationality and status	
9	NBC considerations	

Figure 1.1: 9 line request for Army aeromedical evacuation

due to their speed and capability to provide en route medical care [17]. In addition to administering first level medical care, aeromedical evacuation assets also transport casualties between treatment facilities that offer higher levels of medical care as needed. Due to the severe nature of combat injuries, the most urgent casualties will not survive without the prompt evacuation and timely medical care provided by aeromedical evacuation assets. Therefore, this dissertation introduces new operations research models to study the hard but important logistical problems concerning aeromedical evacuation assets found in military medical systems. Decision support tools are needed to identify how to strategically locate and utilize scarce military aeromedical evacuation assets to maximize the survivability of the most urgent casualties and provide prompt evacuation to all casualties.

The following sections in this chapter explain the motivation for research on military medical systems. The first section summarizes the history of military aeromedical evacuation. The next section presents a literature review of the existing operations research models

to illustrate what and where this dissertation contributes to the literature. The last section presents a brief overview of the contents of each chapter of this dissertation.

## 1.1 History of Military Aeromedical Evacuation

The origin of military aeromedical evacuation of casualties has roots in the history of heavier-than-air flight itself. In 1908, the Wright brothers were awarded a contract from the U.S. War Department for the Army's first airplane five years after achieving their iconic first powered and sustained flight in Kitty Hawk, North Carolina [65]. Shortly afterwards, the U.S. Army Medical Corps lobbied the U.S. War Department for an aeromedical evacuation asset saying, "that thousands of hours and ultimately thousands of patients would be saved through the use of airplanes in air evacuation." [65]. By 1920 the Army had built the DeHavilland DH-4A, its first aircraft designed as an air ambulance [65]. The widespread use of fixed-wing aircraft for military aeromedical evacuation began in 1942 when the first U.S. aeromedical evacuation unit was activated [65]. Despite most medical evacuation from the front line by ground ambulance during this time, over one and a half million casualties were medically transported from front line to rear hospitals by airplane [65].

Despite the emergence of a dedicated aeromedical evacuation unit in World War II, the transition from airplane to helicopter was paramount in the development of effective aeromedical evacuation systems. The helicopter is a more flexible air asset, especially regarding take off and landing. The first flight of an U.S. helicopter was achieved by the Sikorsky's prototype VS-300 in 1939 [65]. The military began to utilize the functionality of the helicopter in 1943 when they ordered thirty four Sikorsky helicopters officially titled the R4 and nicknamed the Hoverfly, as shown in figure 1.2. The R4 was not originally designed to be an evacuation asset. In order to be converted to an air ambulance, patient transport litters had to be attached to the sides of the R4. The converted air ambulance was utilized by

Lt. Carter Harman who in 1944 flew the US Army's first helicopter aeromedical evacuation mission in Burma [65].



Figure 1.2: The Sikorsky VS-300 helicopter nicknamed the Hoverfly. Used in service circa 1943. Photo from National Museum of the USAF [85].

The first systematic use of helicopters for aeromedical evacuation of military casualties occurred in the Korean War because ground evacuation proved difficult due to tough terrain [65]. In 1950, the Army deployed four helicopter detachments each consisting of four H-13 Sioux helicopters to Korea with the primary mission of aeromedical evacuation of casualties [65]. By the end of the war about seventeen thousand and seven hundred casualties had been evacuated by helicopter [65].

Building upon the initial success during the Korean War, helicopters demonstrated their high levels of functionality as aeromedical evacuation assets during the Vietnam War. Bell UH-1 Iroquois helicopters, as shown in figure 1.3, were used to move almost nine hundred thousand U.S. and allied casualties during the Vietnam War [65]. The main contribution helicopter ambulances established during the Vietnam War was the now modern day practice of providing on board medical care during flight. The Bell UH-1 Iroquois was nicknamed

the Huey and was the Army's primary aeromedical evacuation asset until 1979. Army aeromedical evacuation missions adopted the call sign "Dustoff" for their motto: Dedicated, Unhesitating Service to Our Fighting Forces [65]. "Dustoff" is still the present day call sign used by all Army aeromedical evacuation flights [17].



Figure 1.3: The Bell UH-1 Iroquois helicopter commonly known as the Huey. Used in service circa 1979. Photo from The Aviation History Online Museum [70].

Since replacing the Bell UH-1 Huey in 1979, the Sikorsky UH-60 Black Hawk helicopter has been the Army's front line general purpose helicopter [17]. The Black Hawk greatly outperforms previous generations of service helicopters with a cruising speed of one hundred and fifty knots and max range of three hundred and fifteen nautical miles [1]. The Sikorsky UH-60Q (see figure 1.4) is an aeromedical evacuation variant of the UH-60. The specific medical features of the Sikorsky UH-60Q include six patient litters, on-board oxygen generation, and medical suction [1]. These on board medical systems allow medical technicians to administer life saving care immediately. The Sikorsky UH-60Q has red cross markings to designate an unarmed medical transport, per the Geneva Convention [17]. Since the Sikorsky UH-60Q is not armed, it must be escorted by security assets while operating in hostile enemy territory [17].



Figure 1.4: The Sikorsky UH-60Q Black Hawk helicopter. Used in service present day. Photo from Federation of American Scientists [1]

The other modern U.S. military asset frequently used in the aeromedical evacuation of casualties is the Air Force's HH-60G Pave Hawk helicopter [17]. This asset is a weaponized variant of the Sikorsky UH-60 Black Hawk. Specifically, the Pave Hawk can be outfitted with side window mounted M-60, M-240, or GAU-2B machine guns; two 7.62 mm mini guns mounted in the cabin windows; and two 0.50 caliber machine guns which can be mounted in the cabin doors [1]. In contrast to the UH-60Q, the self protection obtained from weapon systems allow the HH-60G Pave Hawk to fly solo without a security escort [17].

The aeromedical evacuation of military casualties has been linked to heavier-than-air flight since its inception. Transportation of casualties by helicopter provides a faster medium to begin medical treatment and deliver casualties to a higher level of medical care. As long as military conflicts exist, so too will the need to effectively evacuate casualties. Aeromedical evacuation will mirror modern aviation as it continues to improve and evolve. Future aeromedical evacuation assets will remain critical in the successful medical treatment of military casualties.

## 1.2 Literature Review

This dissertation introduces new operations research models to improve strategic decision making in military medical systems. The models focus on the management of aeromedical evacuation assets and investigate how to improve decision making. The literature described in this section is broken down into three categories of existing research: the first explicitly analyzes military aeromedical evacuation systems, the second examines how to dispatch servers to patients, and the third reviews facility location problems.

### 1.2.1 Military Aeromedical Evacuation

A number of existing research papers focus explicitly on systematically improving aeromedical evacuation of military casualties. Higgins [44] introduces the operational issues of U.S. Army aeromedical evacuation helicopters. Helicopter operational issues are pivotal in any research study involving casualties and aeromedical evacuation systems because speed of response directly increases likelihood of survival. Bastian [6] presents a multi-criteria decision analysis (MCDA) model using stochastic, mixed-integer goal programming to determine the minimum number of air assets needed at each medical treatment facility. This in turn maximizes the coverage of the theater-wide casualty demand and the probability of meeting that demand. Fulton et al. [34] introduce a two-stage stochastic optimization modeling framework for the medical evacuation (ground and air) of casualties. This framework identifies optimal casualty evacuation sites and medical treatment facility sites in response to stochastic demands for service. Bastian and Fulton [7] present a geospatial-based decision-support tool that identifies which air asset to launch in response to a casualty event given knowledge of terrain, aircraft location, and aircraft capabilities. Zeto et al. [87] examine the pre-location of air assets, along with type and quantity, to maximize the theater-wide coverage while balancing air asset reliability. Bouma [16] develops the Medical Evacuation

and Treatment Capabilities Optimization Model (METCOM) that considers policy effects on key measures of effectiveness, and then optimizes treatment and facility capacities for given casualty flows. Fulton et al. [35] present a Monte Carlo simulation to evaluate rules of allocation and planning considerations for Army air ambulance companies during major combat operations (MCO). Bastian et al. [8] examine the required capabilities of aeromedical evacuation of future U.S. aeromedical evacuation platforms by identifying the zero-risk aircraft ground speed. Bastian et al. [9] further examine three research issues surrounding future U.S. aeromedical evacuation platforms including optimal operational capabilities; trade-off considerations of different aircraft engines; and the effect of weaponizing the current MEDEVAC asset fleet on range, coverage radius, and response time. The models in this dissertation extend the work mentioned above by identifying new methodology to improve the performance of aeromedical evacuation of military casualties.

### 1.2.2 Dispatching Servers to Patients

Military medical systems share characteristics with civilian emergency medical systems. In particular, both systems must decide which server to dispatch to a call for service. The models summarized in this section focus on dispatching servers in civilian emergency medical systems. To improve response and transport times to the truly most critical patients, it is important to understand when to dispatch the closest server and when to ration that asset. McLay and Mayorga [63] present a Markov decision process model for dispatching servers to spatially-distributed patients that maximizes the fraction of patients who are responded to within a fixed time frame while allowing for the possibility of classification errors in initial patient classification. Several other papers have examined dispatch issues for civilian EMS and fire departments. Jarvis [46] introduces a Markov decision process for determining optimal dispatching policies for a single type of server. Daskin and Haghani [30] estimate the distribution of the first vehicle's arrival time at the scene when multiple vehicles are



simultaneously dispatched. Swersey [83] develops a Markov model for determining how many fire engines to send to prioritized fire calls that balances the costs associated with dispatching too few or too many. Ignall et al. [45] extend this model to account for calls and fire engines that are spatially distributed as well as provide a “preparedness” heuristic instead of exploring an optimal solution. Both Andersson and Varbrand [2] and Lee [54] propose similar “preparedness” heuristics for dispatching ambulances to calls. Ansari et al. [3] examine a dispatching list for multiple servers at a location and balance the workload among responding servers. Majzoubi et al. [55] introduce integer, linear, and nonlinear programming models in addition to an approximation model for dispatching and relocating ambulance vehicles. Emergency medical service systems also identify which hospital to transport patients, a decision similar to initial dispatching. In civilian systems, protocols from the medical director dictate which hospital an ambulance takes a patient leaving little room for selective decision-making. Shunko et al. [82] explore hospital transport decisions using game theory in the context of two competing hospitals that can send delay signals to turn away incoming ambulances. This dissertation introduces new models to more effectively dispatch servers to patients and transport patients to hospitals in military medical systems.

### 1.2.3 Facility Location Problems

In military medical systems, assets such as helicopters and medical treatment facilities must be located strategically to reduce the time until medical treatment of casualties. If assets are treated as facilities on a network, the rich facility location literature can be leveraged to improve logistics management in military medical systems. In this section, the facility location literature is summarized including minisum, minimax, and stochastic facility location problems.

Minisum location-allocation problems find medians among existing points to minimize the sum of the distances from each of the points to the median point. The  $p$ -median problem is a minisum problem that attempts to locate  $p$  facilities in relation to a set of weighted demand nodes such that the sum of the weighted distance between the nodes and facilities is minimized. The  $p$ -median problem has been used extensively in facility location models. Tansel et al. [84] survey the theory and development of the  $p$ -median problem. Also, Mladenović et al. [69] survey metaheuristic approaches for the  $p$ -median problem. Solution methods to the  $p$ -median problem are surveyed in Reese [73]. Hakimi [39, 41] proved one of the most famous results for the  $p$ -median problem on networks: when facilities are allowed to be located anywhere along the edges of a graph, there is always a collection of  $p$  vertices that minimizes the objective.

In contrast to minisum problems, minimax problems find centers among points that minimize the maximum distance from all demand points to the center point. The  $p$ -center problem is an example of a minimax problem used in facility location models. Kariv and Hakimi [50] present an algorithmic approach to finding  $p$ -centers in networks. The problem of finding a  $p$ -center was originated by Hakimi [39, 41]. Minieka [66] extends previous results for calculating the centers of a graph so that every point on every edge as well as all vertices are served. Hakimi et al. [40] discuss improvements and generalizations of techniques to find the 1-center of a network and a  $p$ -center of a tree. The  $p$ -center problem is also discussed in minimax location articles by Dearing and Francis [31], Handler [42, 43], and Goldman [37].

Stochastic facility location models consider the real-world situation where one or more of the problem input values are not known with certainty. The following models all incorporate stochastic input. Facility location models have considered stochastic demand (Manne [57], Bean et al. [11], and Carbone [19]). Stochastic travel times in a network is considered by Mirchandani [67] and Mirchandani and Odoni [68]. Berman and Odoni [15] and Berman and

LeBlanc [12] consider the effect of uncertain travel times in a network on facility location-relocation problems. Daskin [27, 29] considers the situation where facilities are assumed to be busy with probability  $p$  and formulates an extension of the maximal covering problem (see above). In a different approach than Daskin, ReVelle and Hogan [75, 76] and ReVelle and Marianov [77] handle the stochastic nature of server availability by locating servers that covers demand with a reliability threshold. Marionov and ReVelle [58] extend the reliability model to incorporate multiple vehicle types for a given joint reliability threshold. As described above the  $p$ -median problem is central to facility location models. Weaver and Church [86] present computational procedures for the solution of stochastic  $p$ -median problems.

A subset of stochastic facility location models leverage spatial queuing theory to compute system statistics that improve location decision making. Larson [51] is the seminal reference in the literature, presenting the Hypercube model. The approximate Hypercube also developed by Larson [52] has  $N$  (number of servers) non-linear equations versus  $2^N$  equations in the exact model. There have been several extensions to the Hypercube model. Jarvis [47] incorporates dependencies on casualty call type for service times with servers that are distinguishable by location. Chelst and Barlach [21] examine two units dispatched to a single call. Batta et al. [10] include the dispatch probability estimates from the Hypercube model to locating ambulances. Budge et al. [18] examine multiple servers co-located at a single station. McLay [61] also develops an extension to the Hypercube model for two types of servers while responding to multiple types of customers. Mandell [56] also considers multiple types of medical units through a two dimensional queueing model. The Hypercube spatial queuing model has also been extended in Jarvis [46], Chelst and Jarvis [22], and Larson and McKnew [53]. Marianov and ReVelle [59] also utilize spatial queuing theory to develop the queueing maximal availability location problem. Berman [13] and Berman et al. [14] leverage spatial queuing theory to explore the location of a single facility on networks

with random travel times and demands.

Another way stochastic facility location models account for unknown input values is through scenario planning. In this context, different future scenarios represent the possible values for parameters that can vary. An overview of scenario planning is found in Georgantzas and Acar [36] and Vanston et al. [49]. Scenario planning can be used to minimize the expected cost of facility location planning over all scenarios (Sheppard [80]) or maximize the number of covered demands over all possible scenarios (Schilling [78]). Unknown travel times and/or demand has been examined via scenario planning by Serra and Marianov [79] and Carson and Batta [20]. In the dynamic, or time-dependent, location problem, Jornsten and Bjorndal [48] use scenario planning to minimize the expected cost across all scenarios. Current et al. [26] use scenario planning to consider the problem when the number of facilities sited is unknown. Daskin et al. [28] present the  $\alpha$ -reliable  $p$ -minimax regret problem that optimizes the worst-case performance over a set of scenarios. Averbakh and Berman [4, 5] focus on polynomial time algorithms to solve minimax regret problems.

### 1.3 Dissertation Overview

This dissertation introduces novel operations research models for the dispatching, delivering, and locating logistics of aeromedical evacuation assets in military medical systems. Discrete optimization models that leverage Markov decision processes, integer programming, and spatial queuing theory are developed to provide critical insights into the operation of aeromedical evacuation of casualties. The models in this dissertation are applied to realistic examples to illustrate the impact and relevance of the results.

This dissertation is organized as follows. Chapter 2 formulates an undiscounted, infinite horizon Markov decision process (MDP) model that examines the interrelated decisions of how to optimally dispatch aeromedical evacuation assets to calls for service and transport

casualties to medical treatment facilities. The model accounts for errors made during triage of casualties to investigate the revelation of information over time. The model also allows the batch arrival of casualties to the system. The MDP is solved with a value iteration algorithm. The optimal policy is compared to three heuristic casualty transport policies.

Chapter 3 develops a novel binary linear programming (BLP) model to optimally locate two types of aeromedical evacuation assets and construct response districts using a dispatch preference list. Additionally, the BLP model balances the workload among assets and enforces contiguity in the first assigned locations for each air asset. The objective of the BLP model is to maximize the proportion of high-priority casualties responded to within a prespecified time threshold while meeting performance benchmarks to other types of casualties. A spatial queuing approximation model is derived to provide inputs to the BLP model, which reflects the underlying spatial queuing dynamics of the system. The models of Chapter 3 are applied to a computational example that reflects realistic military data.

In Chapter 4, a IP model examines the hard problem of co-locating aeromedical evacuation assets and security escorts in military medical systems. Aeromedical evacuation assets that are unarmed must be escorted when operating in enemy territory. Waiting for an available escort asset increases the time until medical treatment for casualties. The IP model strategically locates both armed and un-armed aeromedical evacuation assets as well as security escorts to maximize the survivability of casualties in military medical systems.

The dissertation is summarized in Chapter 5. Disclosure of the drawbacks and limitations of the operations research models in this dissertation are identified and discussed. Several extensions to the research in this dissertation are described for future work. Extensions allow the research to continue to evolve and increase the usefulness of the models to decision makers. The models in this dissertation are derived specifically for military medical systems. However, the applicability of the models to other problem domains is discussed.

## Dispatching Military Aeromedical Evacuation Assets and Transporting Casualties

### 2.1 Introduction

Effective medical evacuation (MEDEVAC) of wounded soldiers (casualties) in military operations is important to the survivability of the combat soldier [88]. Transporting casualties to a medical treatment facility in a timely manner prevents the deteriorating health and potential death of casualties. The effective MEDEVAC of casualties also contributes to the potential psychological advantage for those participating in combat, who understand that medical assistance will come quickly once requested [8].

This chapter focuses on the dispatch and transport of casualties in United States military systems. Casualties arrive as calls for service, where dispatchers interpret the call detailing a casualty event and make a resource allocation decision regarding which MEDEVAC asset to dispatch to the casualty and later select which medical treatment facility to transport the casualty [17]. The MEDEVAC asset dispatched also transports the casualty to the medical treatment facility (i.e., a different MEDEVAC asset would not transport the casualty based on additional information collected at the scene) and therefore, the dispatch and transport decisions are interrelated.

While this chapter focuses on a model configuration for the United States, the model is applicable to other countries. Effective evacuation of casualties is an important problem shared across all countries that support combat troops. Moreover, issues examined in this chapter, such as imperfect initial triage, the collection of new information over time, and medical guidelines for transporting military casualties, are generally shared across countries,

as evidenced by a recent MEDEVAC summit held in London, England in October 2013 [64].

Identifying effective policies for dispatching air MEDEVAC assets and transporting casualties can be counter-intuitive. A fleet of potential air MEDEVAC assets are distinguishable by their base location, and therefore articulate different response times to casualties. Likewise, medical treatment facilities are distinguishable by both the capable level of care, i.e., a role 2 medical treatment facility versus a role 3 medical treatment facility, and the proximity of the medical treatment facility to the casualty location. Further, there exists a triage scheme within military evacuation systems, in which the categories used to rank injuries for precedence in evacuation are as follows [17]:

- “CAT A”: Alpha category includes urgent casualties that need to be treated within one hour.
- “CAT B”: Bravo category includes priority casualties that need to be treated within four hours.
- “CAT C”: Charlie category includes routine casualties that need to be treated within twenty-four hours.

The evacuation triage system lends itself to sub-categorizing casualties based upon priority. For example, a system that identifies only high-priority and low-priority casualties, the calls for service which have been categorized as CAT A could be seen as high-priority, while all other calls for service could be seen as low-priority. An alternative classification would be treating both the CAT A and CAT B calls for service as high-priority and the CAT C calls for service as low-priority. The prioritization scheme can be generalized for systems with more than two priority levels.

Military aeromedical evacuation systems aim to transport CAT A casualties to a medical treatment facility within one hour, a practice commonly known as the Golden Hour [17].

The fundamental idea of the Golden Hour is that mortality is least likely to occur if initial treatment of a severe trauma casualty begins within one hour post injury. A military evacuation system is evaluated against the Golden Hour standard to increase survivability of the most urgent CAT A casualties. Improving the logistics of MEDEVAC systems to meet the Golden Hour standard is an important problem found frequently in the popular press ([71, 32, 81]).

The Golden Hour performance measure evaluates time until treatment of a casualty, and it is in contrast with performance measures used by civilian emergency medical systems (EMS). Nearly all civilian EMS systems evaluate performance according to response times as opposed to casualty delivery times [62]. As a result, nearly all research in civilian systems focuses on triage accuracy and initial dispatch decisions. However, the importance of triage on resource allocation decisions is well-documented area in civilian systems [see 24, 33], and this issue becomes more complex in military MEDEVAC systems because dispatch and transport decisions are interrelated and more accurate information is collected at the scene.

This chapter formulates a Markov decision process (MDP) model to solve a MEDEVAC asset dispatching and casualty transporting problem with two interrelated types of decisions: first, how to initially dispatch location-dependent air MEDEVAC assets to location-dependent casualties, and second, how to identify distinguishable hospitals to transport the casualties. Both types of decisions indirectly affect the high-priority casualty's likelihood of survival, which is dependent upon time until treatment in a medical treatment facility [25]. To gain insight into military aeromedical evacuation systems, the MDP model determines how to maximize the long-run average Golden Hour reward over the truly high-priority casualties while also providing timely evacuation to low-priority casualties. The MDP model allows for classification errors in the initial triage, in which a truly low-priority casualty may be initially classified as high-priority, and vice-versa, thus leading to dispatch decisions with imperfect information. However, upon arrival at the scene, it is assumed that the medics



from the responding air MEDEVAC asset accurately diagnose the severity of each call thus make transport decisions to the medical treatment facility with perfect information. The MDP model also accounts for batch, or multiple, casualties in a call for service.

This chapter is organized as follows. Section 2.2 outlines the novel MDP model. A computational example of the U.S. configuration is included in Section 2.3. The Value Iteration algorithm is given in section 2.4 and some theoretical results are presented in section 2.5. Concluding remarks and future work are given in Section 2.6.

## 2.2 Markov Decision Process Model

This section presents the MDP model for dispatching air MEDEVAC assets and transporting prioritized casualties in a military aeromedical evacuation system with imperfect information during triage. The model parameters depend on the elapsed time during the treatment of each casualty, because military medical evacuation systems are evaluated by the transfer of high-priority casualties to the medical treatment facility before the Golden Hour. There are seven time steps to a military aeromedical evacuation [6]:

1. Notification time (call arrival).
2. MEDEVAC asset departure time ("wheels up").
3. Arrival at the scene.
4. Departure from the scene.
5. Arrival at the medical treatment facility.
6. Transfer of the casualty to medical treatment facility.
7. Arrival at MEDEVAC asset home location (return to service).

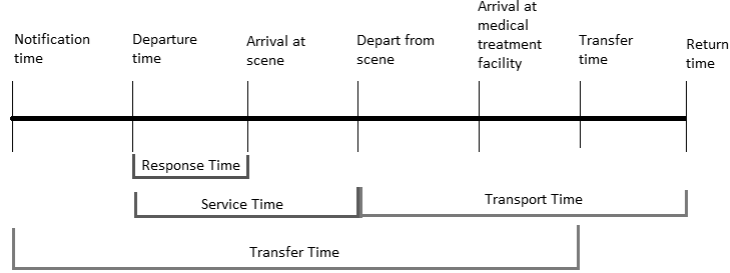


Figure 2.1: Time line during military aeromedical evacuation

Figure 2.1 presents the four time intervals used throughout the remainder of this chapter. *Response time* is the length of time from departure time (2) to the MEDEVAC asset arrival at the scene (3). *Service time* is the length of time from departure time (2) to leaving the scene (4). *Transport time* is length of time from the MEDEVAC asset leaving the scene (4) to returning to its home station after transporting a casualty (7). *Transfer time* is the length of time from injury (1) to the casualty being transferred to a medical facility (6).

The input parameters of the MDP model are summarized next, followed by the system dynamics.

$n$  = the number of casualty locations,

$m$  = the number of air MEDEVAC assets, each at a fixed home location,

$d$  = the number of medical treatment facilities,

$R$  = the *classified* risk level during triage, with  $R \in \{H, L\}$ , where  $H$  ( $L$ ) denotes classified high-risk (low-risk),

$r$  = the *true* risk level, with  $r \in \{H', L'\}$ , where  $H'$  ( $L'$ ) denotes truly high-risk (low-risk),

$\lambda$  = the call arrival rate,

$X$  = random variable representing the number of casualties  $X \in \{1, 2, \dots, N\}$  arriving in a batch arrival,

$P_i^X$  = the conditional probability that a batch call for service with  $X$  casualties arrives at location  $i$ , given that a call arrives,  $i = 1, 2, \dots, n, X = 1, 2, \dots, N$ ,

$P_{R|i}^X$  = the conditional probability that a batch call for service with  $X$  casualties arrives at location  $i$  has classified risk level  $R \in \{H, L\}$ , given that a call arrives,  $i = 1, 2, \dots, n, X = 1, 2, \dots, N$ ,

$P_{r|R \cap i}^X$  = the conditional probability that a batch call for service with  $X$  casualties and classified risk level  $R \in \{H, L\}$  has true risk level  $r \in \{H', L'\}$ ,  $i = 1, 2, \dots, n$ ,

$\mu_{ij}^X$  = the expected service time when MEDEVAC asset  $j$  responds to a batch call for service with  $X$  casualties at location  $i$ ,  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, m, X = 1, 2, \dots, N$ ,

$\delta_{ijk}$  = the expected transport time when MEDEVAC asset  $j$  transports a batch of casualties from location  $i$  to medical treatment facility  $k$  where  $j = 1, 2, \dots, m, i = 1, 2, \dots, n$ , and  $k = 1, 2, \dots, d$ ,

$u_{ijk}^X$  = the expected utility when MEDEVAC asset  $j$  transports a batch of  $X$  casualties with true risk level  $r \in \{H', L'\}$  from location  $i$  to medical treatment facility  $k$ , where  $j = 1, 2, \dots, m, i = 1, 2, \dots, n, X = 1, 2, \dots, N$ , and  $k = 1, 2, \dots, d$ .

The state variable reflects the positions of each of the  $m$  MEDEVAC assets and thus can be represented by the  $m$ -dimensional vector  $s$ , with  $s(t) = (s_1, s_2, \dots, s_m)$ . To describe the state space in a succinct way, we describe all possible values that each component of the

state space can take. In state  $s(t)$ , MEDEVAC asset  $j$  has three possible types of values corresponding to the three possible events in the system 1) asset  $j$  can be sent to a call for service that arrives, while servicing this call at the scene,  $s_j$  is described by a call location  $i$ , a classified priority  $R$ , and a batch size  $X$ , 2) asset  $j$  finishes service at the scene of service and begins transporting casualties to a medical treatment facility denoted by  $D_k$ , and 3) a busy MEDEVAC asset becomes free and returns to its home location ( $s_j = 0$ ). Note that these possible values for  $s_j$ — $(i, R, X)$ ,  $D_k$ ,  $0$ —are all distinct values that are mapped to integers in the computational implementation of the value iteration algorithm used to solve the model. The total number of states is equal to  $(1 + 2Nn + nd)^m$ . Although the MDP model suffers from the so-called “curse of dimensionality,” we will explore conditions under which the state space can be made smaller in Section 4.

Only one component of the state variable changes after an event occurs at time  $t$ . Therefore,  $s(t+1) = s(t)$  except for component  $s_j$  in  $s(t+1)$ . Let  $\phi$  denote the new value of  $s_j$  in state  $s(t+1)$ , either  $(i, R, X)$ ,  $D_k$ , or  $0$ . Let the transition function  $s(t+1) = S^M(s(t)|s_j = \phi)$  capture the new state at time  $t+1$ .

The following assumptions are made in the model. First, if a call for service arrives, an available MEDEVAC asset must be dispatched to the casualty if any are available. Otherwise the call is assumed lost to our system. This assumption is acceptable because in practice, military systems leverage other assets to treat these casualties [17]. Second, service cannot be preempted and air MEDEVAC assets cannot be rationed in expectation of in-coming calls for service. Third, a MEDEVAC asset selects a medical treatment facility destination immediately prior to departing from the scene, and immediately after reassessing the casualty risk to obtain the true risk level  $r$ . Therefore, transportation of casualties is made with information of the true risk level  $r \in \{H', L'\}$ . This can be contrasted with the dispatch decision which is made with the potentially inaccurate triage classification. Thus, the interrelated decisions of dispatch and transport capture the revelation of information

of each casualty's risk level over time. Fourth, the MEDEVAC asset that responds to the casualty must transport the casualty to the medical treatment facility. Fifth, batch arrivals of casualties at a location can be transported by a single responding MEDEVAC asset, that is the capacity of an air MEDEVAC asset is greater than or equal to the number of casualties in a batch. This assumption is reasonable, since in practice, the capacity of MEDEVAC asset is larger enough to transport virtually all batched casualties that arrive [6]. Lastly, risk levels are assessed on a batch level, not a casualty level. A single asset responds to a batch, not individual casualties, and therefore, risk levels assigned on the batch level are practical and easier to operationalize.

The objective of the MDP model is to determine which MEDEVAC asset to dispatch to a casualty and identify which medical treatment facility to transport a casualty, for each state in the state space, so that the expected number of truly high-priority calls that arrive at a medical treatment facility within one hour per stage is maximized.

The optimality equations for the infinite-horizon, average cost MDP model are given next, where  $I_{\{s_j=(i,R,X)\}}$ ,  $I_{\{s_j=D_k\}}$ , and  $I_{\{s_j=0\}}$  are indicator functions representing MEDEVAC asset  $j$  is serving  $X$  casualties at location  $i$ , traveling to medical treatment facility  $k$ , and being idle, respectively. An infinite-horizon MDP model with steady-state parameters is appropriate because of the duration of military operations. We use uniformization to convert a continuous time MDP model into an equivalent discrete time MDP model. To apply uniformization, the maximum rate of transitions is determined to be  $\gamma = \lambda + \sum_{j=1}^m \beta_j$ , where  $\beta_j = \max \left\{ \max_{i,X} \left\{ \frac{1}{\mu_{ij}^X} \right\}, \max_{i,k} \left\{ \frac{1}{\delta_{ijk}} \right\} \right\}$ ,  $j = 1 \dots m$ . Note that  $g$  is the optimal average utility per stage and  $v_t(s(t))$  is a relative value function in state  $s(t) = (s_1, s_2, \dots, s_m)$ , and  $A_1(s(t))$  and  $A_2(s(t))$  represent the set of dispatching and transporting actions available in state  $s(t)$  during iteration  $t$ , respectively.

$$g + v(s(t)) = \frac{1}{\gamma} \left[ \sum_{k=1}^d \sum_{j=1}^m \sum_{i=1}^n (\delta_{ijk})^{-1} I_{\{s_j=D_k\}} v(S^M(s(t)|s_j=0)) \right] \quad (2.1a)$$

$$+ \sum_{i=1}^n \sum_{X=1}^N \sum_{R \in \{H,L\}} \lambda P_i^X P_{R|i}^X \max_{j \in A_1(s(t))} \{v(S^M(s(t)|s_j=(i,R,X)))\} \quad (2.1b)$$

$$+ \sum_{i=1}^n \sum_{X=1}^N \sum_{j=1}^m \sum_{R \in \{H,L\}} \sum_{r \in \{H',L'\}} (\mu_{ij}^X)^{-1} I_{\{s_j=(i,R,X)\}} P_{r|R \cap i}^X \max_{D_k \in A_2(s(t))} \{v(S^M(s(t)|s_j=D_k)) + \gamma u_{ijk}^X\} \quad (2.1c)$$

$$+ \left( \gamma - \lambda - \sum_{i=1}^n \sum_{X=1}^N \sum_{j=1}^m \sum_{R \in \{H,L\}} (\mu_{ij}^X)^{-1} I_{\{s_j=(i,R,X)\}} - \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^d (\delta_{ijk})^{-1} I_{\{s_j=D_k\}} \right) v(s(t)) \right] \quad (2.1d)$$

The four lines in the value functions (2.1) represent the three events that result in a change in the state variable plus the fourth “event” that nothing changes between stages (line (2.1d)). The first line (2.1a) accounts for the event of busy air MEDEVAC assets completing service after transporting a casualty, where  $(\delta_{ijk})^{-1}$  captures the mean transport time and  $v(S^M(s(t)|s_j=0))$  represents the value of the state when MEDEVAC asset  $j$  returns home and is available for service. Line (2.1b) accounts for the dispatch of a MEDEVAC asset to a batch of casualties of size  $X$  that arrives to the system with classified priority  $R$ . Here,  $\lambda P_i^X P_{R|i}^X$  captures the probability of a call for service of  $X$  casualties at casualty location  $i$  and of classified risk level  $R$  arrives to the system. The second part of line (2.1b) selects the MEDEVAC asset  $j$  to the incoming call for service  $(i, R, X)$  that maximizes the value of dispatching a MEDEVAC asset. Line (2.1c) accounts for the decision to transport a batch of casualties to a medical treatment facility, which occurs when service at the scene is completed. Here,  $(\mu_{ij}^X)^{-1}$  represents the mean service time at the scene, and  $P_{r|R \cap i}^X$  captures the conditional probability of the casualty’s true risk level  $r$  given its classified as risk level  $R$  and location  $i$ . The second part of line (2.1c) selects the medical treatment facility that maximizes the value of transporting casualties to a medical treatment facility, where  $u_{ijk}^X$

represents the reward received when MEDEVAC asset  $j$  transports  $X$  casualties at casualty location  $i$  of true risk level  $r$  to medical treatment facility  $k$ . Section 2.4 summarizes a value iteration convergence algorithm using the corresponding  $N$ -stage finite-horizon optimality equations [see 72].

Lastly, the assessment of the severity of the calls is imperfect during triage, resulting in possible mismatches between the classified risk level  $R \in \{H, L\}$  and the true risk level  $r \in \{H', L'\}$  as captured by the  $P_{r|R \cap i}$  parameters. The accuracy of triage classification is assumed known, such as from past system performance data. Since military aeromedical evacuation systems are evaluated according to the response to casualties that are truly the most critical ( $H'$ ), it is of interest to match the classified risk levels  $R$  to the true risk levels  $r$ . Let  $\alpha$  denote the ratio of the proportion of classified high-risk casualties that are truly high-risk to the proportion of classified low-risk casualties that are truly high-risk,

$$\alpha^X = \frac{P_{H'|H}^X}{P_{H'|L}^X}.$$

Therefore,  $\alpha^X$  can be interpreted as the accuracy of the triage for high-priority casualties, which we assume is independent of the casualty location. When  $\alpha^X = 1.0$ , the classified high-risk casualties are at least as likely to be truly high-risk as classified low-risk casualties. As  $\alpha^X \rightarrow \infty$ , the set of truly high-risk casualties is a subset of classified high-risk casualties (when  $P_{H'} \leq P_H$ ). In the MDP model, input parameter  $P_{r|R \cap i}^X$  is a function of  $\alpha^X$ , and can be computed as follows. Note that  $\alpha^X = \frac{P_{H'|H \cap i}^X}{P_{H'|L \cap i}^X}$  since triage accuracy is independent of the casualty location. Rearranging and applying Bayes rule yields:

$$P_{H'|H \cap i}^X = \frac{\alpha^X P_{H'}^X}{P_{L \cap i}^X} (P_{i|H'}^X - P_{H \cap i|h}^X).$$

Rearranging, noting that  $P_{L \cap i | H'}^X + P_{H \cap i | H'}^X = P_{i | H'}^X$ , and applying Bayes rule again yields:

$$P_{H' | H \cap i}^X = \frac{\alpha^X P_{H' | i}^X}{P_{H | i}^X + (1/\alpha^X) P_{L | i}^X}.$$

The analogous procedure can be applied to classified low-risk calls, yielding  $P_{H' | L \cap i}^X$ .

Section 2.5 contains theoretical results related to transportation policies in the MDP model proposed in this chapter. The first result indicates that if the expected times to transfer a casualty at two medical treatment facilities are the same, it is optimal to transport to the medical treatment facility with the highest utility. The second result indicates that if the utilities for transferring a casualty at two medical treatment facilities are the same, it is optimal to transport to the medical treatment facility with the shortest expected time until transfer. However, these results are not actionable when there are trade offs between transfer time and quality of care. Therefore, we examine the trade offs in the computational example in the following section.

## 2.3 Computational Example

### 2.3.1 Problem Setup

Consider a military aeromedical evacuation system example in support of an U.S. Army brigade, where the location of casualties to be evacuated and medical treatment facilities are both known. As described in [8], the area of operations for future U.S. Army brigades (a military unit with over 3,000 personnel) is 90,000 km<sup>2</sup>, and a sub-area of 900 km<sup>2</sup> containing four air MEDEVAC assets and four casualty locations ( $m = n = 4$ ) is proposed for analysis here (see Figure 2.2).

Each square in Figure 2.2 is 15 kilometers long and 15 kilometers wide. Travel times are computed using the Euclidean distance between locations and MEDEVAC assets have a



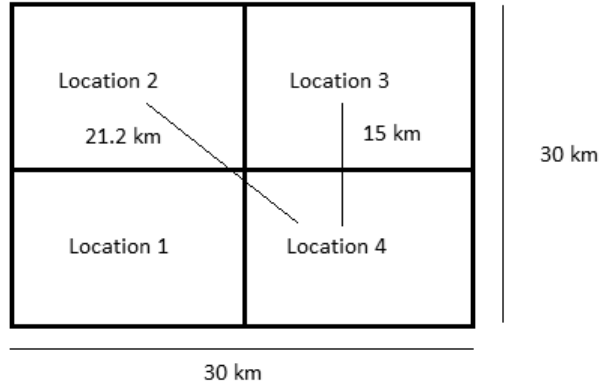


Figure 2.2: Geography of military aeromedical evacuation system with four casualty locations and four aeromedical evacuation assets

flight speed of 155 nautical miles per hour (knots) [6].

There are two distinguishable medical treatment facilities (i.e.,  $d = 2$ ) available in support of the Army brigade—the first is a role 2 medical treatment facility denoted  $k = 2$ , and the second is a role 3 medical treatment facility denoted  $k = 3$ . The utilities  $u_{i,j,k,H'}$  are set based on the time it takes for asset  $j$  to transfer a casualty at location  $i$  to medical treatment facility  $k$ . Suppose this transport time takes  $t$  hours, on average, to a role 2 medical treatment facility. Then, the modified Golden Hour utility function as a function of  $t$  is  $\max\{\frac{-t}{60} + 1, 0\}$ . The utilities for transporting a casualty to a role 3 medical treatment facility is assumed to be a factor of  $\Gamma$  increase over the role 2 utility. Moreover, we assume the utility for transporting true low-priority casualties is zero, since these casualties are expected to survive regardless of where they are transported.

We focus on the disparity in proximity and medical treatment quality between the role 3 and role 2 medical treatment facilities in this example. The relative utilities and travel times between these two types of facilities are pertinent when managing military aeromedical

evacuation system logistics. Therefore, define the *distance ratio*  $\theta$  as the relative travel distance to the role 3 medical treatment facility as compared to the role 2 medical treatment facility for each call location  $i$  and responding MEDEVAC asset  $j$ :

$$\theta = \frac{\delta_{i,j,3}}{\delta_{i,j,2}}, i = 1, 2, \dots, n, j = 1, 2, \dots, m.$$

When  $\theta = 2$ , the relative transport time to the role 3 medical treatment facility is twice the relative transport time to the role 2 medical treatment facility, given the same MEDEVAC asset  $j$  and demand location  $i$ .

Define the *reward ratio*  $\Gamma$  to distinguish between the system utility received transporting a truly high-priority casualty to the role 3 medical treatment facility and the utility received transporting a truly high-priority casualty to the role 2 medical treatment facility.

$$\Gamma = \frac{u_{i,j,3,H'}}{u_{i,j,2,H'}}, i = 1, 2, \dots, n, j = 1, 2, \dots, m.$$

When  $\Gamma = 2$ , the role 3 medical treatment facility is twice as medically capable as the role 2 medical treatment facility, due to better resources, surgeons on staff, etc.

Table 2.1 reports the average utility when transporting to the role 2 medical treatment facility or the role 3 medical treatment facility, for each MEDEVAC asset  $j$  and casualty location  $i$  under the base case of the military aeromedical evacuation system in this example.

Many of the transition probabilities depend on the length of time until a busy MEDEVAC asset becomes free, for each dispatching or transporting action, or a call for service ends. Table 2.2 presents the average service and transport time when responding to a call for service under the base case of the military aeromedical evacuation system in this example.

The remaining transition probabilities depend on the rate at which casualties arrive to the

Table 2.1: Average utility when transporting a true high-priority casualty to the role II medical treatment facility and the role III medical treatment facility

$u_{i,j,2,H'}$	$i = 1$	$i = 2$	$i = 3$	$i = 4$
$j = 1$	0.364	0.364	0.291	0.291
$j = 2$	0.312	0.417	0.312	0.269
$j = 3$	0.291	0.364	0.364	0.291
$j = 4$	0.312	0.343	0.312	0.343
$u_{i,j,3,H'}$	$i = 1$	$i = 2$	$i = 3$	$i = 4$
$j = 1$	0.456	0.456	0.363	0.363
$j = 2$	0.390	0.520	0.390	0.336
$j = 3$	0.363	0.456	0.456	0.363
$j = 4$	0.390	0.429	0.390	0.429

Table 2.2: Average service times and average transport times (in hours)

$\mu_{i,j}$	$i = 1$	$i = 2$	$i = 3$	$i = 4$
$j = 1$	0.500	0.552	0.574	0.552
$j = 2$	0.552	0.500	0.552	0.574
$j = 3$	0.574	0.552	0.500	0.552
$j = 4$	0.552	0.574	0.552	0.500
$\delta_{i,j,2}$	$i = 1$	$i = 2$	$i = 3$	$i = 4$
$j = 1$	0.188	0.136	0.188	0.210
$j = 2$	0.136	0.083	0.136	0.157
$j = 3$	0.188	0.136	0.188	0.210
$j = 4$	0.210	0.157	0.210	0.231
$\delta_{i,j,3}$	$i = 1$	$i = 2$	$i = 3$	$i = 4$
$j = 1$	0.376	0.271	0.376	0.419
$j = 2$	0.271	0.167	0.271	0.315
$j = 3$	0.376	0.271	0.376	0.419
$j = 4$	0.419	0.315	0.419	0.462

system. Calls arrive according to a Poisson process with parameter  $\lambda = 3.0$  calls per hour. The distribution of calls  $P_i$  are unevenly spaced across the four casualty locations. There is a “hotbed” of activity and more frequent calls for service in location 2. Table 2.3 presents the base case set of input parameters and the corresponding ranges used for sensitivity analysis in this example.

Table 2.3: Input parameters and ranges considered for sensitivity analysis

Input parameter	Base case value	Parameter range
Medical treatment facilities $d$	2	-
Casualty locations $n$	4	-
Air MEDEVAC assets $m$	4	-
Classified risk levels $R$	2	-
Reward Ratio $\Gamma$	1.25	[1, 1.5]
Distance Ratio $\theta$	2	[2, 6]
Triage accuracy $\alpha$	$10^6$	[1, $10^6$ ]
Call arrival rate per hour $\lambda$	3.0	[1.5, 3.25]
Casualties in a batch $X$	1	-
Probability of casualty at each location $P_i$	[0.25 0.50 0.10 0.15]	-

All computations are performed on dual servers with Quad-Core 3.00 GHz processors and 16GB RAM. To solve the Markov decision process model [see 72] a value iteration convergence algorithm with tolerance of  $10^{-5}$  is used. The value iteration convergence algorithm is presented in section 2.4. The run time for the value iteration algorithm is approximately 250 minutes for the base case model with 83,521 states and 30 replications of a simulation for 10,000 casualties has a run time of approximately 18 minutes.

### 2.3.2 Policy Comparison

The optimal Markov decision process solution is compared to three heuristic policies: 1) transport all casualties to the most rewarding medical treatment facility 2) transport all casualties to the closest medical treatment facility and 3) transport low-priority casualties to the closest medical treatment facility and transport high-priority casualties to the most

rewarding medical treatment facility. All three heuristics dispatch the closest available server. Figure 2.3 compares the objective function value of the MDP to the performance of the three heuristics. The objective function values are rescaled so that values reflect the average modified Golden Hour utility received per casualty. Insights to be gained from figure 2.3 include the magnitude of improvement in system performance from leveraging optimization techniques versus heuristic policies. The optimal policy yields solution values that are, on average, 4.55%, 21.32%, and 0.72% better than heuristics 1, 2, and 3, respectively, as the distance ratio  $\theta$  increases. It is also of note that when the distance ratio is small, e.g.  $\theta = 2$ , the MDP only mildly outperforms heuristic 3 and heuristic 1. However, as the distance ratio  $\theta$  increases, the margin in which the MDP outperforms the best heuristic increases.

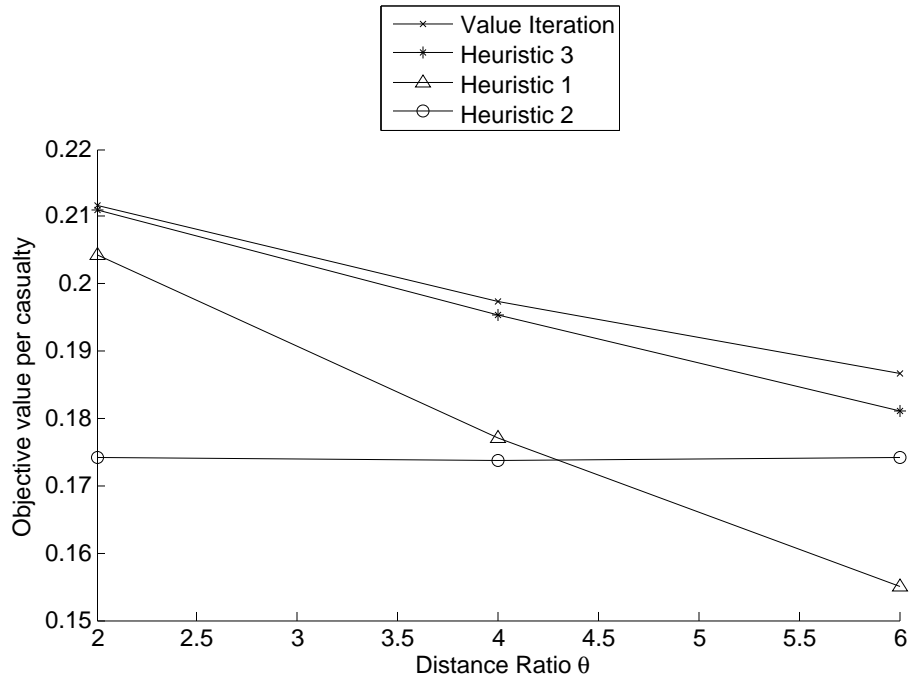


Figure 2.3: Simulated policy comparison

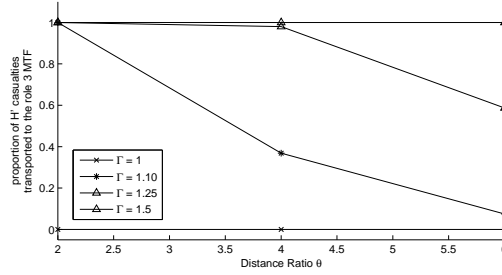
The loss rate is defined as the percentage of calls for service lost within the military aeromedical evacuation system, due to the system being overcrowded with no available

air MEDEVAC assets. For the MDP model, under the base case with  $\lambda = 3.0$  and  $\theta = 2$ , the loss rate is 12.97%. Likewise, when  $\theta = 6$  the loss rate is 16.38%. In practice, the so-called “lost” calls are delegated to non-traditional MEDEVAC assets so that all casualties receive timely service. Military aeromedical evacuation systems differ from their civilian counterparts in that every effort is made to keep the queue for service at zero [17].

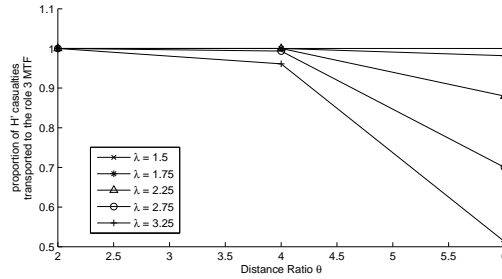
### 2.3.3 Dispatching and Transporting Sensitivity

The MDP model optimizes over both the dispatching and transporting decisions, and therefore it is of interest to know when to transport casualties to the different medical treatment facilities. Figures 2.4(a) - 2.4(c) illustrate the sensitivity of the proportion of high-priority casualties delivered to the role 3 medical treatment facility as a function of the distance ratio  $\theta$  for different  $\Gamma$ ,  $\lambda$ , and  $\alpha$  values. A main insight of military aeromedical evacuation systems, seen in figure 2.4(a), is the impact of the reward ratio  $\Gamma$  on the proportion of casualties transported to the more rewarding facility. Specifically, when the role 3 medical treatment facility is 50% better than the role 2 medical treatment facility ( $\Gamma = 1.5$ ), all high priority casualties are transported to the more rewarding role 3 medical treatment facility, regardless of whether the role 3 medical treatment facility is close or far. Figure 2.4(b) provides insights on how the call arrival rate  $\lambda$  effects the system. An increase in  $\lambda$  floods the system with more casualties and more high priority casualties are transported to the role 2 medical treatment facility that is closer, so that servers can end service and be available to respond. It is optimal to deliver almost all true high-priority casualties to the role 3 medical treatment facility unless the role 3 medical treatment facility is very distant and there are few marginal benefits or high call volume. In sum, these results suggest that a heuristic that transports truly high-priority casualties to the medical treatment facility with the higher utility (unless it is extremely distant) and truly low-priority casualties to the nearest medical treatment facility, as done by Heuristic 3, can be used by military decision makers as a near

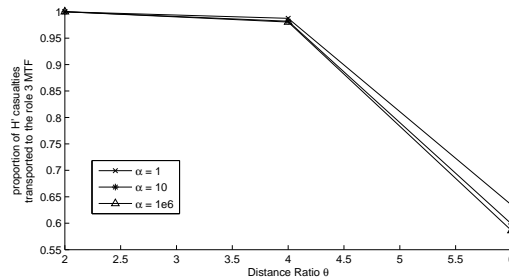
optimal transportation policy. A heuristic transport policy has the added benefit of greatly reducing the state space to  $(1 + nN)^m$  states, which helps to improve model scalability. However, it is less clear which asset to send upon initial dispatch. The dispatch decision is largely responsible for the difference in performance between the optimal policy and Heuristic 3 (see Figure 2.3), and therefore, we examine this issue next.



(a)  $\Gamma$  sensitivity



(b)  $\lambda$  sensitivity



(c)  $\alpha$  sensitivity

Figure 2.4: Sensitivity analysis on the proportion of true high-priority casualties  $H^l$  transported to the role 3 medical treatment facility, with respect to distance ratio  $\theta$

Table 2.4 presents the proportion of classified high-priority and low-priority calls to

whom the closest MEDEVAC asset is dispatched, which captures system insights on whether to send the closest server or ration it instead. We note that the closest MEDEVAC asset is not always available, so it is impossible for these values to be equal to 1.0. We examine this decision across different levels of triage accuracy from a worst-case lower bound  $\alpha = 1$  to  $\alpha = 100$ . Consider the classified  $H$  casualties in Table 2.4. The general insight we gain is that the frequency in which the closest server is dispatched decreases for locations 3 and 4, and increases for locations 1 and 2. The model is accounting for the call location distribution  $P_i$ , where location 1 and 2 have the greatest probability of a call for service, and reducing the response time to classified  $H$  calls for service in location 1 and 2. Reducing the response time by sending the closest server allows the servers to finish service and become available sooner for the additional calls expected in location 1 and 2.

In a similar manner, Table 2.4 also presents the closest MEDEVAC asset dispatching frequency for classified low-priority casualties. As  $\alpha$  increases, each MEDEVAC asset generally responds to fewer calls in its “home” location. Therefore, when information is more accurate (i.e., when  $\alpha$  is large), the system “saves” some MEDEVAC assets for responding to nearby truly high-priority casualties by strategically sending more distant MEDEVAC assets to low-priority casualties. This suggests that as classification accuracy improves, it is optimal to ration MEDEVAC assets in areas with the largest rate of truly high-priority calls.

We further study the impact of the initial dispatch decision on later transport decisions by examining whether the transport decisions depend on the MEDEVAC asset dispatched. Recall that the MEDEVAC asset dispatched to a call for service also transports the casualty. Table 2.5 shows the proportion of true high-priority casualties that are transported to role 3 medical treatment facility in two cases: when the closest MEDEVAC asset responds and when further MEDEVAC assets respond. Here, we see that when  $\theta = 2$  all responding MEDEVAC assets transport casualties to the role 3 medical treatment facility, both when



Table 2.4: Proportion of calls for service at each location that are responded to by the closest aeromedical evacuation asset

$\theta$	$\alpha$	Location $i$	classified $H$ calls	classified $L$ calls
$\theta = 2$	$\alpha = 1$	1	0.506	0.506
		2	0.447	0.447
		3	0.503	0.503
		4	0.555	0.555
	$\alpha = 10$	1	0.516	0.416
		2	0.535	0.071
		3	0.460	0.456
		4	0.484	0.480
	$\alpha = 100$	1	0.534	0.256
		2	0.535	0.069
		3	0.461	0.430
		4	0.452	0.461
$\theta = 6$	$\alpha = 1$	1	0.434	0.434
		2	0.388	0.389
		3	0.422	0.422
		4	0.478	0.478
	$\alpha = 10$	1	0.447	0.349
		2	0.463	0.077
		3	0.409	0.406
		4	0.431	0.440
	$\alpha = 100$	1	0.457	0.237
		2	0.464	0.076
		3	0.415	0.347
		4	0.415	0.411

Table 2.5: Proportion of true high-priority casualties transported to role III medical treatment facility based on the responding aeromedical evacuation asset

$\theta$	$\alpha$	$i$	Closest MEDEVAC asset responds	More distant MEDEVAC asset responds
2	$\alpha = 1, 10, \text{ and } 100$	1	1.000	1.000
		2	1.000	1.000
		3	1.000	1.000
		4	1.000	1.000
6	$\alpha = 1$	1	0.541	0.249
		2	1.000	0.780
		3	0.809	0.231
		4	0.480	0.037
	$\alpha = 10$	1	0.560	0.139
		2	1.000	0.761
		3	0.775	0.145
		4	0.429	0.042
	$\alpha = 100$	1	0.563	0.105
		2	1.000	0.756
		3	0.751	0.107
		4	0.317	0.053

dispatch classification is poor ( $\alpha = 1$ ) and when dispatch classification is better ( $\alpha = 100$ ). Also, when the role 3 medical treatment facility is further and  $\theta = 6$ , we see that more distant responding MEDEVAC assets are less likely to later transport truly high-priority casualties to role 3 medical treatment facilities. There is less incentive (in terms of the Golden Hour performance measure) to transport casualties to the role 3 medical treatment facility when more distant MEDEVAC assets respond to a casualty. Tables 2.4 and 2.5 together shed light on how the revelation of information affects decisions throughout the treatment and delivery of each casualty. In particular, when there are more initial classification errors, distant assets are more likely to respond to  $H'$  casualties, who are then less likely to be transported to medical treatment facilities with the best capabilities.

## 2.4 Value Iteration Convergence Algorithm

To solve for the optimal policy, the relative value function algorithm [see 72] is run using the finite-horizon value functions. Therefore,  $t$  is the iteration here, not time. To do so, define  $v_t(s(t))$  as the value of being in state  $s(t)$  during iteration  $t$ , for  $t = 0, \dots, N-1$ , and  $v_0(s(t)) = 0$  for all  $s(t) \in S$ .

$$v_{t+1}(s(t)) = \frac{1}{\gamma} \left[ \sum_{k=1}^d \sum_{j=1}^m \sum_{i=1}^n (\delta_{ijk})^{-1} I_{\{s_j=D_k\}} v_t(S^M(s(t)|s_j=0)) \right] \quad (2.2a)$$

$$+ \sum_{i=1}^n \sum_{X=1}^N \sum_{R \in \{H,L\}} \lambda P_i^X P_{R|i}^X \max_{j \in A_1(s(t))} \{v_t(S^M(s(t)|s_j=(i,R,X)))\} \quad (2.2b)$$

$$+ \sum_{i=1}^n \sum_{X=1}^N \sum_{j=1}^m \sum_{R \in \{H,L\}} \sum_{r \in \{H',L'\}} (\mu_{ij}^X)^{-1} I_{\{s_j=(i,R,X)\}} P_{r|R \cap i}^X \max_{D_k \in A_2(s(t))} \{v_t(S^M(s(t)|s_j=D_k)) + \gamma u_{ijk}^X\} \quad (2.2c)$$

$$+ \left( \gamma - \lambda - \sum_{i=1}^n \sum_{X=1}^N \sum_{j=1}^m \sum_{R \in \{H,L\}} (\mu_{ij}^X)^{-1} I_{\{s_j=(i,R,X)\}} - \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^d (\delta_{ijk})^{-1} I_{\{s_j=D_k\}} \right) v_t(s(t)) \right] \quad (2.2d)$$

To achieve the optimal policy, the relative value iteration algorithm is run until the upper and lower bounds converge to the optimal average utility per stage  $g$

$$L_t \leq L_{t+1} \leq g \leq U_{t+1} \leq U_t,$$

with lower bound

$$L_t = \min_{s(t) \in S} [v_{t+1}(s(t)) - v_t(s(t))]$$

and upper bound

$$U_t = \max_{s(t) \in S} [v_{t+1}(s(t)) - v_t(s(t))].$$

The value iteration algorithm is executed until  $U_{t+1} - L_{t+1} \leq \varepsilon$ , for a given  $\varepsilon$ .

## 2.5 Theoretical Results

We consider the finite stage optimality equations and consider the limit. The  $N$ -stage case MDP equations are in section 2.4. Note that in section 3, equation 2.1a - 2.1d capture the exact, infinite horizon, average cost optimality equations. In contrast, equation 2.2a - 2.2d capture the optimality equations for the finite-horizon case.

Next we exploit the finite case optimality equations to analyze the MDP structural properties. The first lemma shows that it is always optimal to choose to transport casualties to the more rewarding medical treatment facilities when two medical treatment facilities have the same expected transport time. This suggests that transporting to the closest facility is optimal.

**LEMMA 2.1.** Suppose a MEDEVAC asset  $j$  finishes service at location  $i$  and needs to transport a casualty with risk level  $r$  to one of two medical treatment facilities, labeled as 1 and 2,  $r \in \{H', L'\}$ . If both facilities have the same expected transport time, i.e.,  $\delta_{ij1} = \delta_{ij2}$  and facility 1 has a higher utility than facility 2, i.e.,  $u_{ij1r}^X \geq u_{ij2r}^X$ , then it is always better to deliver to the facility with the highest utility.

*Proof.* Without loss of generality, assume that the system is in state  $s(t)$  with value  $v_t(s(t))$ . The set of available transport decisions here are  $A_2(s(t)) = \{D_1, D_2\}$ , which correspond to facilities 1 and 2. Let  $s_1(t+1)$  and  $s_2(t+1)$  denote the states when facilities 1 and 2 are selected, respectively (see (2.1c)). It is sufficient to show that  $v_t(s_1(t+1)) + \gamma u_{ij1r}^X \geq v_t(s_2(t+1)) + \gamma u_{ij2r}^X$ . Note that in this case, the state in place  $j$  corresponding to asset  $j$  moves into the same transport state (i.e.,  $s_j = D$  whether medical treatment facility 1 or 2 is selected. Then, we can rearrange this to obtain

$$\gamma u_{ij1r}^X - \gamma u_{ij2r}^X \geq v_t(s_2(t+1)) - v_t(s_1(t+1)).$$

Since  $\delta_{ij1} = \delta_{ij2}$  for  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ , then the value functions in these two states entirely cancel, yielding  $\gamma u_{ij1r}^X - \gamma u_{ij2r}^X \geq 0$ , which is true since  $\gamma > 0$  and  $u_{ij1r}^X \geq u_{ij2r}^X$  for  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ ,  $r \in \{H', L'\}$ .  $\square$

The next proposition shows that the average utility per stage is higher in a state when a MEDEVAC is available as compared to when it is busy transporting a casualty.

**PROPOSITION 1.** Let state  $s(t)$  be a state where server  $j$  is available, i.e.,  $s_j(t) = 0$ . Let state  $\hat{s}(t)$  be the corresponding state where server  $j$  is transporting a casualty, i.e.,  $\hat{s}_j(t) = D_k$  for some  $k$  and  $\hat{s}_l(t) = s_l(t)$ ,  $l = 1, \dots, m$  and  $l \neq j$ . Then  $v_t(s(t)) - v_t(\hat{s}(t)) \geq 0$  for all  $t \geq 0$ .

*Proof.* The claim will be shown by induction. First, note that the claim is trivially true for  $t = 0$  since  $v_t(s(0)) = 0$  for all states  $s(0)$ . Let  $s(t)$  and  $s(t+1)$  denote identical states at different times. After some rearranging:

$$\begin{aligned} & \gamma(v_{t+1}(s(t+1)) - v_{t+1}(\hat{s}(t+1))) = \\ & \sum_{k=1}^d \sum_{j=1}^m \sum_{i=1}^n (\delta_{ijk})^{-1} I_{\{s_j=D_k\}} (v_t(S^M(s(t)|s_j=0)) - v_t(S^M(\hat{s}(t)|\hat{s}_j=0))) \\ & + \sum_{i=1}^n \sum_{X=1}^N \sum_{R \in \{H,L\}} \lambda P_i^X P_{R|i}^X \left( \max_{j \in A_1(s(t))} \{v_t(S^M(s(t)|s_j=(i,R,X)))\} - \max_{j \in A_1(\hat{s}(t))} \{v_t(S^M(\hat{s}(t)|\hat{s}_j=(i,R,X)))\} \right) \quad (2.3a) \\ & + \sum_{i=1}^n \sum_{X=1}^N \sum_{j=1}^m \sum_{R \in \{H,L\}} \sum_{r \in \{H',L'\}} \frac{P_{r|R \cap i}}{\mu_{ij}^X} I_{\{s_j=(i,R,X)\}} \left( \max_{D_k \in A_2(s(t))} \{v_t(S^M(s(t)|s_j=D_k)) + \gamma u_{ijkr}^X\} - \right. \\ & \quad \left. \max_{D_k \in A_2(\hat{s}(t))} \{v_t(S^M(\hat{s}(t)|\hat{s}_j=D_k)) + \gamma u_{ijkr}^X\} \right) \quad (2.3b) \\ & + \left( \gamma - \lambda - \sum_{i=1}^n \sum_{X=1}^N \sum_{j=1}^m \sum_{R \in \{H,L\}} (\mu_{ij}^X)^{-1} I_{\{s_j=(i,R,X)\}} - \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^d (\delta_{ijk})^{-1} I_{\{s_j=D_k\}} \right) (v_t(s(t)) - v_t(\hat{s}(t))). \end{aligned}$$

Note that in line (2.3a), the set of actions in  $A_1(s(t))$  is a subset of those in  $A_1(\hat{s}(t+1))$ . Let  $j^* = \arg \max_j \{A_1(\hat{s}(t+1))\}$ . We can bound the expression above from below by setting both decisions in (2.3a) to  $j^*$ . Likewise, we can apply this same idea to the actions in  $A_2(s(t))$  selected in both maximizations in (2.3b). Let  $d^* = \arg \max_{D_k} \{A_2(\hat{s}(t+1))\}$ , and set the destination in the first maximization to  $d^*$ . Then,

$$\begin{aligned} & \gamma(v_{t+1}(s(t+1)) - v_{t+1}(\hat{s}(t+1))) = \\ & \sum_{k=1}^d \sum_{j=1}^m \sum_{i=1}^n (\delta_{ijk})^{-1} I_{\{s_j=D_k\}} (v_t(S^M(s(t)|s_j=0)) - v_t(S^M(\hat{s}(t)|\hat{s}_j=0))) \end{aligned}$$

$$\begin{aligned}
& + \sum_{i=1}^n \sum_{X=1}^N \sum_{R \in \{H,L\}} \lambda P_i^X P_{R|i}^X (v_t(S^M(s(t)|s_{j^*} = (i,R,X))) - \{v_t(S^M(\hat{s}(t)|\hat{s}_{j^*} = (i,R,X)))\}) \\
& + \sum_{i=1}^n \sum_{X=1}^N \sum_{j=1}^m \sum_{R \in \{H,L\}} \sum_{r \in \{H',L'\}} \frac{P_{r|R \cap i}}{\mu_{ij}^X} I_{\{s_j = (i,R,X)\}} (v_t(S^M(s(t)|s_j = d^*)) - v_t(S^M(\hat{s}(t)|\hat{s}_j = d^*))) \\
& + \left( \gamma - \lambda - \sum_{i=1}^n \sum_{X=1}^N \sum_{j=1}^m \sum_{R \in \{H,L\}} (\mu_{ij}^X)^{-1} I_{\{s_j = (i,R,X)\}} - \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^d (\delta_{ijk})^{-1} I_{\{s_j = D_k\}} \right) (v_t(s(t)) - v_t(\hat{s}(t))).
\end{aligned}$$

Here, all four lines are non-negative using by the induction assumption. Therefore, the claim is true.  $\square$

The second lemma shows that it is always optimal to choose to transport casualties to the “closer” medical treatment facility if both facilities have the same utility, where a facility is “closer” to a casualty at location  $i$  with asset  $j$  if its expected transport time is smaller,

LEMMA 2.2. Suppose in state  $s(t)$  a MEDEVAC asset finishes service at the scene and needs to transport a casualty with risk level  $r$  to one of two medical treatment facilities, labeled as 1 and 2,  $r \in \{H', L'\}$ . If both facilities have the same utility, i.e.,  $u_{ij1r}^X = u_{ij2r}^X$  for  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ ,  $r \in \{H', L'\}$ , and the expected transport time is shorter for facility 1, i.e.,  $\delta_{ij1} < \delta_{ij2}$  for  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ , then  $v_t(S^M(s(t)|s_j = D_1)) - v_t(S^M(s(t)|s_j = D_2)) \geq 0$  for all  $t \geq 0$  and it is always better to deliver to the facility with the smaller expected service time.

*Proof.* The claim will be shown by induction. First, note that the claim is trivially true for  $t = 0$  since  $v_t(s(0)) = 0$  for all states  $s(0)$ . We assume that  $v_t(s1(t)) - v_t(s2(t)) \geq 0$  for all states  $s1$  and  $s2$  that are identical. Let MEDEVAC asset  $j^*$  be the asset that must transport casualties to a medical treatment facility. Let the state  $s1(t+1) = S^M(s(t)|s_j = D_1)$  and let  $s2(t+1) = S^M(s(t)|s_j = D_2)$ . Next, after some rearranging:

$$\begin{aligned}
& \gamma(v_t(s1(t+1)) - v_t(s2(t+1))) = \\
& \sum_{k=1}^d \sum_{j=1, j \neq j^*}^m \sum_{i=1}^n (\delta_{ijk})^{-1} I_{\{s_j = D_k\}} (v_t(S^M(s1(t)|s1_j = 0)) - v_t(S^M(s2(t)|s2_j = 0))) \\
& + \sum_{i=1}^n \sum_{X=1}^N \sum_{R \in \{H,L\}} \lambda P_i^X P_{R|i}^X \left( \max_{j \in A_1(s1(t))} \{v_t(S^M(s1(t)|s1_j = (i,R,X)))\} - \max_{j \in A_1(s2(t))} \{v_t(S^M(s2(t)|s2_j = (i,R,X)))\} \right) \quad (2.4a)
\end{aligned}$$

$$\begin{aligned}
& + \sum_{i=1}^n \sum_{X=1}^N \sum_{j=1}^m \sum_{R \in \{H,L\}} \sum_{r \in \{H',L'\}} \frac{P_{r|R \cap i}}{\mu_{ij}^X} I_{\{s1_j=(i,R,X)\}} \left( \max_{D_k \in A_2(s1(t))} \{v_t(S^M(s1(t)|s_j=D_k)) + \gamma u_{ijk}^X\} - \right. \\
& \quad \left. \max_{D_k \in A_2(s2(t))} \{v_t(S^M(s2(t)|s2_j=D_k)) + \gamma u_{ijk}^X\} \right) \\
& + \left( \gamma - \lambda - \sum_{i=1}^n \sum_{X=1}^N \sum_{j=1, j \neq j^*}^m \sum_{R \in \{H,L\}} (\mu_{ij}^X)^{-1} I_{\{s_j=(i,R,X)\}} - \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^d (\delta_{ijk})^{-1} I_{\{s_j=D_k\}} \right) (v_t(s1(t)) - v_t(s2(t))) \\
& + ((\delta_{ij1})^{-1} - (\delta_{ij2})^{-1}) v(s(t+1)) - ((\delta_{ij1})^{-1} v(s1(t)) + (\delta_{ij2})^{-1} v(s2(t))).
\end{aligned} \tag{2.4b}$$

As in Proposition 1, let  $j^* = \arg \max_j \{A_1(s2(t))\}$ . We can bound the expression above from below by setting both decisions in (2.4a) to  $j^*$ . Likewise, we can apply this same idea to the actions selected in both maximizations in (2.4b). Let  $d^* = \arg \max_{D_k} \{A_2(s2(t))\}$ , and set the destination in the first maximization to  $d^*$ . Moreover, the last line can be rearranged to yield:

$$(\delta_{ij1})^{-1} (v(s0(t)) - v(s1(t))) - (\delta_{ij2})^{-1} (v(s0(t)) - v(s2(t)))$$

after noting that  $s0(t) = S^M(s1(t)|s1_j=0) = S^M(s2(t)|s2_j=0)$ . Moreover, we can bound this below by applying the induction assumption, with

$$(\delta_{ij1})^{-1} (v(s0(t)) - v(s1(t))) - (\delta_{ij2})^{-1} (v(s0(t)) - v(s2(t))) \geq ((\delta_{ij1})^{-1} - (\delta_{ij2})^{-1}) (v(s0(t)) - v(s1(t))).$$

This yields:

$$\begin{aligned}
& \gamma(v_t(s1(t+1)) - v_t(s2(t+1))) \geq \\
& \sum_{k=1}^d \sum_{j=1, j \neq j^*}^m \sum_{i=1}^n (\delta_{ijk})^{-1} I_{\{s_j=D_k\}} (v_t(S^M(s1(t)|s1_j=0)) - v_t(S^M(s2(t)|s2_j=0))) \\
& + \sum_{i=1}^n \sum_{X=1}^N \sum_{R \in \{H,L\}} \lambda P_{R|i}^X P_{R|i}^X (v_t(S^M(s1(t)|s1_{j^*}=(i,R,X))) - v_t(S^M(s2(t)|s2_{j^*}=(i,R,X)))) \\
& + \sum_{i=1}^n \sum_{X=1}^N \sum_{j=1}^m \sum_{R \in \{H,L\}} \sum_{r \in \{H',L'\}} \frac{P_{r|R \cap i}}{\mu_{ij}^X} I_{\{s1_j=(i,R,X)\}} (v_t(S^M(s1(t)|s_j=d^*)) - v_t(S^M(s2(t)|s2_j=d^*))) \\
& + \left( \gamma - \lambda - \sum_{i=1}^n \sum_{X=1}^N \sum_{j=1, j \neq j^*}^m \sum_{R \in \{H,L\}} (\mu_{ij}^X)^{-1} I_{\{s_j=(i,R,X)\}} - \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^d (\delta_{ijk})^{-1} I_{\{s_j=D_k\}} \right) (v_t(s1(t)) - v_t(s2(t))) \\
& + ((\delta_{ij1})^{-1} - (\delta_{ij2})^{-1}) (v(s0(t)) - v(s1(t))).
\end{aligned}$$

The first four lines are greater than or equal to zero by the induction assumption. The last line is greater than or equal to zero by Proposition 1 and by noting that  $(\delta_{ij1})^{-1} - (\delta_{ij2})^{-1} \geq$

## 2.6 Conclusions

This chapter models and analyzes optimal dispatching and transporting policies in military aeromedical evacuation systems. Timely transportation of casualties motivates the need to examine how to make better interrelated decisions—how to dispatch MEDEVAC assets to casualties and then transport casualties to medical treatment facilities—given the revelation of information over the duration of each call. An undiscounted, infinite horizon, average-cost MDP model is formulated to identify optimal policies, which is solved using a value iteration algorithm. In the computational example, a situation where two medical treatment facilities are distinguishable by both their proximity to calls for service (distance) and treatment capability (reward) is considered. Each dispatching and transporting decision affects system resources being busy or available to respond to additional calls for service. Optimal decision policies utilize the better role 3 medical treatment facility with varying frequency, as system input parameters such as call volume and dispatcher classification ability are varied. The optimal policy outperforms three heuristics considered in this chapter on average by 4.55%, 21.32%, and 0.72%, respectively. The initial dispatch decisions account for much of the improvement over the heuristic policies. The computational results suggest that for most settings, a heuristic policy could be used for the transport decisions, which would greatly reduce the state space and improve model scalability.



## Locating and Dispatching Two Classes of Military Aeromedical Evacuation Assets

### 3.1 Introduction

Military medical evacuation (MEDEVAC) systems save lives by responding to casualty incidents and transporting the most urgent casualties to a medical treatment facility (MTF) via a fleet of distinguishable, geographically-dispersed air ambulance assets (servers) dedicated to the response of casualties. It is important to know where to locate distinct classes of air assets and how to dispatch them in order to maximize the likelihood of survival of the most urgent casualties. To do so, it becomes necessary to triage casualties so that soldiers with the most severe, life-threatening injuries – where death can occur if response is not immediate – receive the most timely treatment. This chapter examines the following hard problems found in military MEDEVAC systems: 1) how to geographically locate two classes of air assets and 2) how to construct response districts for all air assets in the system. The problems mentioned above are hard because it is not obvious where to locate scarce air assets or when to dispatch further air assets (and ration closer ones) to casualty incidents in response time-dependent military medical systems.

Optimal decision-making within MEDEVAC systems is complicated due to multiple types of casualties and distinguishable air assets. The ranking of evacuation precedence is handled through a straightforward casualty triage scheme:

- CAT A: alpha category includes urgent and urgent-surgical casualties that need to be treated within one hour.

- CAT B: bravo category includes priority casualties that need to be treated within four hours.
- CAT C: charlie category includes routine casualties that need to be treated within 24 hours.

The fleet includes two classes of air assets to manage:

- U.S. Army helicopters such as the UH-60A/L Black Hawk or HH-60 MEDEVAC platforms are normally leveraged to respond to casualties based on their speed, range, and en route medical capability.
- U.S. Air Force helicopters such as the HH-60G Pave Hawk are also available to respond to casualties but are slower because they carry heavy weapon systems and have a larger flight crew.

When effectively employed together, both Army and Air Force air MEDEVAC assets can provide a more efficient and responsive military MEDEVAC system for overseas combat and stability operations.

We propose a novel binary linear programming (BLP) model that optimally locates two classes of air assets within a military medical system deployed overseas and assigns these assets to casualty locations using a dispatch preference list. A dispatch preference list is an ordered ranking of air assets for each casualty location. Additionally, the BLP model balances the workload among assets and enforces contiguity amongst the first assigned location for each air asset. The objective of the BLP model is to maximize the proportion of high-priority casualties (CAT A) responded to within a given response time threshold. The BLP model incorporates coverage thresholds for low-priority casualties to encourage prompt service for all casualties. Accounting for the queuing dynamics in military medical systems results in a more realistic optimization model in which air assets are not always available

for service. Therefore, we also propose a queuing model to derive performance statistics, including busy probabilities and independence correction factors, for two classes of air assets. Realistic busy probabilities and removing the assumption of asset independence leads to more effective resource allocation decision-making for military medical planners. The main contribution of this chapter is a BLP model that constructs a complete set of MEDEVAC response districts (modeled as dispatch preference lists) of air assets to casualties that explicitly account for two types of air assets and multiple casualty types.

This chapter is organized as follows. Section 3.2 provides a literature review of service system location models and the probabilistic nature of military MEDEVAC systems. Section 3.3 introduces the approximate Hypercube algorithm [52, 47] used to compute queuing system performance statistics. Section 3.4 introduces the BLP model to locate air assets and construct dispatch preference lists. We present computational results for a military medical system in section 3.5. Conclusions and future work are presented in Section 3.6.

### 3.2 Literature Review

A number of existing research papers focus on air MEDEVAC asset optimization models for military medical systems. Bastian [6] presents a multi-criteria decision analysis (MCDA) model using stochastic, mixed-integer goal programming to determine the minimum number of air assets needed at each MTF. This in turn maximizes the coverage of the theater-wide casualty demand and the probability of meeting that demand. Simultaneously, Bastian's model also minimizes the maximal MTF evacuation site total vulnerability to enemy attack. Fulton et al. [34] introduce a two-stage stochastic optimization modeling framework for the medical evacuation (ground and air) of casualties, which identifies optimal casualty evacuation sites and MTF sites in response to stochastic demands for service. While Bastian's and Fulton's papers also account for the stochastic nature of military medical

systems, this paper is distinct because it leverages queuing theory to derive busy probabilities to more accurately model the stochastic nature and availability of air assets. Bastian and Fulton [7] present a geospatial-based decision-support tool that identifies which air asset to launch in response to a casualty event given knowledge of terrain, aircraft location, and aircraft capabilities. Zeto et al. [87] examine the pre-location of air assets, along with type and quantity, to maximize the theater-wide coverage while balancing air asset reliability. Bastian and Fulton [7] and Zeto et al. [87] examine location and dispatch problems, but differ from the model presented in this paper because they do not construct a dispatch preference list and complete ordering of all air assets for each casualty location. Bastian et al. [8, 9] examine medical evacuation platforms of the Future of Vertical Lift (FVL) program in support of brigade operations. The specific operations analyzed include the zero-risk aircraft ground speed; different aircraft engine trade-off considerations; and the effects of weaponizing the current air asset fleet will have on the fleet's range, coverage radius, and response time. In contrast to the evaluation of future platforms, this paper presents a model that can be used at the tactical or operational level to decide where to locate air assets and to identify how to dispatch air assets to casualty incidents. Bouma [16] develops the Medical Evacuation and Treatment Capabilities Optimization Model (METCOM) that considers policy effects on key measures of effectiveness, and then optimizes treatment and facility capacities for given casualty flows. This paper does not consider the impact of policy on military medical systems or examine the impact of casualty flows on facility capacities. Fulton et al. [35] present a Monte Carlo simulation to evaluate rules of allocation and planning considerations for Army air ambulance companies during major combat operations (MCO). Rather than incorporating a simulation model, this paper uses queuing theory and binary linear programming to model and examine military medical systems.

There has been much research on civilian emergency medical service (EMS) systems which share similar characteristics as military medical systems. There are many ambulance

location models that focus on “covering” patient demand based on response time thresholds that are (usually) nine minutes from dispatch. Many of the models are extensions and variations on the seminal paper by Church and ReVelle [23] that introduces the Maximal Covering Location Problem (MCLP) as a basic facility location model for EMS systems. An extension to MCLP is the Maximal Expected Covering Location Problem (MEXCLP) developed by Daskin [29], which attempts to maximize the expected number of calls covered in a given amount of time, and makes the following assumptions: servers operate independently, each server has the same busy probability, and busy probabilities do not depend on server location. The model in this paper lifts Daskin’s last two assumptions. "Assets of different types may be used in ways that result in large differences in the busy probabilities between the asset types. Likewise, there are often “hot-beds” of activity within military systems (e.g., near an enemy stronghold) that would naturally result in some assets to have larger busy probabilities due to their frequent selection as the responding assets. This paper addresses both of these issues." Batta et al. [10] further extend MEXCLP by introducing the Adjusted MEXCLP (AMEXCLP) to lift the three previously mentioned assumptions of MEXCLP by use of Hypercube correction factors (see below) for one type of server and a single casualty type. Our model extends the work by Batta et al. [10] by examining more than one type of server, multiple casualty types, and probabilistic travel times. The papers thus far do not consider distinguishable types of ambulances as we do in this paper.

ReVelle and Marianov [77] and Marianov and ReVelle [58] examine the location of multiple types of servers—fire engines and fire trucks—to maximize the number of calls requiring both fire engines and fire trucks. ReVelle and Marianov’s work assumes that one server of both types is required for each call for service, where this model dispatches one server to each call for service. Mandell [56] also considers multiple types of medical units but does not consider multiple casualty types. Marianov and Serra [60] introduce a

model that maximizes the demand covered when the customer does not have to wait in line (due to congestion) with more than a prespecified number of other customers. In contrast to Marianov and Serra's work, there is a zero length queue in this model. McLay [61] introduces a model that locates two types of ambulances while responding to multiple types of customers but does not assign response zones to ambulances. This paper extends the work of McLay [61] to construct dispatch preference lists modeled as contiguous response zones while also balancing asset workloads.

Spatial queuing models have been widely used to describe the underlying dynamics of military medical and civilian EMS systems. The exact and approximate Hypercube models by Larson [51, 52] are the most well-known spatial queuing models. The Hypercube model assumes a multi-server queuing system with indistinguishable servers. Jarvis [47] extends the Hypercube queuing model with an approximation algorithm to incorporate dependencies on casualty call type for service times with servers that are distinguishable by location. Chelst and Barlach [21], Larson and McKnew [53], and Budge et al. [18] consider other extensions to the Hypercube model. Several papers use Hypercube model outputs as BLP model inputs (see Batta et al. [10], McLay [61], and Ansari et al [3]).

This work adds to the literature by introducing a new BLP model that optimally locates two classes of air assets and assigns assets to casualty locations in a preference list. This work also includes an extension to the approximate Hypercube queuing model that accounts for multiple casualty types to derive system statistics that are used as input to the BLP. The new BLP model in this work better captures the military medical evacuation problem studied than previous models. For example, recently military medical planners and senior decision-makers must manage two distinct classes of air assets (both U.S. Army and U.S. Air Force) to respond to triaged casualty events. Further, contiguous response zones are a realistic detail of this model that reflect the practice of a determining dedicated, compact area of responsibility for helicopter pilots. Accounting for the queuing dynamics in military

medical systems results in a more realistic optimization model, in which air assets are not always available for service.

### 3.3 Queuing Model

In this section, we introduce a queuing model that computes system performance statistics to reflect the complex dynamics of military medical systems. Air assets are not always available for service and, therefore, computing busy probabilities allows for more realistic air asset management in military systems. Servers do not act independently of one another in service systems; thus, the queuing model in this section also develops independence correction factors to account for the assumption of server independence. A list of the symbols used in this chapter is given next, followed by a discussion of the system dynamics and assumptions of the queuing model.

$s_A$  ( $s_B$ ) – Number of type A (type B) air assets in the system with  $s_A + s_B = s$ ,

$I_A$  ( $I_B$ ) – Set of all potential type A (type B) MTF locations,

$J$  – Set of all casualty locations,

$C$  – Set of casualty types that can arrive partitioned into four classes  $C_A, C_{AB}, C_{BA}, C_B$ ,

$C_A$  ( $C_B$ )  $\subseteq C$  – Set of casualties that require type A (type B) air asset evacuation,

$C_{AB}$  ( $C_{BA}$ )  $\subseteq C$  – Set of casualties that prefer type A air asset evacuation over type B air evacuation (prefer type B over type A),

$\lambda_c$  – Call arrival rate of casualty type  $c$ , for  $c \in C$ , with  $\lambda_c = \sum_{j \in J} \lambda_{jc}$ , and  $\lambda_{jc}$  is the call arrival rate of a casualty at location  $j$  of type  $c$ , for  $j \in J, c \in C$ ,

$\lambda$  – System-wide total call arrival rate with  $\lambda = \sum_{j \in J, c \in C} \lambda_{jc}$ ,

$\tau_{ijc}^A$  ( $\tau_{ijc}^B$ ) – Service time when type A (type B) air asset location  $i$  responds to casualty of type  $c$  at location  $j$ , for  $c \in C$ ,  $j \in J$ ,  $i \in I_A(I_B)$ ,

$\tau^A$  ( $\tau^B$ ) – System-wide average type A (type B) air asset service time,

$\theta_c$  – Minimum performance benchmark for the fraction of casualty type  $c$  calls that must be responded to in a prespecified time threshold, for  $c \in C$ ,

$\rho^A$  ( $\rho^B$ ) – Traffic intensity for type A (type B) air assets,

$r_i^A$  ( $r_i^B$ ) – Type A (Type B) air asset location  $i$  busy probability, for  $i \in I_A(I_B)$ ,

$r^A$  ( $r^B$ ) – System-wide average type A (type B) air asset busy probability,

$a_{jck}$  – Ordered dispatch preference list where  $a_{jck}$  is the  $k^{th}$  preferred type A air asset for casualty of type  $c$  at location  $j$ , for  $j \in J$ ,  $c \in C_A \cup C_{AB}$ ,  $k = 1, \dots, s_A$  or for  $c \in C_A \cup C_{BA}$ ,  $k = s_B + 1, \dots, s_B + s_A$ ,

$f_{ijck}^A$  ( $f_{ijck}^B$ ) – The dispatch probability of type A air asset location  $i \in I_A$  (type B air asset  $i \in I_B$ ) to casualty  $j \in J$  of type  $c \in C$  as the  $k^{th} = 1, \dots, s_A$  ( $s_B$ ) priority,

$P_k^A$  ( $P_k^B$ ) – Loss probability that  $k$  type A (type B) air assets are busy,

$R_{ijc}^A$  ( $R_{ijc}^B$ ) – Fraction of time type A (type B) air asset location  $i$  can respond to casualties at location  $j$  of type  $c$  in a prespecified time threshold, for  $i \in I_A(I_B)$ ,  $j \in J$ ,  $c \in C$ ,

$Q^A(s_A, \rho^A, k)$  or  $Q^B(s_B, \rho^B, k)$  – Hypercube correction factor as a function of the number of air assets  $s_A$  or  $s_B$ , traffic intensity  $\rho^A$  or  $\rho^B$ , and priority  $k$ , for  $k = 0, \dots, s_A - 1$  or  $s_B - 1$ ,

$N_j$  – A neighbor incident matrix used for contiguity, i.e., the set of casualty locations that are adjacent to casualty location  $j$  in the usual sense, for  $j \in J$ .



Our queuing model is an extension to the approximate Hypercube queuing model given in Jarvis [47] that preserves the location of assets. The extension considers two types of air assets—type A and type B—and multiple casualty types in a military MEDEVAC system. Consider a system with  $s_A$  type A air assets,  $s_B$  type B air assets,  $J$  casualty locations, and  $C$  casualty types. Classification by triage leads to four mutually exclusive casualty types: (1) casualties that require type A air asset response denoted  $C_A$ , (2) casualties that prefer type A air asset response over type B air asset response denoted  $C_{AB}$ , (3) casualties that prefer type B air asset response over type A air asset response denoted  $C_{BA}$ , and (4) casualties that require type B air asset response denoted  $C_B$ . Note that  $C = C_A \cup C_{AB} \cup C_{BA} \cup C_B$ . Casualties at location  $j \in J$  of type  $c \in C$  arrive according to a Poisson process with rate  $\lambda_{jc}$  independent of other casualty locations or types. The total arrival rate of casualties in the system is  $\lambda = \sum_{j \in J} \sum_{c \in C} \lambda_{jc}$ . Exactly one air asset is assigned to each casualty for evacuation unless all air assets are busy, in which case a non-traditional evacuation air asset will be called upon for assistance. In military medical systems there is a zero length queue for service. Note that in this section we develop formulas for the type A air assets. The equations for the type B air assets are omitted but computed analogously.

In military medical systems, air assets do not act independently from one another and thus an optimization model to locate and dispatch air assets should account for asset dependencies. As in Jarvis [47], the type A air asset independence correction factors  $Q_A(s_A, \rho^A, j)$  quantify the correction to the probability of obtaining  $j$  busy type A air assets followed by an available type A air asset when assuming that air assets operate independently. Note that  $P_{s_A}^A$  is the probability that all  $s_A$  type A air assets are busy.

$$Q^A(s_A, \rho^A, k) = \sum_{j=k}^{s_A-1} \frac{(s_A - k - 1)!(s_A - j)(s_A)^j(\rho^A)^{j-k}P_0^A}{(j - k)!s_A!(1 - P_{s_A}^A)^k(1 - \rho^A(1 - P_{s_A}^A))}, \text{ for } k = 0, 1, \dots, s_A - 1. \quad (3.1)$$

It is desirable to know the probability of assigning specific air assets to specific customers. Denote the dispatch probability of type A air asset location  $i \in I_A$  to casualty  $j \in J$  of type  $c \in C$  as the  $k^{th} = 1, \dots, s_A$  priority as  $f_{ijck}^A$ . For type A air assets,  $f_{ijck}^A$  is computed differently for casualty types  $C_A$  and  $C_{AB}$  than casualty type  $C_{BA}$  since type A air assets can only be dispatched to casualties of type  $C_{BA}$  if all type B air assets are busy in service. The remaining system statistics depend on  $f_{ijck}^A$ . Assume there is a known dispatch preference list, and let  $a_{jck}$  equal the air asset that is  $k^{th}$  preferred to respond to a casualty of type  $c$  at location  $j$ . For example if  $a_{1A2} = 3$ , then air asset 3 is the second preferred air asset to respond to a casualty of type A at location 1. In the computation of  $f_{ijck}^A$ ,  $r_i^A$  is the busy probability for the type A air asset location  $i \in I_A$ . Note that  $f_{ijck}^A = 0$  for  $c \in C_B$  since type B air assets are required to respond to casualties of type  $C_B$ .

$$f_{ijck}^A \approx Q^A(s_A, \rho^A, k-1)(1-r_i^A) \prod_{l=1}^{k-1} r_{a_{jcl}}^A, \text{ for } c \in C_A \cup C_{AB}, j \in J, k = 1, \dots, s_A, i \in I_A \quad (3.2)$$

$$f_{ijck}^A \approx P_{s_B}^B Q^A(s_A, \rho^A, k-1)(1-r_i^A) \prod_{l=s_B+1}^{k-1} r_{a_{jcl}}^A, \text{ for } c \in C_{BA}, j \in J, k = s_B+1, \dots, s_B+s_A, i \in I_A \quad (3.3)$$

The dispatch probabilities computed in equations (2) and (3) depend on three values: 1) correction of the assumption of independence among air assets  $Q^A(s_A, \rho^A, k-1)$ , 2) the probability that the specific type A air asset is available  $(1-r_i^A)$ , and 3) the probability that the  $k-1$  more preferred air assets are busy  $\prod_{l=1}^{k-1} r_{a_{jcl}}^A$ . Note that  $a_{jcl} = i$  is a dispatch preference list, so the formula for  $f_{ijck}^A$  is multiplying together the busy probabilities of the  $k-1$  more preferred air assets. Lastly, equation (3) includes the  $P_{s_B}^B$  term because a type A air asset is only assigned to a  $C_{BA}$  type casualty once all  $s_B$  type B air assets are busy.

The computation of the type A air asset busy probabilities  $r_i^A$  is given by equation (3.4).

$$r_i^A = \frac{V_i^A}{(1 + V_i^A)}, i \in I_A \quad (3.4)$$

In equation (3.4),  $V_i^A$  is the rate at which type A air asset  $i$  is assigned to a call for service and is computed by equation (3.5):

$$V_i^A = \sum_{k=1}^{s_A} \left( \sum_{c \in C_A \cup C_{AB}} \sum_{j: a_{jck}=i} \lambda_{jc} \tau_{ijc}^A Q^A(s_A, \rho^A, k-1) \prod_{l=1}^{k-1} r_{a_{jcl}}^A \right) + P_{s_B}^B \sum_{k=s_B+1}^{s_B+s_A} \left( \sum_{c \in C_{BA}} \sum_{j: a_{jck}=i} \lambda_{jc} \tau_{ijc}^A Q^A(s_A, \rho^A, k-1) \prod_{l=s_B+1}^{k-1} r_{a_{jcl}}^A \right), \quad (3.5)$$

for  $i \in I_A$ . Equation (3.5) is comprised of two pieces separated by addition. The first piece captures the rate at which type A air asset  $i$  is assigned to casualty types  $C_A$  and  $C_{AB}$ . Type A air asset  $i$  is assigned to casualties only when the more preferred air assets are busy, captured by  $\prod_{l=1}^{k-1} r_{a_{jcl}}^A$ . The first summation shows that type A air asset  $i$  can be any of the first  $1..s_A$  preferred air assets. The third summation is over the casualty location  $j$  only if the dispatch preference list maps type A air asset  $i$  to casualty location  $j$  as the  $k^{th}$  preferred asset. The second piece of equation (3.5) captures the rate at which type A air asset  $i$  is assigned to casualty types  $C_{BA}$ . Note that the preference index  $k$  start at  $s_B + 1$  because type A air assets cannot be assigned to the first  $s_B$  spots of the dispatch list for casualty type  $C_{BA}$ . Also,  $P_{s_B}^B$  is included in this second piece because a type A air asset is only assigned to a  $C_{BA}$  type casualty once all  $s_B$  type B air assets are busy.

The average service time for type A air assets  $\tau^A$  is given by equation 3.6.

$$\tau^A = \sum_{j \in J} \sum_{c \in C} (\lambda_{jc}/\lambda) \sum_{i=1}^{s_A} \sum_{k=1}^s \tau_{ijc}^A f_{ijck}^A / (1 - P_{s_A}^A) \quad (3.6)$$

In equation (3.6), The average service time is a product of the service times  $\tau_{ijc}^A$  and the normalized dispatch probabilities  $f_{ijck}^A / (1 - P_{s_A}^A)$ . The average service time is weighted by the call arrival rate to casualties  $(\lambda_{jc}/\lambda)$ .

The approximate algorithm is run until the maximum change in air asset busy probabilities falls below a pre-specified  $\varepsilon$ . The approximate algorithm contains an initialization step where the closest air asset is assumed always available.

**Given:**

1. Call arrival rates  $\lambda_{jc}$ , for  $j \in J, c \in C$
2. Type A air asset service times  $\tau_{ijc}^A$ , for  $i \in I_A, j \in J, c \in C_A \cup C_{AB} \cup C_{BA}$

**Initialize:**

$$r_i^A = \sum_c \sum_{j: i \text{ is the closest}} \lambda_{jc} \tau_{ijc}^A, \quad i \in I_A \quad (3.7)$$

$$\tau^A = \sum_{j \in J} \sum_{c \in C} \sum_{k=1}^s (\lambda_{jc} / \lambda) \tau_{ijc}^A \quad (3.8)$$

**Iterate:**

*Step 1.* Compute  $Q^A(s_A, \rho^A, k)$  for  $k = 0, 1, \dots, s_A - 1$  using equation (3.1). Use  $\rho^A = \lambda \tau^A / s_A$ .

*Step 2.* For  $i = 1, \dots, s_A$ , compute  $r_i^A$  with equation (3.4), where  $V_i^A$  is given by equation (3.5).

*Step 3.* Stop if maximum change in  $r_i^A$  is less than a tolerance.

*Step 4.* Else, compute  $\tau^A$  by equation (3.6), where  $f_{ijck}^A$  is given by equations (3.2) and (3.3).

*Step 5.* Return to Step 1.

To compute the type B correction factors and average air asset busy probabilities also used as inputs to the BLP model, the analogous type B approximate algorithm is run with the corresponding type B air asset equations. The following differences are necessary to derive the type B equations. Type B air assets are preferred first when responding to casualties of type  $c \in C_B \cup C_{BA}$  as priority  $k = 1, \dots, s_B$ . The type B air assets responding to casualties of type  $c \in C_{AB}$  are preferred after type A air assets in the dispatch list as priority

$k = s_A + 1, \dots, s_A + s_B$ . The type A air asset independence correction factors  $Q^A(s_A, \rho^A, k)$  and average air asset busy probabilities  $r^A = \sum_{i \in I_A} r_i^A / s_A$  are used as input to the BLP model formulated in the next section. Likewise, the type B air asset independence correction factors  $Q^B(s_B, \rho^B, k)$  and average air asset busy probabilities  $r^B = \sum_{i \in I_B} r_i^B / s_B$  are also used as input to the BLP model. The next section introduces an optimization model that utilizes the queuing correction factors and dispatch probabilities.

The outputs of the spatial queuing approximation can be used to compute several inputs to the BLP model derived in the next section. The coefficients  $h_{ijk}^A$  and  $h_{ijk}^B$  found in the objective function and constraints of the BLP in the next section represent the fraction of casualties of type  $c$  at location  $j$  that air asset  $i$  can reach in a prespecified time threshold as the  $k$ th preferred asset to send, weighed by the demand at location  $j$  (captured by the  $\lambda_{jc}/\lambda_c$  term). These coefficients utilize dispatch probabilities (see (3.2 and 3.3) when assuming that busy probabilities of assets of the same type are equal. The coefficients for type A assets  $h_{ijk}^A$  are as follows:

$$h_{ijk}^A = (\lambda_{jc}/\lambda_c) R_{ijc}^A \left[ Q^A(s_A, \rho^A, k) (1 - r^A) (r^A)^{k-1} \right], i \in I_A, j \in J, c \in C_A \cup C_{AB}, k = 1, \dots, s_A \quad (3.9)$$

$$h_{ijk}^A = (\lambda_{jc}/\lambda_c) R_{ijc}^A P_{s_B}^B \left[ Q^A(s_A, \rho^A, k) (1 - r^A) (r^A)^{k-1} \right], i \in I_A, j \in J, c \in C_{BA}, k = s_B + 1, \dots, s_B + s_A \quad (3.10)$$

These coefficients take two sources of uncertainty into account: (1) uncertain asset availability reflected in the dispatch probabilities and correction factors and (2) uncertain travel times that lead to imperfect coverage (reflected by the terms  $R_{ijc}^A$  that represent the fraction of type  $c$  calls that can be reached in the fixed response time threshold).

In a similar manner, the coefficients  $g_{ijk}^A$  and  $g_{ijk}^B$  represent the asset workloads for type A and type B air assets, respectively, and the coefficients for type A assets  $g_{ijk}^A$  are defined in the following

equations:

$$g_{ijk}^A = (\lambda_{jc}/\lambda_c) \tau_{ijc}^A \left[ Q^A(s_A, \rho^A, k)(1 - r^A) (r^A)^{k-1} \right], i \in I_A, j \in J, c \in C_A \cup C_{AB}, k = 1, \dots, s_A \quad (3.11)$$

$$g_{ijk}^A = (\lambda_{jc}/\lambda_c) \tau_{ijc}^A P_{s_B}^B \left[ Q^A(s_A, \rho^A, k)(1 - r^A) (r^A)^{k-1} \right], i \in I_A, j \in J, c \in C_{BA}, k = s_B + 1, \dots, s_A + s_B \quad (3.12)$$

where  $\tau_{ijc}^A$  and  $\tau_{ijc}^B$  are the service times associated with type A and type B assets, respectively. These coefficients are used to balance the workload among the servers of the same type, and therefore, they also reflect the queuing dynamics.

### 3.4 Binary Linear Programming Model

We next present the BLP model used to locate two classes of air MEDEVAC assets and construct dispatch preference lists while balancing the workload among assets and enforcing contiguity amongst the first assigned location for each air asset. The objective of the BLP model is to maximize the proportion of high-priority casualties ( $C_A$  calls) responded to within a pre-specified time threshold. The BLP model also incorporates coverage thresholds for low-priority casualties to provide prompt service for all casualties.

To describe the BLP model we first introduce the decision variables. The first two sets of binary variables  $y_i^A$  and  $y_i^B$  indicate the set of MTF locations utilized:  $y_i^A = 1$  if type A MTF location  $i \in I_A$  is utilized and 0 otherwise. Likewise,  $y_i^B = 1$  if type B MTF location  $i \in I_B$  is utilized and 0 otherwise.

The next sets of binary variables  $x_{ijk}^A$  and  $x_{ijk}^B$  capture the dispatch preference list assignment of type A and type B air assets to casualties. These decision variables are defined to reflect the response protocols to casualties of different types. For example, only type A air assets can be dispatched to casualties of type  $C_A$ , and type A air assets must be dispatched to casualties of type  $C_{AB}$  before type B air assets are dispatched as backup coverage to casualties of type  $C_{AB}$ . Likewise, only type B air assets can be dispatched to casualties of type  $C_B$ , and type B air assets must be dispatched to

casualties of type  $C_{BA}$  before type A air assets are dispatched in secondary support to casualties of type  $C_{AB}$ . Therefore,  $x_{ijck}^A = 1$  if type A air asset  $i \in I_A$  is the  $k^{th}$ ,  $k = 1, \dots, s_A$ , preferred air asset to respond to a casualty at location  $j \in J$  of type  $c \in C_A \cup C_{AB}$ , or if type A air asset location  $i \in I_A$  is the  $k^{th}$ ,  $k = s_B + 1, \dots, s_B + s_A$ , preferred air asset to respond to a casualty of type  $c \in C_{BA}$ , and 0 otherwise. Likewise,  $x_{ijck}^B = 1$  if type B air asset location  $i \in I_B$  is the  $k^{th}$ ,  $k = 1, \dots, s_B$ , preferred air asset to respond to a casualty at location  $j \in J$  of type  $c \in C_B \cup C_{BA}$ , or if type B air asset location  $i \in I_B$  is the  $k^{th}$ ,  $k = s_A + 1, \dots, s_A + s_B$ , preferred air asset to respond to a casualty of type  $c \in C_{AB}$ , and 0 otherwise.

Now, we formally state the binary linear program.

$$\max_{c \in C_A} \sum_{j \in J} \sum_{i \in I_A} \sum_{k=1}^{s_A} h_{ijck}^A x_{ijck}^A \quad (3.13)$$

$$\text{subject to } \sum_{i \in I_A} x_{ijck}^A = 1 \quad \forall j \in J, c \in C_A \cup C_{AB}, k = 1, \dots, s_A \quad (3.14)$$

$$\sum_{i \in I_A} x_{ijck}^A = 1 \quad \forall j \in J, c \in C_{BA}, k = s_B + 1, \dots, s_B + s_A \quad (3.15)$$

$$\sum_{i \in I_B} x_{ijck}^B = 1 \quad \forall j \in J, c \in C_B \cup C_{BA}, k = 1, \dots, s_B \quad (3.16)$$

$$\sum_{i \in I_B} x_{ijck}^B = 1 \quad \forall j \in J, c \in C_{AB}, k = s_A + 1, \dots, s_A + s_B \quad (3.17)$$

$$\sum_{k=1}^{s_A} x_{ijck}^A = y_i^A \quad \forall j \in J, c \in C_A \cup C_{AB}, i \in I_A \quad (3.18)$$

$$\sum_{k=s_B+1}^{s_B+s_A} x_{ijck}^A = y_i^A \quad \forall j \in J, c \in C_{BA}, i \in I_A \quad (3.19)$$

$$\sum_{k=1}^{s_B} x_{ijck}^B = y_i^B \quad \forall j \in J, c \in C_B \cup C_{BA}, i \in I_B \quad (3.20)$$

$$\sum_{k=s_A+1}^{s_A+s_B} x_{ijck}^B = y_i^B \quad \forall j \in J, c \in C_{AB}, i \in I_B \quad (3.21)$$

$$\sum_{j \in J} \left( \left( \sum_{i \in I_A} \sum_{k=1}^{s_A} h_{ijck}^A x_{ijck}^A \right) + \left( \sum_{i \in I_B} \sum_{k=s_A+1}^{s_A+s_B} P_{s_A}^A h_{ijck}^B x_{ijck}^B \right) \right) \geq \theta_c \quad \forall c \in C_{AB} \quad (3.22)$$

$$\sum_{j \in J} \left( \left( \sum_{i \in I_A} \sum_{k=s_B+1}^{s_B+s_A} P_{s_B}^B h_{ijck}^A x_{ijck}^A \right) + \left( \sum_{i \in I_B} \sum_{k=1}^{s_B} h_{ijck}^B x_{ijck}^B \right) \right) \geq \theta_c \quad \forall c \in C_{BA} \quad (3.23)$$

$$\sum_{j \in J} \sum_{i \in I_B} \sum_{k=1}^{s_B} h_{ijk}^B x_{ijk}^B \geq \theta_c \quad \forall c \in C_B \quad (3.24)$$

$$\sum_{i \in I_A} y_i^A = s_A \quad (3.25)$$

$$\sum_{i \in I_B} y_i^B = s_B \quad (3.26)$$

$$x_{ijk}^A \leq y_i^A \quad \forall i \in I_A, j \in J, c \in C_A \cup C_{AB}, k = 1, \dots, s_A \quad (3.27)$$

$$x_{ijk}^A \leq y_i^A \quad \forall i \in I_A, j \in J, c \in C_{BA}, k = s_B + 1, \dots, s_B + s_A \quad (3.28)$$

$$x_{ijk}^B \leq y_i^B \quad \forall i \in I_B, j \in J, c \in C_B \cup C_{BA}, k = 1, \dots, s_B \quad (3.29)$$

$$x_{ijk}^B \leq y_i^B \quad \forall i \in I_B, j \in J, c \in C_{AB}, k = s_A + 1, \dots, s_A + s_B \quad (3.30)$$

$$\sum_{j \in J} \sum_{c \in C_A \cup C_{AB}} \sum_{k=1}^{s_A} g_{ijk}^A x_{ijk}^A + \sum_{j \in J} \sum_{c \in C_{BA}} \sum_{k=s_B+1}^{s_B+s_A} p_{s_B}^B g_{ijk}^A x_{ijk}^A \leq (r^A + \delta^A) y_i^A \quad \forall i \in I_A \quad (3.31)$$

$$\sum_{j \in J} \sum_{c \in C_B \cup C_{BA}} \sum_{k=1}^{s_B} g_{ijk}^B x_{ijk}^B + \sum_{j \in J} \sum_{c \in C_{AB}} \sum_{k=s_A+1}^{s_A+s_B} p_{s_A}^A g_{ijk}^B x_{ijk}^B \leq (r^B + \delta^B) y_i^B \quad \forall i \in I_B \quad (3.32)$$

$$\sum_{j' \in N_j} x_{ij'c1}^A \geq x_{ijc1}^A \quad \forall i \in I_A, c \in C_A \cup C_{AB}, j \in J \quad (3.33)$$

$$\sum_{j' \in N_j} x_{ij'c1}^B \geq x_{ijc1}^B \quad \forall i \in I_B, c \in C_B \cup C_{BA}, j \in J \quad (3.34)$$

$$x_{ijk}^A \in \{0, 1\} \quad \forall i \in I_A, j \in J, c \in C, k = 1, \dots, s_A + s_B \quad (3.35)$$

$$x_{ijk}^B \in \{0, 1\} \quad \forall i \in I_B, j \in J, c \in C, k = 1, \dots, s_A + s_B \quad (3.36)$$

$$y_i^A \in \{0, 1\} \quad \forall i \in I_A \quad (3.37)$$

$$y_i^B \in \{0, 1\} \quad \forall i \in I_B \quad (3.38)$$

The objective function in (3.13) maximizes the expected coverage of  $C_A$  casualties. In the BLP model, only type A air assets respond to  $C_A$  casualties.  $C_{AB}$  casualties prefer type A air assets due to operational characteristics such as in-flight medical treatment ability and flight speed. Constraints (3.14) assign one type A air asset in each of the first  $1, \dots, s_A$  positions in the preference list to casualties of type  $C_A$  or  $C_{AB}$ . Casualties of type  $C_{BA}$  are assigned all of the type B air assets first in the preference list before the type A air assets are assigned in the preference list as seen in constraints (3.15). Likewise, only type B air assets respond to  $C_B$  casualties and  $C_{BA}$  casualties prefer type B



air assets. Constraints (3.16) assign one type B air asset in each of the first  $1, \dots, s_B$  positions of the preference list to casualties of type  $C_B$  or  $C_{BA}$ . Casualties of type  $C_{AB}$  are assigned all of the type A air assets in the preference list first before the type B air assets are assigned in the preference list as seen in constraints (3.17). Constraints (3.18) – (3.21) assign each air asset in the system to only one position in the preference list. In other words, a specific air asset cannot be both the first and third preferred air asset for the same casualty location. Constraints (3.22), (3.23), and (3.24) enforce minimum pre-specified thresholds  $\theta_c$  of the expected coverage for the non-objective casualty types  $c \in C_{AB} \cup C_{BA} \cup C_B$ . The minimum levels of service are typically defined by a division surgeon cell or senior medical planners. Constraints (3.25) and (3.26) locate  $s_A$  type A and  $s_B$  type B air assets at MTFs in the system, respectively. Constraints (3.27) – (3.30) make sure that only air assets located at open MTF facility locations are utilized in construction of the preference list. Constraints (3.31) compute and balance the type A workloads within a pre-specified  $\delta^A$  of the average workload. In a similar manner, constraints (3.32) compute and balance the type B workloads. Constraints (3.33) and (3.34) enforce contiguity amongst each type A and type B air asset's set of first priority casualty locations. Two locations are contiguous to each other if the two locations differ by one unit in either the  $x$  or  $y$  direction. Lastly, constraints (3.35) – (3.36) and (3.37) - (3.38) require the preference list and the locating variables to be binary, respectively. In the next section, we present an example to apply the BLP model to a military medical system.

### 3.5 Computational Example

We now present a computational example to illustrate how to locate two classes of air assets and assign them to multiple types of casualties in a dispatch preference list. In the ongoing stability operations in support of Operation Enduring Freedom (OEF) in Afghanistan, coalition forces from the North Atlantic Treaty Organization (NATO) leverage both U.S. Army and U.S. Air Force helicopters in support of combat troops conducting operations on the ground. Specifically, the U.S. Army air assets (UH-60A/L and HH-60 MEDEVAC) and the U.S. Air Force air asset (HH-60G Pave Hawk) are primarily responsible for the aeromedical evacuation of casualties [6]. The HH-60 MEDEVAC

platform – referred to as air asset type A – has been specifically modified with a state-of-the-art medical interior to accommodate acute care casualties and has a cruising speed of 278 km/hour. The HH-60G Pave Hawk air asset – referred to as air asset type B – is a slower air asset with a cruising speed of 240 km/hour and does not have the in-flight medical capabilities to accommodate acute care casualties [1]. In this chapter, we assume that both type A and type B air assets fly single-ship (i.e., solo) during MEDEVAC operations, in contrast to tandem operations in which type A air assets must be accompanied by a security/escort asset due to risk of enemy action. Knowing where to locate both type A and type B air assets and how to dispatch them to battle triaged casualties will help increase survivability of the most urgent casualties.

We apply our air MEDEVAC asset optimization model to NATO Regional Command - South (RC-South), one of the four regional commands which span several Afghanistan provinces, to illustrate how air asset management can lead to more effective military medical systems. Figure 1 depicts the area of RC-South divided into seventy-five distinct casualty locations. Twenty-five potential sites for medical treatment facilities (MTF) are distributed throughout RC-South, represented by the red crosses in Figure 1. The base case example solved in this section will optimally locate ten type A air assets and four type B air assets at MTFs within RC-South. Note that the casualty data (from 2001 through 2012) was one-time extracted from the Defense Casualty Analysis System (DCAS).

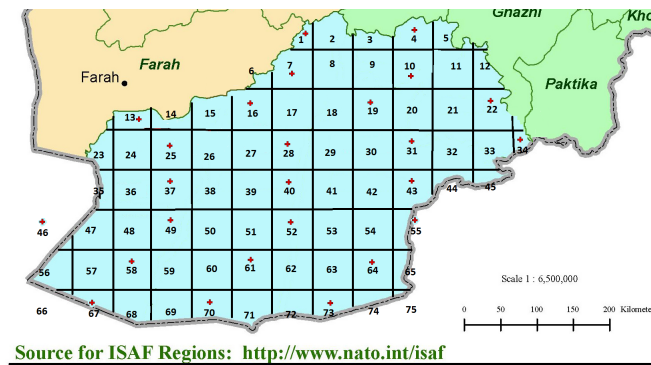


Figure 3.1: 25 Potential MTF sites in NATO Regional Command-South (represented by red crosses)

The following organization of casualty types is used to be consistent with the notation of sections

3.3 and 3.4. Casualty type  $C_A$  contains the CAT A casualties,  $C_{AB}$  contains the CAT B casualties, and  $C_B$  contains the CAT C casualties. Note that there are no  $C_{BA}$  call types in this example.

We next derive the BLP model inputs. The type A service times  $\tau_{ijc}^A$  (hours) are computed by dividing the Euclidean distance (kilometers) between a MTF location  $i \in I_A$  and a casualty location  $j \in J$  by the type A air asset speed (kilometers/hour). The type B service times  $\tau_{ijc}^B$  are computed similarly except that the service time calculation is multiplied by two since medical treatment does not begin until a casualty returns back to the MTF (in contrast to the medical treatment provided at the casualty location with a type A air asset). The distribution of casualty type is as follows: CAT A 75%, CAT B 20%, and CAT C 5%. Further, the total casualty arrival rate  $\lambda = 3.0$  and each casualty location  $j \in J$  has equal chance of a casualty event. Evacuation of a CAT A casualty by a type A air asset results in an expected coverage between 0.8 – 1.0, depending on how many minutes have elapsed up to the threshold of one hour. Therefore, we use an exponential decreasing curve to model the rapid deteriorating health of CAT A casualties after one hour. Similar translated exponential decreasing curves are used for the evacuation of CAT B and CAT C casualties to reflect time-standards of four hours and twenty-four hours, respectively. We use a time-standard of six hours for CAT C casualties to encourage the rapid evacuation of all casualty types.  $R_{ijc}^A$  and  $R_{ijc}^B$  depend on the expected type A and type B air asset service time between casualty location  $j$  and air asset location  $i$  (denoted  $t$  in the equations below) as well as the casualty type  $c \in C$ .  $R_{ijc}^A$  and  $R_{ijc}^B$  are given by the following equations. For  $c \in \text{CAT A}$ :  $R_{ijc} = -0.2t + 1$  if  $t \leq 1$  and  $e^{-1.6(t-1)}$  otherwise, for  $c \in \text{CAT B}$ :  $R_{ijc} = -0.05t + 1$  if  $t \leq 4$  and  $e^{-1.6(t-4)}$  otherwise, and for  $c \in \text{CAT C}$ :  $R_{ijc} = -0.05t + 1$  if  $t \leq 6$  and  $e^{-1.6(t-6)}$  otherwise.

We perform all computations on laptop with a AMD A6 2.10 GHz processor and 8GB RAM, leveraging Python 2.6 for parameter calculations and Gurobi 5.6.2 for optimization of the BLP model. The following algorithm is executed to solve the optimization model. Note that each optimization iteration contains an instance of the approximate Hypercube algorithm. Optimize the BLP using the most recent values of the following system statistics: 1) the average busy probabilities  $r^A$  and  $r^B$ , 2) the independence correction factors  $Q_A$  and

$Q_B$ , and 3) the offered load imbalance thresholds  $\delta_A$  and  $\delta_B$ . Optimize a new instance of the problem until the workload of each type A and type B air assets is less than 5% more than the average air asset workload. After each optimization iteration, the current solution contains the location of the air assets and the dispatch preference list. Using the dispatch preference list, execute the approximate Hypercube algorithm until the max change in type A and type B busy probabilities is less than 0.001. The system performance statistics  $(r^A, r^B, Q_A(s_A, \rho^A, k), Q_B(s_B, \rho^B, k))$  are updated and passed as input to the next instance of the BLP model. The offered load imbalance thresholds  $\delta_A$  and  $\delta_B$  are tightened to the new standard deviation of the type A and type B busy probabilities. For the base case example, the optimization algorithm converged in four iterations. The approximate Hypercube algorithm converged in six or fewer iterations within each iteration of the optimization algorithm. The base case example corresponds to a BLP with 67,757 constraints and 52,602 variables. Solving the optimization algorithm on the base case ten times resulted in the following run time metrics: a minimum CPU time of 908 seconds (15.1 minutes), a maximum CPU time of 1,132 seconds (18.9 minutes), and an average CPU time of 1,001 seconds (16.7 minutes).

Table 3.1 presents the ten type A and four type B air asset locations selected by the BLP model as well as the corresponding busy probabilities. MTF43 is the only location with both a type A and a type B air asset. The type A busy probabilities range from the low of 0.051 (MTF07) to the high of 0.103 (MTF25). The type B busy probabilities range from the low of 0.224 (MTF37) to the high of 0.358 (MTF43). All air asset busy probabilities are balanced to be no greater than 5% more than the respective type A or type B average busy probability.

Table 3.2 presents the independence correction factors for type A and type B air assets. The type A factors quantify the correction to the probability of obtaining  $k$  busy type A air assets followed by an available type A air asset when assuming that air assets operate independently. Note that  $Q_A(s_A, \rho^A, 0) = Q_B(s_B, \rho^B, 0) = 1.0$  because the probability of

Table 3.1: Location and Busy Probabilities for Type A and Type B Air Assets in the Base Case Example

The location of the $s_A = 10$ type A air assets	MTF25	MTF22	MTF28	MTF07	MTF46	MTF61	MTF64	MTF43	MTF10	MTF58
Type A air asset busy probabilities $r_i^A$	0.103	0.058	0.098	0.051	0.068	0.086	0.076	0.061	0.081	0.070
Average type A air asset busy probability $r^A$	0.075									
The location of the $s_B = 4$ type B air assets	MTF43	MTF73	MTF37	MTF19						
Type B air asset busy probabilities $r_i^B$	0.358	0.339	0.224	0.320						
Average type B air asset busy probability $r^B$	0.310									

obtaining an available air asset when no air assets are busy is equal to one.

Table 3.2: Independence Correction Factors for Type A and Type B Air Assets in the Base Case Example

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$
$Q_A(s_A, \rho^A, k)$	1.0	0.907	0.878	0.907	0.995	1.155	1.411	1.805	2.405	3.318
	$k=0$	$k=1$	$k=2$	$k=3$						
$Q_B(s_B, \rho^B, k)$	1.0	0.856	0.807	0.830						

The objective function of the BLP is the expected fraction of calls covered when dispatching type A air assets to CAT A casualties. It is desirable to know the performances changes in the system as additional type A air assets are introduced or removed from the system. Figure 2 presents a sensitivity analysis on the expected coverage of CAT A casualties with respect to the number of type A air assets in the system as well as the call arrival rate  $\lambda$ . Figure 2 can be used by decision-makers to identify how many type A air assets are needed for a given benchmark. For example, if the desired system performance is 90% expected coverage, then six type A air assets will be enough when the call volume  $\lambda$  is between 2.5 and 3.5 calls per hour. However, to achieve 95% expected coverage in the same call volume range, nine type A air assets are required. In figure 2, there is a diminishing return on the objective function value with respect to the number of type A air assets. For example, when  $\lambda = 3.0$ , consider the large increase in expected coverage when  $s_A = 2$  (53%) and when  $s_A = 5$  (89%) to the small increase when  $s_A = 8$  (94%) and when  $s_A = 10$  (95%).

Figures 3.3(a) - 3.3(c) illustrate the contiguity of the 1st priority response districts created for CAT A, CAT B, and CAT C casualties, respectively. Figures 3.3(a) and 3.3(b) show type A air assets because the first priority for CAT A and CAT B casualties must be a type A

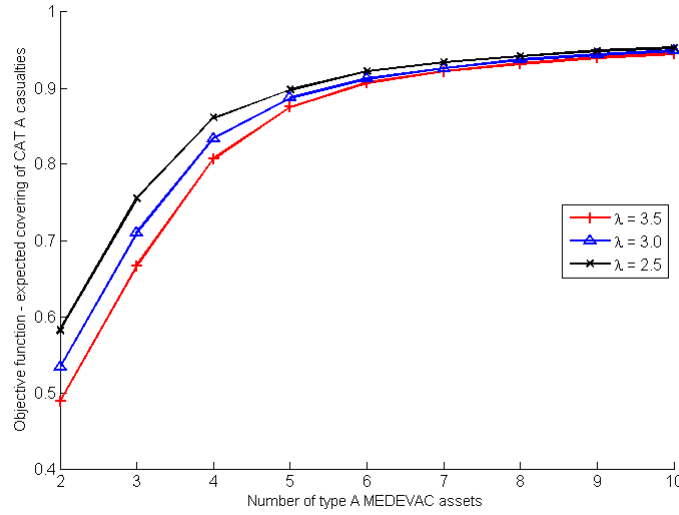
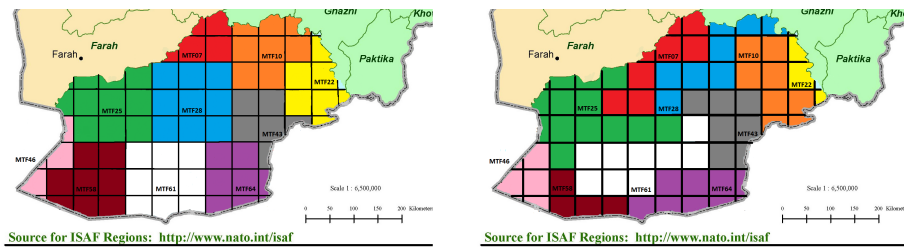


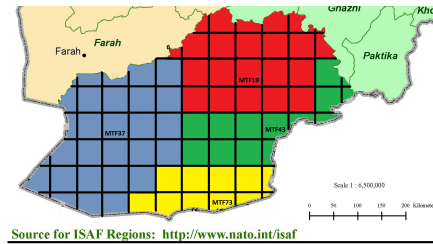
Figure 3.2: Sensitivity Analysis of the Objective Function for the Number of Type A Air Assets and Call Arrival Rate  $\lambda$ .

air asset. The BLP model identifies the set of type A air asset locations that maximize the fraction of CAT A casualties covered within 60 minutes. The set of locations selected must also balance air asset workloads while considering busy probabilities and accounting for air assets that are dependent. Likewise, figure 3.3(c) shows type B air assets because the first priority for each CAT C casualty must be a type B air asset.

The objective function used in this computational example encourages dispatch of the closest air asset to a casualty incident. Dispatching the closest air asset leads to the shortest service time and subsequently the largest expected coverage. However, due to the unavailability of air assets in a crowded system, it is sometimes optimal to ration the closest available air asset. Table 3 presents the difference in objective function value between optimality and the heuristic policy of dispatch the closest air asset to a casualty. To implement the closest air asset heuristic,  $s_A$  type A and  $s_B$  type B air assets must be selected for use in the priority list. Therefore, a rank ordering of all potential MTF sites is created based upon the average distance to all seventy five casualty locations. Once the air asset



(a) 1st Priority Response Districts of Type A Air Assets Assigned to CAT A Casualties (b) 1st Priority Response Districts of Type A Air Assets Assigned to CAT B Casualties



(c) 1st Priority Response Districts of Type B Air Assets Assigned to CAT C Casualties

Figure 3.3: Contiguous 1st priority dispatching assignments for type A and type B air assets in the base case example. Casualty locations that comprise an air asset's 1st priority district are the same color.

locations are known, the priority list is created based on proximity between each casualty location and each asset location. To evaluate the objective function value of the heuristic, to locating and dispatching variables are set to 1.0, the system input parameters are calculated, and the instance is solved in Gurobi. Table 3 shows that the heuristic of dispatching the closest air asset performs very well. The largest difference in objective function value between optimality and the heuristic over the range of  $s_A$  and  $s_B$  values considered is 3.99%. In some scenarios, such as  $s_A = 2$  and  $s_B = 4$ , the heuristic policy is optimal. Note that in Table 3 the base case value of  $\lambda = 3$  is used.

Table 3.3: Optimal Solution Versus Closest Server Heuristic Solution with Additional Type A Air Assets

$s_A$	$s_B$	Heuristic Obj Value	Optimal Obj Value	Obj Improvement (percent)
2	4	0.533	0.533	0%
3	4	0.710	0.710	0%
4	4	0.805	0.834	3.60%
5	4	0.853	0.887	3.99%
6	4	0.884	0.912	3.17%
7	4	0.906	0.926	2.21%
8	4	0.916	0.937	2.29%
9	4	0.922	0.944	2.39%
10	4	0.933	0.949	1.71%

The comparison of optimality to the dispatch closest heuristic is examined further in Table 4. The call arrival rate is varied from  $\lambda = 1$  to 10. Under this range of values, the heuristic of dispatching the closest air asset again performs very well. The largest difference in heuristic objective function value from optimality is 1.71%. Note that in Table 4 the base case values of  $s_A = 10$  and  $s_B = 4$  are used. These improvements, while small in absolute value, are large relative to other methodological improvements in this domain.

This computational example illustrates how to locate two classes of air assets and construct an air MEDEVAC asset dispatch preference list for all casualty locations. The base case example developed in this section illustrates how to locate and dispatch 10 type



Table 3.4: Optimal Solution Versus Closest Server Heuristic Solution with Varying Call Arrival Rate  $\lambda$

Call Arrival Rate $\lambda$	Heuristic Obj Value	Optimal Obj Value	Obj Improvement (percent)
1	0.941	0.957	1.70%
3	0.933	0.949	1.71%
6	0.918	0.932	1.53%
10	0.908	0.911	0.33%

A and 4 type B air assets to maximize the expected coverage of CAT A casualties within sixty minutes, while balancing the air asset workload. A dispatch preference list is valuable to senior military decision-makers and medical planners to identify preferred back-up air assets in a crowded military MEDEVAC system.

### 3.6 Concluding Remarks

In this chapter, we introduced a binary linear programming model that optimally locates two classes of military air MEDEVAC assets and assigns air assets to casualties in a dispatch preference list. Queuing system performance statistics, including busy probabilities and independence correction factors, are derived for air assets and are leveraged as inputs to the BLP model. Accounting for the queuing dynamics in military medical systems results in a more realistic optimization model in which air assets are not always available for service. The problem of military air MEDEVAC asset management is important because military medical systems save lives by responding to multiple types of casualty incidents. By providing prompt en route medical care to the most urgent casualties, the likelihood of survival is increased. The effective and efficient medical evacuation of casualties serves to keep troop morale high while maintaining a healthy and sustained fighting force.

The model in this chapter can be extended to consider more than two types of air assets. The resulting spatial queuing model and BLP can be derived using methods that are analogous to those in this chapter.

This chapter also can be extended to account for batch arrivals of casualties to the system where the volume of casualties requires the dispatch of more than one air asset. Also, the initial triage of casualties is not perfect and can lead to misclassification of the severity of a casualty incident. The logistics of military medical systems are even more complicated when there are multiple air assets and then the accuracy of triage affects dispatch decisions. Also, in some cases, air assets must be assigned to and wait for security/escort assets. An optimization model to identify which air asset and security/escort asset pair to dispatch to casualty incidents to minimize MEDEVAC response time of urgent casualties would be of value to military decision-makers. Work is currently in progress to address these issues.

## Co-locating Military Aeromedical Evacuation Assets and Security Escorts

### 4.1 Introduction

Military aeromedical evacuation (medevac) systems save lives by responding to casualty incidents and transporting the most urgent casualties to a medical treatment facility (MTF) via a fleet of distinguishable, geographically-dispersed evacuation assets (servers) dedicated to the response of casualties. There are distinguishable classes of evacuation assets within military medical systems. The first class of assets (type A) are fast and outfitted with the medical equipment needed to care for the most urgent casualties. However, the actionable service time of these fast assets is slowed by a mandatory wait time for an available escort asset to fly in tandem and provide security. In contrast, the next class of air assets (type B) are weaponized and therefore heavy and slow but do not have to wait for an escort before servicing a casualty incident. Also, type B air assets are not medically equipped to evacuate the most urgent casualties. Military medical systems are evaluated against a specified time standard, usually sixty minutes for high-priority casualties and a few hours for low-priority casualties. Therefore, minimizing the evacuation time of all casualties, including the wait time for an available escort asset, improves the performance of military medical systems.

The ranking of evacuation precedence is handled through a straightforward casualty triage scheme. CAT A: alpha category includes urgent and urgent-surgical casualties that need to be treated within one hour, CAT B: bravo category includes priority casualties that need to be treated within four hours, and CAT C: charlie category includes routine casualties that need to be treated within 24 hours. The fleet includes two classes of air assets to manage:

U.S. Army helicopters such as the UH-60A/L Black Hawk or HH-60 MEDEVAC platforms are normally leveraged to respond to casualties based on their speed, range, and en route medical capability. U.S. Air Force helicopters such as the HH-60G Pave Hawk are also available to respond to casualties but are slower because they carry heavy weapon systems and have a larger flight crew. When used together effectively, both Army and Air Force air assets can provide a more effective military MEDEVAC system.

We propose a novel integer programming (IP) model that optimally locates type A and type B evacuation assets, as well as the security escort assets necessary to accompany type A evacuation assets. The objective of the MIP model is to maximize the number of demands covered within a given time threshold. The model accounts for the probabilistic nature of service systems in which a facility is not always available for service.

This chapter is organized as follows. Section 4.2 provides a review of the relevant literature on military medical systems and stochastic facility location models. Section 4.3 introduces the medevac and escort locating integer programming model. Section 4.4 applies the model to a realistic military data set and provides a sensitivity analysis useful to senior military decision makers. Lastly, section 4.5 provides concluding remarks and directions for future research.

## 4.2 Literature Review

A number of existing research papers focus explicitly on systematically improving aeromedical evacuation of military casualties. Higgins [44] introduces the operational issues of U.S. Army aeromedical evacuation helicopters. Helicopter operational issues are pivotal in any research study involving casualties and aeromedical evacuation systems because speed of response directly increases likelihood of survival. Bastian [6] presents a multi-criteria decision analysis (MCDA) model using stochastic, mixed-integer goal programming to

determine the minimum number of air assets needed at each medical treatment facility. This in turn maximizes the coverage of the theater-wide casualty demand and the probability of meeting that demand. Fulton et al. [34] introduce a two-stage stochastic optimization modeling framework for the medical evacuation (ground and air) of casualties. This framework identifies optimal casualty evacuation sites and medical treatment facility sites in response to stochastic demands for service. Bastian and Fulton [7] present a geospatial-based decision-support tool that identifies which air asset to launch in response to a casualty event given knowledge of terrain, aircraft location, and aircraft capabilities. Zeto et al. [87] examine the pre-location of air assets, along with type and quantity, to maximize the theater-wide coverage while balancing air asset reliability. Bouma [16] develops the Medical Evacuation and Treatment Capabilities Optimization Model (METCOM) that considers policy effects on key measures of effectiveness, and then optimizes treatment and facility capacities for given casualty flows. Fulton et al. [35] present a Monte Carlo simulation to evaluate rules of allocation and planning considerations for Army air ambulance companies during major combat operations (MCO). Bastian et al. [8] examine the required capabilities of aeromedical evacuation of future U.S. aeromedical evacuation platforms by identifying the zero-risk aircraft ground speed. Bastian et al. [9] further examine three research issues surrounding future U.S. aeromedical evacuation platforms including optimal operational capabilities; trade-off considerations of different aircraft engines; and the effect of weaponizing the current MEDEVAC asset fleet on range, coverage radius, and response time. The models in this paper extend the work mentioned above by examining the different logistical problem of medevac/escort management.

Stochastic facility location models consider the real-world situation where one or more of the problem input values are not known with certainty. The following models all incorporate stochastic input. Facility location models have considered stochastic demand (Manne [57], Bean et al. [11], and Carbone [19]). Stochastic travel times in a network is considered by

Mirchandani [67] and Mirchandani and Odoni [68]. Berman and Odoni [15] and Berman and LeBlanc [12] consider the effect of uncertain travel times in a network on facility location-relocation problems. Daskin [27, 29] considers the situation where facilities are assumed to be busy with probability  $p$  and formulates an extension of the maximal covering problem (see above). In a different approach than Daskin, ReVelle and Hogan [75, 76] and ReVelle and Marianov [77] handle the stochastic nature of server availability by locating servers that covers demand with a reliability threshold. Marionov and ReVelle [58] extend the reliability model to incorporate multiple vehicle types for a given joint reliability threshold. As described above the  $p$ -median problem is central to facility location models. Weaver and Church [86] present computational procedures for the solution of stochastic  $p$ -median problems.

### 4.3 Model

In this section we introduce a model to strategically locate evacuation and escort assets on a military medical network. Type A evacuation assets must be accompanied by an escort asset to provide security. The following sets and parameters are used in the model.

$J$  = the set of casualty demand nodes

$I_A (I_B) (E)$  = the set of potential type A (type B) (escort) facility locations to locate assets

$H_j (L_j)$  = the high-priority (low-priority) demand generated at casualty location  $j \in J$

$T^H (T^L)$  = Time standard for the evacuation of high-priority (low-priority) casualties

$T^E$  = Time standard for the assignment of an escort asset to a type A evacuation asset

$t_{ij}^A (t_{ij}^B)$  = the service time when type A (type B) evacuation asset  $i \in I_A (I_B)$  responds to casualty location  $j \in J$

$t_{ei}^E$  = the service time when escort asset  $e \in E$  is assigned to a type A evacuation asset  $i \in I_A$

$a_{ij}^{AH} = 0$  if  $t_{ij}^A > T^H$  and 1 if  $t_{ij}^A \leq T^H$ , for  $j \in J$  and  $i \in I_A$ .

$a_{ij}^{AL} = 0$  if  $t_{ij}^A > T^L$  and 1 if  $t_{ij}^A \leq T^L$ , for  $j \in J$  and  $i \in I_A$ .

$a_{ij}^{BL} = 0$  if  $t_{ij}^B > T^L$  and 1 if  $t_{ij}^B \leq T^L$ , for  $j \in J$  and  $i \in I_B$ .

$a_{ei}^E = 0$  if  $t_{ei} > T^E$  and 1 if  $t_{ei} \leq T^E$ , for  $i \in I_A$  and  $e \in E$ .

$s_A$  ( $s_B$ ) ( $s_E$ ) = the number of type A (type B) (escort) assets to be located

$p_A$  ( $p_B$ ) = the busy probability of type A (type B) evacuation assets

$N_A$  ( $N_B$ ) ( $N_E$ ) = the upper bound of type A (type B) (escort) assets that can be located at a given facility location

$\alpha$  = the system wide reliability level required for an available escort asset

$\theta_i$  = the number of escorts required to be located within the neighborhood of type A asset  $i \in I_A$  to provide  $\alpha$ -reliability

$F_i$  = the escort service time per hour in neighborhood around type A evacuation asset  $i \in I_A$

The  $a_{ij}^{AH}$ ,  $a_{ij}^{AL}$ ,  $a_{ij}^{BL}$ ,  $a_{ei}^E$  parameters each represent the set of assets that are capable of covering the respective call for service. For example,  $a_{ij}^{AH}$  represents type A assets responding to high-priority casualty demands.  $a_{ij}^{AH} = 1$  if type A evacuation asset  $i \in I_A$  is able to cover high-priority casualty demand location  $j \in J$  within the pre-specified time threshold and 0 otherwise. Note that  $a_{ij}^{AL}$  and  $a_{ij}^{BL}$  are different parameters corresponding to type A or type B evacuation assets (with different operating speeds) covering low-priority casualty demand location  $j \in J$  within the time threshold given for low-priority casualties. Lastly, the  $a_{ei}^E$  parameter is equal to one if an escort asset  $e \in E$  covers a type A evacuation asset  $i \in I_A$  in an acceptable time standard.

The first set of binary variables represent the number of high-priority and low-priority demand coverings when locating type A and type B evacuation assets. For high-priority casualties,  $y_{jk}^H = 1$  if casualty location  $j \in J$  is covered by at least  $k = 1..s_A$  type A evacuation assets and 0 otherwise. Likewise for low-priority casualties,  $y_{jk}^L = 1$  if casualty location  $j \in J$  is covered by at least  $k = 1..s_A + s_B$  type A and type B evacuation assets and 0 otherwise. In this model, the type A evacuation assets must also be covered by security assets. Therefore,

the binary variable  $y_{ik}^E = 1$  if type A medevac location  $i \in I_A$  is covered by at least  $k = 1 \dots s_E$  escort assets and 0 otherwise. The remaining integer variables identify how many assets to locate at each potential facility location:  $x_i^A$  = the number of type A evacuation assets located at node  $i \in I_A$ ,  $x_i^B$  = the number of type B evacuation assets located at node  $i \in I_B$ , and  $x_e^E$  = the number of security escort assets located at node  $e \in E$ .

This model considers the reality of military medical systems in which evacuation assets are not always available for service. Let  $p_A$  represent the probability that a type A asset is busy. If we assume that the number of busy type A assets follows a binomial distribution, then the Prob ( $k$  type A assets busy out of  $M$  located) =  $\binom{s_A}{k} (1 - p_A)^k p_A^{s_A - k}$ ,  $k = 0, 1, \dots, s_A$ . The probability that casualty location  $j \in J$  is covered by a busy type A asset given  $m$  type A assets cover node  $j \in J$  is equal to 1 - prob( $m$  type A assets are not busy) =  $1 - p_A^m$ . Let  $X_{jm}$  be a random variable equal to the number of high priority demands at node  $j \in J$  covered by an available type A asset given  $m$  type A assets are capable of covering node  $j \in J$ .  $X_{jm} = H_j$  with probability  $1 - p_A^m$  and 0 with probability  $p_A^m$ . Therefore  $E(X_{jm}) = H_j(1 - p_A^m)$ , for all  $j, m$ . The increase in expected coverage at node  $j$  that results from increasing the number of type A assets that cover node  $j$  from  $m - 1$  to  $m$ , for  $m = 1, \dots, s_A$  is given by

$$\Delta E(X_{jm}) = E(X_{jm}) - E(X_{jm-1}) = H_j p_A^{m-1} (1 - p_A), m = 1, 2, \dots, s_A \quad (4.1)$$

In military medical systems, valuable response time is lost when a non weaponized evacuation asset has to wait for unavailable escort assets to become available for service. The model in this chapter ensures  $\alpha$ -reliability for escort assets, where escort assets are available for service at a proportion of time greater than  $\alpha$ . To do so, demand (non-weaponized type A evacuation assets) neighborhood specific busy probabilities are used. The following logic was presented in [29], [76] and [74]. It is assumed that the type A evacuation asset neighborhoods are only serviced by escort assets located in the neighborhood. To



compute the estimates of busy fractions,  $q_i$  for  $i \in I_A$ , the total busy time of the escorts in the neighborhood is divided by the total service time in that neighborhood.

$$q_i = \frac{\bar{t} \sum_{e \in E} a_{ei}^E f_k}{\sum_{e \in E} a_{ei}^E x_e^E} = \frac{F_i}{\sum_{e \in E} a_{ei}^E x_e^E} \quad \forall i \in I_A \quad (4.2)$$

$F_i$  is the escort asset service time per hour in the neighborhood around type A evacuation asset  $i \in I_A$  and  $\bar{t} \sum f_k$  is the total required escort asset service time in the neighborhood around type A evacuation asset  $i \in I_A$  per hour. The probability of at least one of the escorts in the neighborhood being available when type A evacuation asset requests service is  $1 - P(\text{all escorts in the neighborhood are busy})$ . Assume the probability of one or more escorts being busy follows a binomial distribution. Using the previous estimate of local busy probability for escorts  $q_i$  as the probability of a single escort being busy, then the probability of at least one escort being available is

$$1 - (q_i)^{\sum_{e \in E} a_{ei}^E x_e^E}. \quad (4.3)$$

Replacing  $q_i$  results in

$$1 - \left[ \frac{F_i}{\sum_{e \in E} a_{ei}^E x_e^E} \right]^{\sum_{e \in E} a_{ei}^E x_e^E} \geq \alpha. \quad (4.4)$$

Because  $\sum_{e \in E} a_{ei}^E x_e^E$  only takes integer values, define the parameter  $\theta_i$  as the smallest integer that satisfies

$$1 - \left[ \frac{F_i}{\theta_i} \right]^{\theta_i} \geq \alpha. \quad (4.5)$$

$\theta_i$  is the smallest number of escorts that must be located within the neighborhood of type A evacuation asset  $i \in I_A$  to ensure  $\alpha$  -reliability. The corresponding linear constraints to be added to the IP are

$$\sum_{e \in E} a_{ei}^E x_e^E \geq \theta_i \quad \forall i \in I_A. \quad (4.6)$$

Now the integer programming (IP) model is presented to strategically locate type A

and type B evacuation assets as well as escort assets to maximize the expected number of coverings of casualty demands in a military medical system. Note that only type A evacuation assets can respond to high priority casualty demand (H). However, both type A and type B evacuation assets can respond to low priority casualty demand (L).

$$\max \sum_{j \in J} \left[ H_j \left( \sum_{k=1}^{s_A} (1-p_A) p_A^{k-1} y_{jk}^H \right) + L_j \left( \sum_{k=1}^{s_A+s_B} (1-p_A)(1-p_B) (p_A p_B)^{k-1} y_{jk}^L \right) \right] \quad (4.7)$$

$$\text{subject to } \sum_{k=1}^{s_A} y_{jk}^H - \sum_{i \in I_A} a_{ij}^{AH} x_i^A \leq 0 \quad \forall j \in J \quad (4.8)$$

$$\sum_{k=1}^{s_A+s_B} y_{jk}^L - \left( \sum_{i \in I_A} a_{ij}^{AL} x_i^A + \sum_{i \in I_B} a_{ij}^{BL} x_i^B \right) \leq 0 \quad \forall j \in J \quad (4.9)$$

$$\sum_{k=1}^{s_E} y_{ik}^E - \sum_{e \in E} a_{ei}^E x_e^E \leq 0 \quad \forall i \in I_A \quad (4.10)$$

$$\sum_{k=1}^{s_E} y_{jk}^E \geq x_j^A \quad \forall j \in I_A \quad (4.11)$$

$$y_{jk}^E \leq y_{jk-1}^E \quad \forall j \in I_A, k \in 2..s_E \quad (4.12)$$

$$y_{jk}^H \leq y_{jk-1}^H \quad \forall j \in J, k \in 2..s_A \quad (4.13)$$

$$y_{jk}^L \leq y_{jk-1}^L \quad \forall j \in J, k \in 2..s_A + s_B \quad (4.14)$$

$$\sum_{i \in I_A} a_{ei}^E x_e^E \geq \theta_e x_e^E \quad \forall e \in E \quad (4.15)$$

$$\sum_{i \in I_A} x_i^A \leq s_A \quad (4.16)$$

$$\sum_{i \in I_B} x_i^B \leq s_B \quad (4.17)$$

$$\sum_{i \in E} x_i^E \leq s_E \quad (4.18)$$

$$x_i^A \leq N_A \quad \forall i \in I_A \quad (4.19)$$

$$x_i^B \leq N_B \quad \forall i \in I_B \quad (4.20)$$

$$x_i^E \leq N_E \quad \forall i \in E \quad (4.21)$$

$$x_i^A, x_k^B, x_e^E \in \mathbf{Z}^+ \quad \forall i \in I_A, k \in I_B, e \in E \quad (4.22)$$

$$y_{jk}^H, y_{ji}^L \in \{0, 1\} \quad \forall j \in J, k \in 1..s_A, i \in 1..s_A + s_B \quad (4.23)$$

The objective function in (4.7) maximizes the expected number of coverings of high-priority and low-priority casualties while taking into account the busy probabilities of type A and type B servers. The objective function coefficients  $(1 - p_A)p_A^{k-1}$  capture the increase in expected coverage that results from increasing the number of type A evacuation assets that cover node  $j \in J$  from  $k - 1$  to  $k$ , for  $k = 1, \dots, s_A$ . The first two sets of constraints given in (4.8) and (4.9) ensure the covering of a casualty demand comes from only the type A and type B assets located at MTF locations capable of providing service within the specified time threshold. The constraints given in (4.10) ensure the covering of a type A evacuation asset comes from only the escort assets located at stations capable of providing security within the specified time threshold. Constraints (4.11) make sure that the number of escort coverings is sufficient for the number of type A assets at each facility location. The constraints in (4.12) - (4.13) are the evacuation asset ordering constraints. Specifically, a casualty location must be covered by  $(k - 1)$  type A or type B assets before being covered by  $k$  type A or type B assets. Likewise in constraint (4.14),  $(k - 1)$  escort assets must cover a type A evacuation asset location before  $k$  escort assets cover a type A evacuation asset location. Constraints (4.15) are the  $\alpha$ -reliability constraints. The number of escort assets that must be located in the neighborhood of each type A evacuation asset to ensure  $\alpha$ -reliability is computed exogenously. Constraints (4.16), (4.17), and (4.18) ensure that no more than the number of available assets of each type are located on the network. Constraints (4.19) - (4.21) enforce upper bounds on the number of type A, type B, and escort assets at each facility location, respectively. Constraints (4.22) restrict the locating variables to be positive integers. Lastly,

constraints (4.23) restrict the demand covering variables to be binary.

#### 4.4 Computational Example

We now present a computational example to illustrate how to locate two types of evacuation assets and security escort assets to maximize expected covering of high-priority and low-priority casualty demand. In the operations in support of Operation Enduring Freedom (OEF) in Afghanistan, coalition forces from the North Atlantic Treaty Organization (NATO) leverage both U.S. Army and U.S. Air Force helicopters in support of combat troops conducting operations on the ground. Specifically, the U.S. Army air assets (UH-60A/L and HH-60 MEDEVAC) and the U.S. Air Force air asset (HH-60G Pave Hawk) are primarily responsible for the aeromedical evacuation of casualties [6]. The HH-60 MEDEVAC platform – referred to as evacuation asset type A – has been specifically modified with a state-of-the-art medical interior to accommodate acute care casualties and has a cruising speed of 278 km/hour. The HH-60G Pave Hawk air asset – referred to as evacuation asset type B – is a slower asset with a cruising speed of 240 km/hour and does not have the in-flight medical capabilities to accommodate acute care casualties [1]. In this chapter, we assume that type A evacuation assets fly tandem operations in which type A air assets must be accompanied by a security/escort asset due to risk of enemy action. The U.S. Army UH-60 Blackhawk is used as an escort asset complete with a self contained weapon system and has a cruising speed of 294 km/hour. Knowing where to locate escort assets in addition to both type A and type B evacuation assets increases the survivability of casualties in military medical systems.

We apply our military asset optimization model to NATO Regional Command - South (RC-South), one of the four regional commands which span several Afghanistan provinces, to illustrate how air asset management can lead to more effective military medical systems.

Figure 1 depicts the area of RC-South divided into seventy-five distinct casualty locations. Twenty-five potential sites for medical treatment facilities (MTF) are distributed throughout RC-South, represented by the red crosses in Figure 1. The category of high priority ( $H$ )

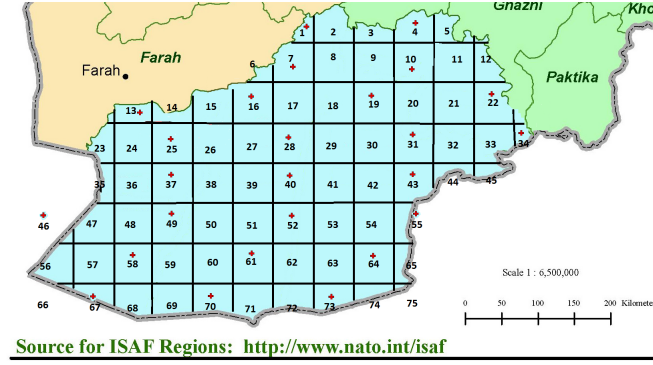


Figure 4.1: 25 Potential MTF sites in NATO Regional Command-South (represented by red crosses)

casualties contain the CAT A triaged casualties while the category of low priority ( $L$ ) casualties contain the CAT B and CAT C triaged casualties.

The set of MTF locations that cover each casualty location depends upon the service time of a responding evacuation asset to the casualty location. The type A service times  $t_{ij}^A$  (hours) are computed by dividing the Euclidean distance (kilometers) between a MTF location  $i \in I_A$  and a casualty location  $j \in J$  by the type A air asset speed (kilometers/hour). The set of MTF locations  $i \in I_A$  that cover casualty location  $j \in J$  are those where the service times are less than a specified time threshold. The type B service times  $t_{ij}^B$  are computed similarly except that the service time calculation is multiplied by two since medical treatment does not begin until a casualty returns back to the MTF (in contrast to the medical treatment provided at the casualty location with a type A air asset). Table 4.1 lists the number of high-priority and low-priority casualty locations  $j \in J$  that are covered by both type A and type B evacuation assets within the respective time standards. The numbers given in the table are the sum over  $j \in J$  of the elements of the parameters  $a_{ij}^{AH}$ ,  $a_{ij}^{AL}$ , and  $a_{ij}^{BL} = 1$  for all

$i \in I_A$  or  $i \in I_B$ . The type A evacuation assets must be covered within ten minutes by escort

Table 4.1: Number of covered casualty locations by evacuation assets

evacuation asset location	number of $H$ coverings by type A	number of $L$ coverings by type A	number of $L$ coverings by type B
01	28	75	31
04	28	68	30
07	35	75	37
10	33	69	34
13	27	74	28
16	41	75	43
19	44	74	44
22	23	62	25
25	36	75	39
28	49	75	52
31	41	72	42
34	19	59	22
37	40	75	43
40	51	75	55
43	41	73	44
46	21	59	23
49	41	75	41
52	46	75	49
55	35	75	37
58	30	68	32
61	40	75	44
64	33	75	34
67	21	61	23
70	30	75	35
73	25	75	30

assets to provide weaponized security. In RC-South all escort asset locations can only cover their own location within ten minutes, except for E01,E04,E07, and E10, each can cover one additional type A asset location.

In this computational example, the distribution of triaged casualty types is CAT A 75%, CAT B 20%, and CAT C 5%. Further, the total casualty arrival rate  $\lambda = 3.0$  calls per hour and each casualty location  $j \in J$  has equal chance of a casualty event =  $1/75 = 0.013$ . Therefore  $H_j = (3.0 * 0.75 * 0.013) = 0.03$  calls per hour for all  $j \in J$ . In a similiar manner,  $L_j = (3.0 * 0.25 * 0.013) = 0.01$  calls per hour for all  $j \in J$ .

In the following subsections, the IP model is applied to several different military medical system configurations. The first subsection contains a base case example to establish a benchmark system performance and asset location decisions. In the following subsection, the IP model inputs are varied over a range of potential values and the effect on the objective value and locating decisions is reported. For the base case example and the sensitivity

analysis, the IP model will locate type A and type B evacuation assets as well as escort assets at a set of potential MTF locations to maximize the expected covering of high-priority and low-priority casualty demand. The distribution of escort assets in the solutions must satisfy  $\alpha$ -reliability constraints for the system to guarantee escort asset availability.

#### 4.4.1 Base Case Example

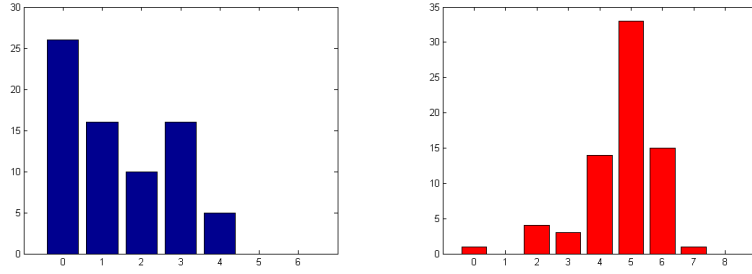
The base case example solved in this section will optimally locate ten type A air assets, four type B air assets, and seven escort assets at MTFs within RC-South. The input parameters of the base case include  $\alpha = 0.95$ ,  $N_A = N_B = N_E = 2$ ,  $p_A = 0.3$ , and  $p_B = 0.4$ . The defined time standards for type A assets is set to fifty minutes, escort assets ten minutes, and type B assets two hours.

The results of the base case example include an objective value of 2.154 and the following distribution of assets at potential MTF locations. For type A evacuation assets: 1 at MTF04, 1 at MTF10, 1 at MTF49, and 1 at MTF52. For type B evacuation assets: 1 at MTF31, 2 at MTF37, and 1 at MTF70. For escort assets: 2 at MTF10, 2 at MTF49, 2 at MTF52, and 1 at MTF73. It is important to note that the other MTF locations not listed in table 4.2 are not utilized to locate assets in the base case example.

Table 4.2: Distribution of each type of asset for the base case example solution

asset type	MTF04	MTF10	MTF31	MTF37	MTF49	MTF52	MTF70	MTF73
type A	1	1	0	0	1	1	0	0
type B	0	0	1	2	0	0	1	0
type E	0	2	0	0	2	2	0	1

Figures 4.2(a) and 4.2(b) illustrate the expected count of the number of high-priority and low-priority casualty demand locations that are covered, respectively. For example, 26 high-priority casualty locations are not covered by any type A assets, and 15 high-priority casualty locations are covered by 3 type A evacuation assets. Likewise, 34 low-priority casualty locations are covered by 5 type A *and* type B evacuation assets.



(a) expected number of high-priority coverings (b) expected number of low-priority coverings

Figure 4.2: Expected number of high and low-priority casualty location coverings

The base case example results in an IP model with 2104 constraints and 2050 variables. We perform all computations on laptop with a AMD A6 2.10 GHz processor and 8GB RAM leveraging Microsoft Excel for parameter calculations and CPLEX for optimization of the IP model. The optimization model was solved in 25 seconds. This subsection establishes a base case example for an average military medical system under “normal” operating characteristics. In the following subsection, we vary the operating characteristics of the data set considered in this computational example to examine how the asset location decision making changes.

#### 4.4.2 Sensitivity Analysis

In this subsection we vary the input parameters and report the changes to the locating decisions and system performance. Varying the input parameters mimics the reality of military medical systems in which system dynamics can change. It is important to have a solution to the medevac and escort locating problem for different potential system configurations. During the following sensitivity analysis it is assumed that a parameter is set to the same value as the base case example, unless otherwise stated. First, in table 4.3 the number of escorts that must be located to achieve different  $\alpha$ -reliability levels is reported. To interpret



Table 4.3: Number of required escort assets to achieve  $\alpha$  reliability

$\alpha = 0.70$	1 escort at each location
$\alpha = 0.80$	2 escorts at E01,E04,E07,E10 and 1 everywhere else
$\alpha = 0.90$	2 escorts at each location
$\alpha = 0.95$	2 escorts at each location
$\alpha = 0.99$	3 escorts at E01,E04,E07,E10 and 2 everywhere else
$\alpha = 0.999$	3 escorts at each location

table 4.3, there must be only one escort asset located in the neighborhood around each type A evacuation asset for  $\alpha$  reliability in the range of 0.7 to 0.79. However to achieve  $\alpha = 0.95$  reliability, two escorts must be located “near” each type A evacuation asset. Note that  $\alpha$  does not ever equal one, but a constraint forcing at least three escort assets to be close enough to each type A evacuation asset results in  $\alpha = 0.999$  reliability. That is an escort is available to provide security at a portion greater than 0.999 when requested.

Next the limiting effect of the number of escort assets on the system performance is illustrated. Figure 4.3 depicts how the objective function value changes with the addition of more type A evacuation assets and escort assets. More type A evacuation assets added to the

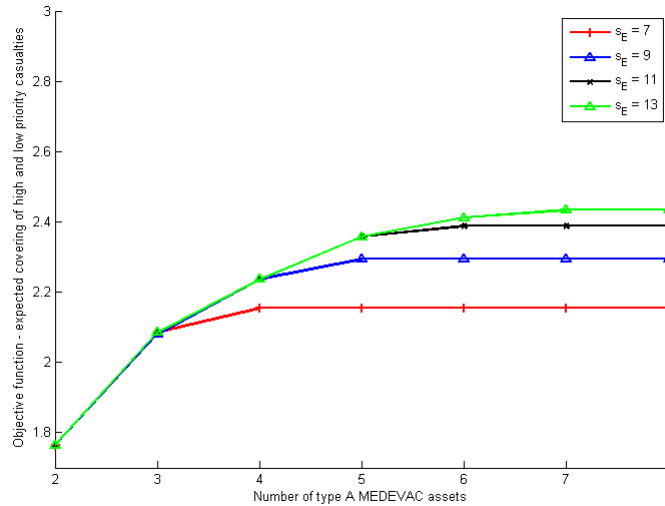


Figure 4.3: Sensitivity of the objective function value to the number of type A evacuation assets and the number of escort assets

system increases the expected number of high and low-priority coverings, as long as there are sufficient escort assets to provide security to the type A evacuation assets. Therefore, the system is quite dependent upon the  $s_E$  – the number of escort assets to be placed in the system.

In table 4.4 the robustness of the distribution of assets with respect to the upper bound on the number of assets that can be located at each potential MTF site is explored. It is

Table 4.4: Distribution of each type of asset with respect to varying upper bounds

number of assets	obj value	upper bound	asset type	distribution of assets
$s_A = 10, s_E = 7$	1.472	N = 1	type A	01,04,07,10
			type B	31,37,49,70
			type E	01,04,07,10
	2.154	N = 2	type A	04,10,49,52
			type B	31,37x2,70
			type E	10x2,49x2,55x2,73
	2.154	N = 3	type A	04,10,49,52
			type B	31,37x2,70
			type E	10x2,49x2,52x2,73
$s_A = 10, s_E = 10$	1.472	N = 1	type A	01,04,07,10
			type B	31,37,49,70
			type E	01,04,07,10
	2.388	N = 2	type A	04,10,37,40,43,49
			type B	31,37x2,70
			type E	10x2,37x2,40x2,43x2,49x2
	2.388	N = 3	type A	04,10,37,40,43,49
			type B	31,37x2,70
			type E	10x2,37x2,40x2,43x2,49x2
$s_A = 20, s_E = 14$	1.472	N = 1	type A	01,04,07,10
			type B	31,37,49,70
			type E	01,04,07,10
	2.467	N = 2	type A	04,10,37,40,43,49,55,58
			type B	31,37x2,70
			type E	04x2,37x2,40x2,43x2,49x2,55,58
	2.467	N = 3	type A	04,10,37,40,43,49,55,58
			type B	31,37x2,70
			type E	04x2,37x2,40x2,43x2,49x2,55,58
$s_A = 20, s_E = 20$	1.472	N = 1	type A	01,04,07,10
			type B	31,37,49,70
			type E	01,04,07,10
	2.507	N = 2	type A	04x2,10x2,37,40,43,46,49,52,55,58
			type B	04,37,49x2
			type E	04x2,10x2,37x2,40,43,46x2,49x2,52x2,55x2,58x2
	2.507	N = 2	type A	04x2,10x2,37,40,43,46,49,52,55,58
			type B	04,37,49x2
			type E	04x2,10x2,37x2,40,43,46x2,49x2,52x2,55x2,58x2

interesting that when  $N = 1$  and only one asset of each type can be located at each potential MTF location, the system performance does not change with the addition of more assets. That is even when twice as many type A and escort assets are available, the system is handcuffed by the upper bound of one asset of each type per location. The system appears to gain a lot in objective function value when the upper bound of the number of assets of each type increases to  $N = 2$ . However, the system shows no gain with additional increases on the upper bound of assets, such as  $N = 3, \dots$

In this subsection we varied several of the key model input parameters to illustrate the effect on system performance and location decisions. In military medical systems, senior decision makers must have a strategic and actionable plan to locate scarce military assets over many possible system configurations.

#### 4.5 Conclusions

This chapter introduces a novel IP model to examine the military evacuation and escort asset location problem. The availability of escorts when requested is an important defense issue because higher level medical treatment of casualties is delayed by unavailable escort assets. The IP model in this chapter locates both evacuation and escort assets to maximize the expected number of high-priority and low-priority casualties that are covered within defined time thresholds. The IP model ensures that enough escorts are located around each type A evacuation asset to achieve a user-specified reliability level. The use of integer programming to gain insight into hard military asset location management problems greatly increase soldier survivability. Senior military decision makers can use the model in this chapter as a guide to locate scarce military assets.

This chapter can also be extended to account for multiple casualty pickups per evacuation mission. Instead of a single casualty location being covered or not within a given time

standard, it would be interesting to see how system performance can improve if multiple casualty locations can be covered by an evacuation asset. Also, the initial triage of casualties is not perfect and can lead to misclassification of the severity of a casualty incident. The logistics of military medical systems are even more complicated when there multiple casualty locations to cover and then the accuracy of triage affects dispatch decisions.

## Summary

This dissertation presents research that examines how to improve dispatch, delivery, and location logistics for aeromedical evacuation of casualties in military medical systems. The aforementioned research highlights an opportunity to increase the survivability of casualties through design of effective aeromedical evacuation systems. The following sections in the chapter discuss the limitations, possible extensions, and applicability to other problem domains of the research in this dissertation.

### 5.1 Limitations and Drawbacks

Despite the quantifiable improvement to military medical systems through the use of operations research, there are practical limitations when applying mathematical modeling to complicated systems. For example, assumptions must be made about the underlying system. The distribution of calls for service in the system and the number of assets dispatched to each call for service are examples of assumptions that must be made when developing a model. The use of mathematical modeling invokes trade-offs between the assumptions necessary to depict realistic systems and the ability to solve the model. The nature of military systems results in input data that are classified or subject to semi-restricted access. Therefore, operations research models should be created and applied to input data that are representative of real-world military medical systems. While more realistic data leads to more realistic insights and conclusions, general insights provide proof of concept that the model works. Proof of concepts give decision makers confidence to analyze real data for

military medical systems and develop decision support systems.

## 5.2 Extensions

The research presented in this dissertation focuses on improving aeromedical evacuation of casualties in military medical systems. The operations research models introduced are a single approach to the problem studied. There are many opportunities to extend the research. This section presents possible extensions to the models in this dissertation. First, the dispatching and transporting MDP model from Chapter 2 can be extended in the following ways. The model does not account for distinguishable classes of aeromedical evacuation assets. In reality there are different types of assets with varying operational characteristics. Further, the model assumes that the location of aeromedical evacuation assets is known a priori. An extension would solve for optimal asset locations within the system instead of pre-locating assets. Also, the model does not consider any real-world sense of risk or mission vulnerability where enemy presence will dictate different decision making.

The locating and dispatching spatial queuing and BLP model from Chapter 3 can be extended in the following ways. First, the model only allows one aeromedical evacuation asset to be dispatched to a call for service. In reality, batch arrivals of casualties to the system where the volume of casualties requires the dispatch of more than one aeromedical evacuation asset is more realistic for military medical systems. Also, the initial triage of casualties is not perfect and can lead to misclassification of the severity of a casualty incident. Lastly, the model size limits its usefulness for larger problem instances. An appropriate extension to the research would develop a heuristic to find good feasible solutions quickly.

The aeromedical evacuation/escort model from Chapter 4 can be extended in the following ways. In reality, an enemy action can result in many casualties (in a batch ) located in hostile territory. In this very complicated situation, several unarmed aeromedical evacuation

assets must be dispatched along with the appropriate number of security assets. An appropriate extension to the research would identify how many evacuation assets and security escorts to dispatch in tandem. Also, demand of casualty evacuation is assumed consistent over time. It would be more realistic to assume that the frequency of calls for service follows a non-uniform distribution over the course of a day, subject to peak demand hours that dictate different system dynamics.

### 5.3 Other Problem Domains

The problem of scarce resource allocation is not restricted to only military medical systems. The following are some suggestions of other problem domains that can benefit from similar models to the ones introduced in this dissertation. First, aeromedical evacuation of casualties is a single element of the medical treatment of casualties within military medical systems. There are many other facets of military medical systems that can also be examined with operations research models to improve soldier survivability. For example, future casualty count estimation (forecasting), decision making when the enemy's strategy is unknown (game theory), and supplying forward surgical teams with necessary medical items while in hostile enemy territory (supply chain management) are all critical to the survivability of soldiers. Furthermore, doctors, surgeons, and nurses must be available (scheduling) at medical facilities available to provide service when requested. Also, the replacement parts for aeromedical evacuation assets are subject to a stochastic life cycle subject to many unknown factors (reliability theory). A maintenance schedule must be strategically developed to minimize the down time of evacuation assets.

Second, military medical systems share the majority of their operating characteristics with civilian emergency medical systems (EMS). The dispatching of civilian ambulances to calls for service suffer from the same spatial queuing dynamics and busy servers becoming

unavailable for service. The strategic locating of assets in anticipation of incoming demand is also central to civilian systems. City planning often tackles hard problems such as citing fire houses and/or police stations to keep assets ready to respond to an event so that the greatest number of citizens are covered within a time threshold or the distance is minimized to all areas of the community. The models developed for military medical systems can be naturally applied to the civilian sector.

Third, the aeromedical evacuation asset and escort model of Chapter 4 is relevant to other domains where an escort is required to protect an asset. For example, consider piracy of cargo containers on the ocean. There are a limited number of gun ships to provide protection and it is desirable to know how to assign escort boats to cargo ships to minimize unprotected wait time. Another application of assigning an escort to assets includes the movement of high ranking officials.

## 5.4 Conclusion

This dissertation examines the important problem of aeromedical evacuation of casualties in military medical systems. Operations research models are introduced to improve decision making in military medical systems and increase soldier survivability. Effective aeromedical evacuation of casualties directly saves lives by ensuring the most urgent casualties are evacuated and receive medical treatment in a timely manner. Further, the effective and efficient aeromedical evacuation of casualties serves to keep troop morale high while maintaining a healthy and sustained fighting force. Operations research was championed by the military during the second World War to gain insight into important logistical problems. New operations research models help gain further insight into one of the most important problems of military medical systems: effective aeromedical evacuation of casualties.



## Bibliography

- [1] Federation of American Scientists. *UH-60 Black Hawk, UH-60L Black Hawk, UH-60Q MEDEVAC, MH-60G Pave Hawk, HH-60G Pave Hawk, CH-60 Sea Hawk*. Accessed November 13, 2013. URL: <http://www.fas.org/programs/ssp/man/uswpns/air/rotary/sh60.html#hh60g>.
- [2] T. Andersson and P. Värbrand. “Decision support tools for ambulance dispatch and relocation”. In: *Journal of the Operational Research Society* 58 (2007), pp. 195–201.
- [3] S. Ansari, L. A. McLay, and M. E. Mayorga. *A maximum expected covering problem for locating and dispatching servers*. Under revision at *Transportation Science*. Virginia Commonwealth University, 2014.
- [4] Igor Averbakh and Oded Berman. “ $(p-1)(p+1)$ -approximate algorithms for  $p$ -traveling salesmen problems on a tree with minmax objective”. In: *Discrete Applied Mathematics* 75.3 (1997), pp. 201–216.
- [5] Igor Averbakh and Oded Berman. “Minmax regret median location on a network under uncertainty”. In: *INFORMS Journal on Computing* 12.2 (2000), pp. 104–110.
- [6] N. Bastian. “A Robust, Multi-criteria Modeling Approach for Optimizing Aeromedical Evacuation Asset Emplacement”. In: *Journal of Defense Modeling and Simulation* 7.1 (2010), pp. 5–23.
- [7] N. Bastian and L.V. Fulton. “Aeromedical Evacuation Planning Using Geospatial Decision-Support”. In: *Military Medicine* 179.2 (2014), pp. 174–182.

- [8] N. Bastian, L. V. Fulton, R. Mitchell, W. Pollard, D. Wierschem, and R. Wilson. “The Future of Vertical Lift: Initial Insights for Aircraft Capability and Medical Planning”. In: *Military Medicine* 177.7 (2012), pp. 863–869.
- [9] N. Bastian, D. Brown, L. Fulton, R. Mitchell, W. Pollard, M. Robinson, and R. Wilson. “Analyzing the Future of Army Aeromedical Evacuation Units and Equipment: A Mixed Methods, Requirements-Based Approach”. In: *Military Medicine* 178.3 (2013), pp. 321–329.
- [10] R. Batta, J. M. Dolan, and N. N. Krishnamurthy. “The maximal expected covering location problem: Revisited”. In: *Transportation Science* 23.4 (1989), pp. 277 –287.
- [11] James C Bean, Julia L Higle, and Robert L Smith. “Capacity expansion under stochastic demands”. In: *Operations Research* 40.3-supplement-2 (1992), S210–S216.
- [12] O Berman and B LeBlanc. “Location-relocation of mobile facilities on a stochastic network”. In: *Transportation science* 18.4 (1984), pp. 315–330.
- [13] Oded Berman. “Locating a facility on a congested network with random lengths”. In: *Networks* 15.3 (1985), pp. 275–293.
- [14] Oded Berman, Richard C Larson, and Samuel S Chiu. “Optimal server location on a network operating as an M/G/1 queue”. In: *Operations Research* 33.4 (1985), pp. 746–771.
- [15] Oded Berman and Amedeo R Odoni. “Locating mobile servers on a network with Markovian properties”. In: *Networks* 12.1 (1982), pp. 73–86.
- [16] Matthew F. Bouma. *Medical Evacuation and Treatment Capabilities Optimization Model METCOM*. Tech. rep. California: Naval Postgraduate School, 2005.
- [17] Joseph M. Bozell. “personal conversation 2013-03-27”. Captain USMC.

- [18] S. Budge, A. Ingolfsson, and E. Erkut. “Approximating vehicle dispatch probabilities for emergency service systems with location-specific service times and multiple units per location”. In: *Operations Research* 57 (2009), pp. 251–255.
- [19] Robert Carbone. “Public facilities location under stochastic demand”. In: *Infor* 12.3 (1974), pp. 261–270.
- [20] Yolanda M Carson and Rajan Batta. “Locating an ambulance on the Amherst campus of the State University of New York at Buffalo”. In: *Interfaces* 20.5 (1990), pp. 43–49.
- [21] K. R. Chelst and Z. Barlach. “Multiple unit dispatches in emergency services: Models to estimate system performance”. In: *Management Science* 27.12 (1981), pp. 1390–1409.
- [22] K. R. Chelst and J. P. Jarvis. “Estimating the probability distribution of travel times for urban emergency service systems”. In: *Operations Research* 27.1 (1979), pp. 199–204.
- [23] R. Church and C. ReVelle. “The maximal covering location problem”. In: *Papers of the Regional Science Association* 32 (1974), pp. 101–108.
- [24] J. J. Clawson, R. L. Martin, G. A. Cady, and R. Sinclair. *EMD: Making the most of EMS*. Report available at [http://firechief.com/ems/firefighting\\_emd\\_making\\_ems/](http://firechief.com/ems/firefighting_emd_making_ems/). 1999.
- [25] Paul Cunningham MD, Robert Rutledge MD, Christopher C. Baker MD, and Thomas V. Clancy MD. “A comparison of the association of helicopter and ground ambulance transport with the outcome of injury in trauma patients transported from the scene”. In: *The Journal of Trauma* 43.6 (1997), p. 940.

- [26] John Current, Samuel Ratick, and Charles ReVelle. "Dynamic facility location when the total number of facilities is uncertain: A decision analysis approach". In: *European Journal of Operational Research* 110.3 (1998), pp. 597–609.
- [27] Mark S Daskin. "Application of an Expected Covering Model to Emergency Medical Service System Design". In: *Decision Sciences* 13.3 (1982), pp. 416–439.
- [28] Mark S Daskin, Susan M Hesse, and Charles S Revelle. " $\alpha$ -reliable p-minimax regret: a new model for strategic facility location modeling". In: *Location Science* 5.4 (1997), pp. 227–246.
- [29] M.S. Daskin. "A maximum expected covering location model: Formulation, properties and heuristic solution". In: *Transportation Science* 17.1 (1983), pp. 48 –70.
- [30] MS Daskin and A Haghani. "Multiple vehicle routing and dispatching to an emergency scene". In: *Environment and Planning A* 16.10 (1984), pp. 1349–1359.
- [31] Perino M Dearing and Richard L Francis. "A minimax location problem on a network". In: *Transportation Science* 8.4 (1974), pp. 333–343.
- [32] S. Doane. *Medevac Team Saves Lives in Afghanistan*. Accessed February 13, 2014. URL: <http://www.cbsnews.com/news/medevac-team-saves-lives-in-afghanistan>.
- [33] J. V. Dunford. "Emergency medical dispatch". In: *Emergency Medicine Clinics of North America* 20 (2002), pp. 859 –875.
- [34] L. Fulton, P. McMurry, and B. Kerr. "A Monte Carlo Simulation of Air Ambulance Requirements During Major Combat Operations". In: *Military Medicine* 174.6 (2009), pp. 610–614.

- [35] L. Fulton, L. Lasdon, R. McDaniel, and M. Coppola. “Two-state Stochastic Optimization for the Allocation of Medical Assets in Steady-state Combat Operations”. In: *Journal of Defense Modeling and Simulation* 7.2 (2010), pp. 89–102.
- [36] Nicholas C Georgantzas and William Acar. *Scenario-driven planning: learning to manage strategic uncertainty*. Quorum Books Westport, CT, 1995.
- [37] AJ Goldman. “Minimax location of a facility in a network”. In: *Transportation Science* 6.4 (1972), pp. 407–418.
- [38] Army Study Guide. *Request Army Air Medevac*. Accessed April 9, 2014. URL: [http://www.armystudyguide.com/content/Leadersbook\\_information/Combat\\_Leaders\\_Guide/medical-request-army-air--2.shtml](http://www.armystudyguide.com/content/Leadersbook_information/Combat_Leaders_Guide/medical-request-army-air--2.shtml).
- [39] S Louis Hakimi. “Optimum locations of switching centers and the absolute centers and medians of a graph”. In: *Operations Research* 12.3 (1964), pp. 450–459.
- [40] S Louis Hakimi, Edward F Schmeichel, and JG Pierce. “On  $p$ -centers in networks”. In: *Transportation Science* 12.1 (1978), pp. 1–15.
- [41] SL Hakimi. “Optimum distribution of switching centers in a communication network and some related graph theoretic problems”. In: *Operations Research* 13.3 (1965), pp. 462–475.
- [42] Gabriel Y Handler. “Minimax location of a facility in an undirected tree graph”. In: *Transportation Science* 7.3 (1973), pp. 287–293.
- [43] Gabriel Y Handler. *Minimax network location: theory and algorithms*. Tech. rep. Cambridge, Mass.: Massachusetts Institute of Technology, Flight Transportation Laboratory,[1974], 1974.
- [44] R. A. Higgins. “MEDEVAC Critical Care Transport From the Battlefield”. In: *AACN Advanced Critical Care* 21.3 (2010), pp. 288–297.

- [45] E. Ignall, G. Carter, and K. Rider. “An algorithm for the initial dispatch of fire companies”. In: *Management Science* 28.4 (1982), pp. 366–372.
- [46] J. P. Jarvis. *Optimization in stochastic systems with distinguishable servers*. Technical Report TR-19-75. Cambridge, MA: MIT, 1975.
- [47] J. P. Jarvis. “Approximating the equilibrium behavior of multi-server loss systems”. In: *Management Science* 31.2 (1985), pp. 235–239.
- [48] Kurt Jornsten and Mette Bjorndal. “Dynamic location under uncertainty”. In: *Studies in Regional and Urban Planning* 3 (1994), pp. 163–184.
- [49] John H. Vanston Jr., W.Parker Frisbie, Sally Cook Lopreato, and Dudley L. Boston Jr. “Alternate scenario planning”. In: *Technological Forecasting and Social Change* 10.2 (1977), pp. 159–180.
- [50] Oded Kariv and S Louis Hakimi. “An algorithmic approach to network location problems I: The  $p$ -centers”. In: *SIAM Journal on Applied Mathematics* 37.3 (1979), pp. 513–538.
- [51] R. C. Larson. “A hypercube queuing model for facility location and redistricting in urban emergency services”. In: *Computers & Operations Research* 1.1 (1974), pp. 67–95.
- [52] R. C. Larson. “Approximating the performance of urban emergency service systems”. In: *Operations Research* 23.5 (1975), pp. 845–868.
- [53] R. C. Larson and M. A. McKnew. “Police patrol-initiated activities within a systems queuing model”. In: *Management Science* 28.7 (1982), pp. 759–774.
- [54] S. Lee. “The role of preparedness in ambulance dispatching”. In: *Journal of the Operational Research Society* 62.10 (2011), pp. 1888–1897.

- [55] F. Majzoubi, L. Bai, and S. S. Heragu. “An optimization approach for dispatching and relocating EMS vehicles”. In: *IIE Transactions on Healthcare Systems Engineering* 2:3 (2012), pp. 211–223.
- [56] M. B. Mandell. “Covering models for two-tiered emergency medical service systems”. In: *Location Science* 6 (1998), pp. 355 –368.
- [57] Alan S Manne. “Capacity expansion and probabilistic growth”. In: *Econometrica: Journal of the Econometric Society* (1961), pp. 632–649.
- [58] V. Marianov and C. ReVelle. “A probabilistic fire-protection siting model with joint vehicle reliability requirements”. In: *Papers in Regional Science* 71 (1992), pp. 217 –241.
- [59] V. Marianov and C. ReVelle. “The queueing maximal availability location problem: a model for the siting of emergency vehicles”. In: *European Journal of Operational Research* 93 (1996), pp. 110 –120.
- [60] V. Marianov and D. Serra. “Hierarchical location-allocation models for congested systems”. In: *European Journal of Operational Research* 135.1 (2001), pp. 195 –208.
- [61] L. A. McLay. “A Maximum Expected Covering Location Model with Two Types of Servers”. In: *IIE Transactions* 41.8 (2009), pp. 730 –741.
- [62] L. A. McLay. “Emergency Medical Service Systems that Improve Patient Survivability”. In: *Encyclopedia of Operations Research in the area of “Applications with Societal Impact”*. Ed. by P. Keskinocak. Hoboken, NJ: published online, 2011, p. 1.
- [63] Laura A McLay and Maria E Mayorga. “A model for optimally dispatching ambulances to emergency calls with classification errors in patient priorities”. In: *IIE Transactions* 45.1 (2013), pp. 1–24.

- [64] 2013 MEDEVAC Summit. *Delivering a swift and effective future medical evacuation capability*. Accessed February 13, 2014. URL: <http://www.medevacevent.com/default.aspx>.
- [65] U.S. Army Medical Department Office of Medical History. *Dustoff: Army Aeromedical Evacuation in Vietnam*. Accessed April 9, 2014. URL: [history.amedd.army.mil/booksdocs/vietnam/dustoff/chapter1.html](http://history.amedd.army.mil/booksdocs/vietnam/dustoff/chapter1.html).
- [66] Edward Minieka. “The centers and medians of a graph”. In: *Operations Research* 25.4 (1977), pp. 641–650.
- [67] Pitu B Mirchandani. “Locational decisions on stochastic networks”. In: *Geographical Analysis* 12.2 (1980), pp. 172–183.
- [68] Pitu B Mirchandani and Amedeo R Odoni. “Locations of medians on stochastic networks”. In: *Transportation Science* 13.2 (1979), pp. 85–97.
- [69] Nenad Mladenović, Jack Brimberg, Pierre Hansen, and José A Moreno-Pérez. “The  $p$ -median problem: A survey of metaheuristic approaches”. In: *European Journal of Operational Research* 179.3 (2007), pp. 927–939.
- [70] The Aviation History Online Museum. *Bell UH-1 Iroquois*. Accessed April 21, 2014. URL: <http://www.aviation-history.com/bell/uh1.htm>.
- [71] E. Pahon. *Best Soldiers For The Worst Days: Medevac Crews in Afghanistan Save Lives Day, Night*. Accessed February 13, 2014. URL: <http://www.army.mil/article/83749/>.
- [72] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York, NY: John Wiley & Sons, Inc., 1994.
- [73] Josh Reese. “Solution methods for the  $p$ -median problem: An annotated bibliography”. In: *Networks* 48.3 (2006), pp. 125–142.



- [74] C. ReVelle and K. Hogan. “A reliability-constrained siting model with local estimates of busy fractions”. In: *Environment and Planning* 15 (1988), pp. 143 –152.
- [75] C. ReVelle and K. Hogan. “The maximum availability location problem”. In: *Transportation Science* 23.3 (1989), pp. 192 –200.
- [76] C. ReVelle and K. Hogan. “The maximum reliability location problem and  $\alpha$ -reliable  $p$ -center problem: derivatives of the probabilistic location set covering problem”. In: *Annals of Operations Research* 18 (1989), pp. 155 –173.
- [77] C. ReVelle and V. Marianov. “A probabilistic FLEET model with individual vehicle reliability requirements”. In: *European Journal of Operational Research* 53 (1991), pp. 93 –105.
- [78] David A Schilling. “Strategic facility planning: The analysis of options”. In: *Decision Sciences* 13.1 (1982), pp. 1–14.
- [79] Daniel Serra and Vladimir Marianov. “The  $p$ -median problem in a changing network: the case of Barcelona”. In: *Location Science* 6.1 (1998), pp. 383–394.
- [80] ES Sheppard. “A conceptual framework for dynamic location-allocation analysis”. In: *Environment and Planning A* 6.5 (1974), pp. 547–564.
- [81] P. D. Shinkman. *Afghan Medevacs Run to Gunfire, Explosions to Rescue Injured Comrades*. Accessed February 13, 2014. URL: <http://www.usnews.com/news/articles/2013/08/08/afghan-medevacs-run-to-gunfire-explosions-to-rescue-injured-comrades>.
- [82] M. Shunko, S. Cho, and A. Scheller-Wolf. *Impact of Delay Signals on Ambulance Traffic Operations*. Cluster on Applied Probability, INFORMS meeting. October 2011.

- [83] A. J. Swersey. “A Markovian decision model for deciding how many fire companies to dispatch”. In: *Management Science* 28.4 (1982), pp. 352–365.
- [84] Barbaros C Tansel, Richard L Francis, and Timothy J Lowe. “State of the Art – Location on Networks: A Survey. Part I: The p-Center and p-Median Problems”. In: *Management Science* 29.4 (1983), pp. 482–497.
- [85] The Official Web Site of National Museum of the USAF. *Fact sheet for SIKORSKY R-4B HOVERFLY*. Accessed April 21, 2014. URL: <http://www.nationalmuseum.af.mil/factsheets/factsheet.asp?id=518>.
- [86] Jerry R Weaver and Richard L Church. “Computational procedures for location problems on stochastic networks”. In: *Transportation Science* 17.2 (1983), pp. 168–180.
- [87] John F. Zeto, Wade Yamada, and Gale Collins. *Optimizing the Emplacement of Scarce Resources to Maximize the Expected Coverage of a Geographically Variant Demand Function*. Tech. rep. Ft. Belvoir, VA: U.S. Center for Army Analysis, 2006.
- [88] Daniel J. Zinder. *Combat Trauma-Placing Surgeons to save the most lives*. Tech. Report. Newport, R.I.: Naval War College, 2007.

## Vita

Benjamin Charles Grannan was born on September 7, 1984 in Orlando, Florida. In 2006, Ben earned his Bachelor of Science degree in Mathematics with a minor in Secondary Education from Lynchburg College in Lynchburg, Virginia. In 2009, Ben earned his Master of Science degree in Mathematics from Virginia Commonwealth University in Richmond, Virginia.

From 2009 to 2010, Ben worked as an IT analyst at the Supreme Court of Virginia. Fortunately, a scheduling optimization project for the Virginia Court of Appeals introduced Ben to the world of operations research. Ben started the Systems Modeling & Analysis Ph.D. program in August of 2010.

Outside of graduate school Ben enjoys watching college basketball, exploring craft beer, gardening, and working out with SEAL Team Physical Training. In May of 2011, Ben married Whitney Malone at their favorite location in Kitty Hawk, North Carolina. In July of 2014, Ben and Whitney are expecting their first child: a girl named Emery.

During the summer of 2014, Ben, Whitney, Emery, and their dog Ace are relocating to Lexington, Virginia. Ben will join the Department of Economics and Business at the Virginia Military Institute as an Assistant Professor of Operations Management. The Grannan family is looking forward to the next chapter of life in Lexington.