OPTIMIZATION OF TRANSPORTATION REQUIREMENTS IN THE DEPLOYMENT OF MILITARY UNITS

A THESIS
SUBMITTED TO THE DEPARTMENT OF
INDUSTRIAL ENGINEERING
AND THE INSTITUTE OF ENGINEERING AND SCIENCE
OF BİLKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By

İbrahim AKGÜN

December 2005

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Barbaros Ç. Tansel (Supervisor)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. İhsan Sabuncuoğlu

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Assoc. Prof. Levent Kandiller

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Assist. Prof. Emre Berk

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Assist. Prof. Bahar Yetiş Kara

Approved for the Institute of Engineering and Sciences:

Prof. Mehmet Baray
Director of Institute of Engineering and Sciences

ABSTRACT

OPTIMIZATION OF TRANSPORTATION REQUIREMENTS IN THE DEPLOYMENT OF MILITARY UNITS

İbrahim Akgün
Ph.D. in Industrial Engineering
Supervisor: Prof. Barbaros Ç. Tansel
December 2005

We study the deployment planning problem (DPP) that may roughly be defined as the problem of the planning of the physical movement of military units, stationed at geographically dispersed locations, from their home bases to their designated destinations while obeying constraints on scheduling and routing issues as well as on the availability and use of various types of transportation assets that operate on a multimodal transportation network. The DPP is a large-scale real-world problem for which no analytical models are existent. In this study, we define the problem in detail and analyze it with respect to the academic literature. We propose three mixed integer programming models with the objectives of *cost*, *lateness* (the difference between the arrival time of a unit and its earliest allowable arrival time at its destination), and *tardiness* (the difference between the arrival time of a unit and its latest arrival time at its destination) minimization to solve the problem. The cost-minimization model minimizes total transportation cost of a deployment and is of use for investment decisions in transportation resources during peacetime and for deployment planning in cases where the operation is not imminent and there is enough time to do deliberate planning that takes costs into account. The lateness and tardiness minimization models are of min-max type and are of use when quick deployment is of utmost concern. The lateness minimization model is for cases when the given fleet of transportation assets is sufficient to deploy units within their allowable time windows and the tardiness minimization model is for cases when the given fleet is not sufficient. We propose a solution methodology for solving all three models. The solution methodology involves an effective use of relaxation and restriction that significantly speeds up a CPLEX-based branch-and-bound. The solution times for intermediate sized problems are around one hour at maximum for cost and lateness minimization models and around two hours for the tardiness minimization model. Producing a suboptimal feasible solution based on trial and error methods for a problem of the same size takes about a week in the current practice in the Turkish Armed Forces. We also propose a heuristic that is essentially based on solving the models incrementally rather than at one step. Computational results show that the heuristic can be used to find good feasible solutions for the models. We conclude the study with comments on how to use the models in the real-world.

Keywords: large-scale optimization; military; transportation; mixed integer programming; min-max; deployment; mobility; restriction and relaxation; branch and bound.

ÖZET

ASKERİ BİRLİKLERİN İNTİKALİNDE ULAŞTIRMA İHTİYAÇLARININ OPTİMİZASYONU

İbrahim Akgün
Endüstri Mühendisliği Bölümü Doktora
Tez Yöneticisi: Prof. Barbaros Ç. Tansel
Aralık 2005

Bu tezde, çok modlu bir ulaştırma ağı üzerinde işletilen farklı tipteki ulaştırma araçlarının çizelgeleme ve rotalama ile kullanım ve hazır bulunma hususlarına ilişkin kısıtları dikkate alarak, farklı coğrafi bölgelerde konuşlu bulunan askeri birliklerin, konuş yerlerinden kendilerine tahsis edilen görev bölgelerine fiziksel hareketlerinin planlanması olarak tanımlanabilecek İntikal Planlama Problemi (İPP) incelenmiştir. Büyük çaplı gerçek bir dünya problemi olan İPP için mevcut literatürde şu ana kadar analitik bir model geliştirilmemiştir. Bu çalışmada, problem detaylı olarak tanımlanmış ve akademik literature göre analiz edilmiştir. Problemin çözümü için, *maliyet*, *en erken varış zamanından sonraki gecikme* (bir birliğin görev bölgesine gerçek varış zamanı ile müsaade edilen en erken varış zamanı arasındaki fark) ve *en geç varış zamanından sonraki gecikme* (bir birliğin görev bölgesine gerçek varış zamanı ile müsaade edilen en geç varış zamanı arasındaki fark) minimizasyonunu hedefleyen üç karışık tamsayı programlama modeli önerilmiştir. Maliyet minimizasyonu modeli, bir intikalin toplam ulaştırma maliyetini minimize eder. Model, barış zamanında ulaştırma kaynaklarına yapılacak yatırım kararlarının tespitinde ve operasyonun kısa zamanda gerçekleşmesinin beklenmediği, maliyetleri dikkate alacak detaylı bir planlama yapmak için yeterli zamanın olduğu durumlarda intikal planlarının hazırlanmasında kullanılır. En erken ve en geç varış zamanından sonraki gecikmeyi hedefleyen modeller, en büyüğün en küçüklenmesi (minimax) tipinde olup hızlı intikalin çok önemli olduğu durumlarda kullanılabilir. En erken varış zamanından sonraki gecikme minimizasyonu modeli, birliklerin müsaade edilen zaman sınırları içinde intikali için ulaştırma araçları filosunun yeterli olduğu, en geç varış zamanından sonraki gecikme minimizasyonu modeli ise araç filosunun yeterli olmadığı durumlarda kullanılabilir. Her üç modeli çözmek için bir çözüm metodolojisi geliştirilmiştir. Çözüm metodolojisi, CPLEX tabanlı dal-sınır yöntemi uygulamasını önemli oranda hızlandıran gevşetme ve sınırlamanın etkin kullanımını içerir. Orta büyüklükteki problemlerin maksimum çözüm zamanları, maliyet ve en erken varış zamanından sonraki gecikme minimizasyonu modelleri için bir saat, en geç varış zamanından sonraki gecikme minimizasyonu modeli için ise iki saat civarındadır. Deneme yanılmaya dayalı Türk Silahlı Kuvvetleri'ndeki mevcut uygulamada, aynı çaptaki bir problem için optimal olmayan bir çözüm üretmek ortalama bir hafta almaktadır. Çalışmada, ayrıca, modellerin tek bir adımda değil, artımsal olarak çözülmesine dayalı bir sezgisel yöntem önerilmiştir. Hesaplama sonuçları, sezgisel yöntemin modeller için olurlu çözümler bulmak için kullanılabileceğini göstermiştir. Çalışmanın sonunda, modellerin gerçek hayatta nasıl kullanılabileceğine ilişkin yorumlar yer almıştır.

Anahtar sözcükler: büyük-ölçekli optimizasyon; askeriye; ulaştırma; karışık tamsayılı programlama, en küçük-en büyük; intikal; sınırlama ve gevşetme; dal ve sınır yöntemi.

ACKNOWLEDGEMENTS

I thank my advisor Prof. Barbaros Tansel for his guidance, expertise, and patience throughout this dissertation research. With his support, this study has been an enjoyable and invaluable learning experience for me. It has really been an honor to work with this consummate professional.

I am indebted to members of my dissertation committee, Prof. İhsan Sabuncuoğlu, Assoc.Prof. Levent Kandiller, Assist.Prof. Emre Berk, and Assist.Prof. Bahar Yetiş Kara, for showing keen interest in the subject matter and for accepting to read and review this thesis. Their remarks and recommendations have been invaluable.

I also would like to express my thanks to my colleagues in the Scientific Decision Support Center and Logistics Department in the Turkish General Staff Headquarters who have supported me in various ways during this study. I especially thank Colonel Ziya İpekkan and Dr. Liteunant Colonel Altan Özkil for their everlasting support. I am also grateful to Captain Ziya Yıldırım for his support and encouragements.

Finally, I thank my mother Halime for her prays and encouragements, my lovely wife Aysun for her love, support, understanding, and patience, and my son Batuhan for his love and patience. Without them, this study would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ADAMS	Allied Deployment and Mobility System
CBO	congressional budget office
CMDPM	cost-minimization deployment planning model
CMDPM-REL	relaxation of cost-minimization deployment planning model
CMDPM-RES	restriction of cost-minimization deployment planning
DNFP	dynamic network flow problem
DPP	deployment planning problem
DRAP	dynamic resource allocation problem
Lin-LMDPM	linearized lateness-minimization deployment planning model
Lin-LMDPM-REL	linearized relaxation of lateness-minimization deployment planning model
Lin-LMDPM-RES	linearized restriction of lateness-minimization deployment planning model
Lin-TMDPM	linearized tardiness-minimization deployment planning model
Lin-TMDPM-REL	linearized relaxation of tardiness-minimization deployment planning model
Lin-TMDPM-RES	linearized restriction of tardiness-minimization deployment planning model
LMDPM	lateness-minimization deployment planning model
LMDPM-REL	relaxation of lateness-minimization deployment planning model
LMDPM-RES	restriction of lateness-minimization deployment planning model
LTL	less-than truckload trucking
MAP	mobility analysis problem

NATO	North Atlantic Treaty Organization
NDP	network design problem
NRMO	NPS/RAND Mobility Optimizer
PMDPM	priority maximization deployment planning model
PnMDPM	penalty minimization deployment planning model
POE	port of embarkation
POD	port of debarkation model
SCM	supply chain management
TL	truckload trucking
TMDPM	tardiness-minimization deployment planning model
TMDPM-REL	relaxation of tardiness-minimization deployment planning model
TMDPM-RES	restriction of tardiness-minimization deployment planning model
TPFDD	time-phased force deployment data
VRP	vehicle routing problem
VRPTW	vehicle routing problem with time windows
VRSP	vehicle routing and scheduling problem

CHAPTER 1

INTRODUCTION

In this dissertation, we study the deployment planning problem (DPP) that may roughly be defined as the problem of the planning of the physical movement of military units, including their troops, weapon systems, vehicles, equipment, and supplies, stationed at geographically dispersed locations, from their home bases to their designated destinations while obeying constraints on scheduling and routing issues as well as on the availability and use of various types of transportation assets that operate on a multimodal transportation network. Large-scale applications arise in moving military forces at a time of conflict, threat, or crisis. Similar planning needs may also arise for planning the movement of emergency response teams, together with their equipment and supplies, at a time of natural disaster.

In this chapter, we define the motivation behind our work, describe the problem in detail, and give an outline of the dissertation together with our contribution to the literature.

1.1. Motivation

The new threat perceptions of the countries since the end of the Cold War have mandated changes in the military strategy and hence in the structure of the armed forces of almost all countries. The strategy of massing up large numbers of troops, weapon systems, equipment, and supplies in regions where an attack is anticipated has been replaced by a new strategy that envisions having smaller but more mobile forces stationed at widely dispersed locations with the capability to deploy (move, transport) troops, weapon systems, equipment, and supplies rapidly to contingency regions at the time they are needed.

This strategy requires heavy investment in acquisition of cargo planes and sealift ships as well as maintaining a well-sized fleet of reliable ground transportation assets. For instance, the US has made plans to spend close to $20 billion dollars from 1998 to 2002, which constitutes 7 percent of proposed military procurement spending over the period, to acquire new strategic (intercontinental) cargo planes and sealift ships (CBO, 1997). This, combined with the fact that we are in an era of intense competition for funding, requires that investment plans for mobility be based on real transportation requirements and not on "gut feel" or "traditional" predictions. This requires that transportation planners use tools based on scientific methods and be capable of creating implementable, effective, and efficient deployment and sustainment (the provision of personnel, equipment, supplies, and other logistics support to the units deployed to the battlefield) plans in a short time and answering what-if questions to predict the number and types of transportation assets needed to support deployment and sustainment operations.

While having to spend large sums of money for acquisition and maintenance of a well-composed pool of transportation resources is a necessary condition for effective deployment, availability alone does not guarantee smooth operation unless supplemented with carefully worked-out deployment plans. In this regard, models and other tools developed for the analysis of deployment operations not only will help with the evaluation and assessment of investment decisions in transportation resources, but also with the planning and execution of cost-effective deployment and sustainment operations that may arise on short notice at a time of threat.

Literature review shows that there are several deployment/mobility analysis models. However, the models are generally simulation based and that the existing simulation and optimization based studies address only certain parts of the problem. McKinzie and Barnes (2003) review existing models and state that the major aspect lacking in the models is the use of advanced optimization techniques for estimating force closure, i.e., the arrival of units at their areas of operations, and that cumbersome, ineffective classical optimization algorithms or simplistic and ineffective greedy approaches are used to find solutions.

Turkish Armed Forces does not have a national deployment and sustainment planning tool. It uses, as all other NATO members except the United States, NATO's Allied Deployment and Mobility System (ADAMS) (Heal and Garnett, 2001) for making both national and international deployment and sustainment plans. ADAMS provides a structured approach to making deployment and sustainment plans; however, it has the same disadvantages pointed out by McKinzie and Barnes (2003). Therefore, Turkish Armed Forces aims to develop optimization and simulation based deployment and sustainment planning tools as part of a capability planning system that is in the works at the Scientific Decision Support Center in the Turkish General Staff Headquarters.

All of these issues motivate us to study the DPP. In the dissertation, we break away from the existing literature and develop an all-encompassing optimization model and its variants all of which can be used to evaluate and assess investment decisions in transportation infrastructure and transportation assets as well as to plan and execute cost-effective deployment operations at different levels of planning. We develop a solution methodology to solve the model and its variants and a heuristic that finds good feasible solutions in a short time. We implement the models under consideration using deployment scenarios obtained from the Turkish Armed Forces.

1.2. Problem Description

The DPP involves many military units stationed at various locations, i.e., their home bases. At a time of crisis, a subset of them, which is determined by the nature and extent of the threat under consideration, is required to move to their specified destinations, i.e., areas of operations. A call for movement is issued for the active set that specifies among other things the earliest times to depart, the earliest and latest times to arrive, and other requirements that must be obeyed during movement.

The deployable military units, generally main battle units such as companies or battalions, are required to develop and submit their deployment plans in compliance with the operational plan. The current practice in Turkey is a bottom-up approach where each unit, starting from the lowest level in the existing military force structure, submits a plan of its own to the next higher level independent of other units. Because plans from subordinating units are conceived independently, conflicts may arise in demanding the usage of the same transportation infrastructure and/or the same transportation assets at the same time. The receiving unit in the hierarchy is expected to resolve these conflicts, readjust plans, and submit the revised plans to the next higher level. The lower level units are notified of any changes that have taken place during the process. In many cases, it is very difficult, if not impossible, to de-conflict submitted plans from subordinating units. It is typically done by manual methods or not done at all. In fact, it is not unusual for plans to move up in the hierarchy to the highest level with no change at all; hence, de-conflicting usually occurs at the highest level. It is a time consuming and tedious activity that may require several rounds of revisions with no guarantee of creating an implementable plan unless demands on the use of common resources at a time of crisis are quite relaxed or nonexistent.

Deployable items that a unit has are pax (personnel/troops) and cargo (weapon systems, equipment, and supplies) that we collectively refer to as *items*. In the deployment-planning context, it is assumed that items that are small enough to be placed in boxes are carried in boxes of certain sizes. For this reason, a planner needs only to deal with well-defined categories of items, e.g., tanks, armored personnel carriers, trucks, and boxes of predefined sizes. Depending on the planning level of detail, items may be aggregated as necessary. For example, items may be given as the total tons of cargo at a high-level planning function and as individual items with specified attributes (such as weight, width, length, and height) at a lower-level planning function.

Because a unit's integrity is of critical importance from a military perspective, it is desired to deploy a unit as a whole. However, this is usually not physically possible or economical. For this reason, a unit is usually deployed in components. Although the number and configuration of components of a unit depend on the doctrine, the nature of the threat, available resources, the unit's size, and other relevant factors, a unit is usually split into three components: an advance party, a pax party, and a cargo party. An advance party consists of a small number of troops and a few cargo items of the unit that arrive at the destination ahead of time to prepare the destination for the arrival of the other two components. A pax party consists of the main body of the troops of the unit while a cargo party consists of the unit's cargo accompanied by a small number of troops. Each deployment component may follow different routes but must be deployed as a whole, e.g., as a convoy. In some cases, where the size of the component, the deploying strategy, and the resources do not allow the movement of a component as a whole, the component may be split into smaller sub-components (e.g., tanks in a cargo party), all of which are required to use the same route collectively or to move in a time-phased manner. A sample splitting of items of a battalion-size unit is given in Figure 1. Notice that a component is not a single entity with a certain size but a mix of different types of items with different quantities.

ITEMS	QUANTITY
Troops	500
Tanks	14
Pax Carrier	45
2.5 Ton Truck	10
5 Ton Truck	10
2 m^3 Box	40

COMPONENTS	ITEMS	QUANTITY
Advance Party	Troops	40
	2.5 Ton Truck	1
	2 m^3 Box	3
Pax Party	Troops	450
Cargo Party	Troops	10
	Tanks	14
	Pax Carrier	45
	2.5 Ton Truck	9
	5 Ton Truck	10
	2 m^3 Box	37

Figure 1. A sample splitting of a unit into deployment components

Certain precedence and/or synchronization relations may be present between components of a unit. For example, the advance party must arrive at the destination before other components of the unit. The pax and cargo parties may arrive at the destination simultaneously or the pax party may arrive before the cargo party. Even a certain time span may be required to pass between the arrivals of components of a unit. Similar relations may also be present between different units.

A deployment planner finds all relevant data regarding deployable units in the operations plans. Main data needed in the planning are where and when units are to be ready, their earliest departure times, their deployment components, what items deployment components of units are comprised of, and precedence/synchronization relations between components of a unit and/or units.

How a unit moves from its home base to its destination depends on the transportation mode selected. Ground transportation, railways, airlift, and shipping lanes are all possibilities. A unit may use one or more of these in succession. If a unit uses a single transportation mode from its origin to its destination, the same set of transportation assets is used during the entire journey. If a combination of different transportation modes is used, then different sets of transportation assets are active on that unit at different time intervals. This requires that the unit's items be transferred from one set of transportation assets to another set at points of connection between different modes. Such points are referred to as *transfer points*. A transfer point is referred to as a *point of debarkation* (POD) for the supplying transport mode and a *point of embarkation* (POE) for the receiving transport mode for the unit. Main transfer points are harbors, airports, and rail stations. Several zones, e.g., staging and marshalling zones, at transfer points prevent congestion and provide uninterrupted flow of items by providing sufficient space in and adjacent to the terminal area to enable deploying and supporting units to carry out loading/unloading, coordination, control, and preparation operations in harmony. A marshalling area can be regarded as a waiting/parking place and a staging area as a service point. In this regard, a capacity may be associated with a transfer point depending on the availability and capacity of material handling equipment and/or its physical characteristics, e.g., a certain number of docks at a seaport. Similar zones may also be operated at home bases and destinations of units.

Even though there is no limit on how many times a unit changes transportation mode, movement pattern of a unit generally includes three movement segments: from home base to a transfer point, from transfer point to another transfer point, and from transfer point to destination. Each such change requires additional planning, coordination, and cooperation activities leading to potential delays and unforeseen problems. For this reason, changes between transport modes are avoided to the extent possible and changes of transportation assets are not allowed while moving in a given mode.

A unit may use transportation assets from three different *sources*: from its own fleet (organic assets), from military transportation units, and/or from civilian transportation companies.

Organic assets in possession of a unit may be used at will by the unit. However, for the usage of transportation assets from the other two sources, a request must be made to a transportation coordination center where all such requests are assessed. In most cases, it is not possible for the center to meet all demands coming from different units due to physical and/or economical limitations. In this regard, the center is expected, if possible, to allocate available transportation resources to demanding units in a time-phased manner in such a way that all units arrive at their destinations at the time they are needed. If this is not possible, the center is expected to make suggestions to carry out the mission successfully, e.g., procuring additional transportation assets. Even a change in the operational plan may be called for.

In meeting the transportation requests of units, the transportation coordination center considers two issues: cost and time. A costing structure is needed to decide how to source the needed transportation assets. In general, a cost structure consisting of a fixed cost and a variable cost is assumed. If a transportation asset is organic or supplied from a military transportation unit, the incurred cost for the transportation asset is generally the variable cost. On the other hand, if a transportation asset belongs to a civilian company, both fixed and variable costs are incurred where the fixed cost is the leasing or procurement cost.

Time is of utmost importance in the deployment-planning context as it is no good if a unit is not ready at its destination on time. In this regard, everything that affects the arrival time of a unit at its destination must be taken into account in the planning. Main factors that affect arrival times are ready times for transportation assets to be available for the first time usage, travel times, and loading, unloading, and idle waiting times at home bases, destinations, or transfer points. A ready time is of critical importance for transportation assets sourced from the civilian sector as companies are contracted to provide transportation assets at specified times and at specified locations. Travel times are dependent on the speed at which transportation assets move. If a unit uses ground transportation, it moves as a convoy and conforms to a pre-specified convoy speed. The same is valid for railway transportation (as a train can also be taken as a convoy). On the other hand, for sea and air transportation, transportation assets move at their regular speeds. Loading and unloading times of transportation assets are known a priori; however, idle waiting times are not known a priori as they are determined by the availability of transportation assets and transportation infrastructure. The availability of a transportation asset is determined by its ready time, travel speed, loading/unloading time, and idle waiting time as it circulates through the network. Availability of the transportation infrastructure is generally related to capacity issues as in the case of transfer points.

In determining how many transportation assets of each type to allocate to a unit, *loadability* features of transportation assets are taken into account. Based on loadability, transportation assets are classified into four groups:

1) Pax: transportation assets that can carry only personnel (e.g., buses) (V_{pax})
2) Cargo: transportation assets that can carry only cargo (e.g., trucks, tank carriers, cargo planes) (V_{cargo})
3) Pax and Cargo: transportation assets that carry cargo and personnel in separate compartments (e.g., ships) (V_{both})
4) Mixed Pax and Cargo: transportation assets that carry cargo and personnel in a single compartment (e.g., trucks, some types of planes) (V_{mix})

For a transportation asset in class V_{pax}, the number of passengers to be carried is determined by the number of seats on the transportation asset while for a transportation asset in class V_{cargo}, the amount of cargo to be loaded is determined by the weight, volume, and/or lanemeter capacities of the transportation asset. Lanemeter capacity (typically expressed in terms of length but not necessarily related to the length of the transportation asset) is similar to the parking capacity of a parking area and generally used while wheeled and/or armored vehicles are loaded onto sealift and/or airlift assets.

For a transportation asset in class V_{both}, the number of passengers and the amount of cargo to transport are determined separately as done for classes V_{pax} and V_{cargo} as passenger and cargo carrying capacities for transportation assets in this class do not interact with each other. Constraints valid for classes V_{pax} and V_{cargo} are also valid for V_{both}.

For a transportation asset in class V_{mix}, the same space is shared by both cargo and personnel and one displaces the other in discrete blocks that can be characterized by a step function. When personnel are to be carried on a transportation asset, seats built in blocks of different sizes are to be installed on it. The number of blocks to be installed is determined by the number of passengers. The portion of the capacity that will be used for cargo is diminished each time a block of seats is added. For example, assuming that seats are built in blocks of 18 seats, one block of seats for 1 through 18 passengers and two blocks of seats for 19 through 36 passengers are installed on the transportation asset. Note that the decrease in capacity is both weight-wise and volume-wise and that the amount of decrease changes depending on the passenger, weight, and volume capacities as well as the number of seats in a block.

One issue that is related both to the movement of items and to the loadability feature is that some cargo are *self-deployable* meaning that they do not need to be carried on a transportation asset on the parts of the transportation network on which they can move by themselves. For example, tanks (generally moved on tank carriers but can self-deploy when necessary) and trucks are self-deployable in ground transportation. A self-deployable can be an organic asset. For example, a truck is self-deployable and an organic asset of a unit. Thus, a self-deployable is treated as a transportation asset on those parts of the network on which it can move by itself. Furthermore, a self-deployable with some cargo on it can be loaded onto another transportation asset, e.g., a truck with boxes may be loaded onto a plane. This requires determining what to load onto a self-deployable item. What is done in practice is to pre-determine the loads of self-deployable items as units are to use their organic assets the first time in transporting some of their items. In this respect, a self-deployable item and what is on it can be regarded as a single entity with a certain weight and volume.

Coupling of a transportation asset and an item based on the loadability features alone is not sufficient. The transportation infrastructure must also support the movement of both transportation assets and items with respect to physical characteristics, e.g., width of a tunnel/dock and strength of a bridge. Thus, analyses such as items-to-routes/locations and transportation assets-to-routes/locations are needed. Such analyses require intensive data and are possible only when supported by a well-organized information system. In addition, prevailing practices based on current policies, strategies, doctrines, and security concerns must be taken account in determining what types of transportation assets and what parts of a transportation network can be used by a unit.

In deployment planning, a planner needs to determine the routes to follow, schedule the movements, and allocate the transportation assets and the transportation infrastructure to the deploying units on a time basis so that all deploying units and their materiel arrive at their destinations at their required times while obeying constraints regarding priorities of the units, availability of resources (transportation assets, transportation infrastructure, material handling equipment, etc.), capacities, and any other specified issues. From a modeling point of view, three main problems are handled simultaneously: routing, scheduling, and resource allocation. Complications arise due to simultaneous handling of two types of flows: those of items and of transportation assets. Transportation assets are the active agents in that they move the items to which they are assigned. They can be repeatedly used for moving different sets of items at different times.

1.3. Outline of the Dissertation and Contribution to the Literature

In Chapter 1, we give a detailed description of the problem. The literature review shows that no model that deals with the DPP as a whole exists. For this reason, there does not exist an academic and detailed description of the problem. Most of the relevant details are found in military field manuals that give information on various aspects of deployment. However, no field manual seems to involve a complete description of the problem either. In this regard, our first contribution is to give a detailed, complete, and academic description of a large-scale, real-world, and complicated problem.

In Chapter 2, we present the literature related to the problem. In compliance with the lack of a complete description of the problem in the literature, there does not exist a scientific analysis of the problem with respect to the scientific literature. For this reason, we do a comprehensive literature review of research areas related to the problem, namely, dynamic network flow, network design, vehicle routing, dynamic resource allocation, and mobility analysis problems. We give a summary of the studies in these areas and explain why the models available in the literature are not able to capture various aspects of the DPP in its entirety. As the DPP is related to the transportation planning problems, we also give an overview of transportation systems in this chapter. In this regard, giving an academic analysis of a complex problem is our contribution in Chapter 2.

In Chapter 3, we first give the abstraction of the problem. Specifically, we define the underlying network, transportation assets, items (commodities) to be moved, and sets and data related to these three. The abstraction is such that it gives a basis for a database. Next, we give the formulation of the first model, a solution methodology to solve the model, and computational results obtained using the solution methodology.

The purpose of the first model, *Cost Minimization Deployment Planning Model (CMDPM)*, is to plan the movements of units with a given fleet of transportation assets such that the sum of fixed and variable transportation costs is minimized. This model may be of use for investment decisions in transportation resources during peacetime and for deployment planning in cases where the operation is not imminent and there is enough time to do deliberate planning taking cost into account.

The solution methodology is based on an effective use of a relaxation and restriction of the model that significantly speeds up a CPLEX-based branch and bound. The solution times for intermediate sized problems are around one hour whereas it takes about a week in the Turkish Armed Forces to produce a suboptimal feasible solution based on trial and error methods.

In Chapter 4, we present two min-max models, *Lateness Minimization Deployment Planning Model (LMDPM)* and *Tardiness Minimization Deployment Planning Model (TMDPM)*. *Lateness* in the LMDPM is defined as the difference between the arrival time of a unit and its earliest allowable arrival time at its destination while *tardiness* in the TMDPM is defined as the difference between the arrival time of a unit and its latest arrival time at its destination. In this regard, the objectives in the LMDPM and TMDPM are to minimize maximum lateness and tardiness, respectively. These models will be of use in cases where quick deployment is of utmost concern. The LMDPM is for cases when the given fleet of transportation assets is sufficient to deploy units within their allowable time windows and the TMDPM is for cases when the given fleet is not sufficient. We solve these models using the test problems and the solution methodology developed in Chapter 3. Solution times for these models are also around one hour for intermediate sized problems.

In Chapter 5, we present a heuristic that involves essentially using the developed models iteratively to obtain quick feasible solutions for the problem.

Our contribution in Chapters 3 through 5 is that we provide manageable and solvable models for a large-scale, real-world problem for which analytical models are nonexistent. In addition, we provide a heuristic algorithm that finds good feasible solutions.

In Chapter 6, we explain how the models can be used with the current practice, bottom-up approach, in the Turkish Armed Forces and with the proposed top-down approach. We next give some what-if questions and explain how they can be answered by using the models and their output. Variations of the models are also given in this chapter. Our contribution in this chapter is to establish the connection between real-world and theoretical work by pointing out how decision making can be improved using the models. In Chapter 7, we summarize the results of our study.

CHAPTER 2

LITERATURE REVIEW

There has recently been a growing interest in the *Supply Chain Management* (SCM) in the business community. This has led to a vast literature both in the theory and practice of the SCM concept. Survey papers by Beamon (1998), Croom, Romano, and Giannakis (2000), Min and Zhou (2003), Slats et al. (1995), Stevens (1998), Tan (2001), and Thomas and Griffin (1996) give an extensive list of studies in this area.

Successful applications of the SCM concept in several business sectors (for example, Arntzen et al. (1995), Cohen et al. (1990), Lee and Billington (1995), and Martin et al. (1993)) have modified the way the military manages its supply chain. The military has adopted business practices to solve some problems it encounters in operating its supply chain during peacetime. However, there are some problems particular to the military supply chain during wartime for which the business SCM theory and practices are insufficient. The DPP is such a problem.

In this chapter, our first goal is to clarify why the business SCM models, specifically transportation planning models, fall short of solving the DPP. To this end, we first establish the relationship between commercial and military supply chains and the DPP and explain why and where the military supply chain is different from the business supply chain. We then review the literature broadly related to the DPP, specifically, the literature on *transportation planning* and *dynamic network flow* problems, and discuss why existing studies in these areas are insufficient for dealing with the level of complexity inherent in the DPP.

The literature that directly addresses the DPP is grouped under the name of *mobility analysis*. Our second goal in this chapter is to give mobility analysis literature and then explain in what ways existing mobility analysis models fall short of dealing with all aspects of the DPP.

The chapter is organized to first place the DPP in the basic context of transportation planning which we view as a subset of the general SCM literature. To this end, the discussion of the directly related literature on the DPP, collectively referred to as the Mobility Analysis Problem (MAP), is deferred to the end of the chapter (Section 2.3.3). The reader interested in the mobility analysis literature may skip to Section 2.3.3 without loss of continuity, but perhaps with some loss of perspective on where the DPP fits in the more general realm of transportation planning and of SCM.

Although there is a literature regarding *SCM deployment*, e.g., Shapiro (2003), this term refers to *inventory deployment* where the modeling focuses on closing/opening plants/distribution centers and determining inventory levels for open plants or distribution centers. In this regard, the DPP and the SCM deployment problem are structurally quite different despite the use of the term "deployment" in both.

2.1. Military and Commercial Supply Chains and Deployment Planning

The main goal of both commercial and military supply chain systems is to ensure that the right commodity is available at the right location, at the right time, and in the right quantity. However, the consequences of not achieving this goal are different for the commercial sector and the military. If this goal is not met in a commercial supply chain system, the cost is essentially a profit loss. On the other hand, if this goal is not met in a military supply chain system, then the cost is human life that results from the failure of a mission, sometimes with catastrophic results. In this regard, no monetary value can justfully be attributed to the success or failure of a military supply chain system and the importance of correct and timely planning in a military context cannot be overemphasized.

Kress (2005) divides the military supply chain into peacetime and wartime supply chains. He points out that peacetime supply chain is similar to a business supply chain and that the military adopts best business supply chain practices to manage its system during peacetime. Thus, planning and operating a military supply chain during peacetime is similar to planning and operating a business supply chain. This similarity allows analytical planning tools available for business supply chains to be used for peacetime supply chain.

Wartime supply chain refers to a supply chain whose malfunction may be disastrous. It has three components: deployment, sustainment, and redeployment of military units. Deployment is simply the physical movement of military units (including troops, equipment and supplies) from their home bases to their areas of operations. During a deployment, units carry their organic equipment and a basic load of supplies so as to be capable of engaging in a confrontation with the enemy or carrying out an operation for a designated length of time (e.g., three days) without relying on external support. However, at the end of the designated time, deployed units must get enough and timely sustainment for subsequent effectiveness in an operation. The sustainment refers here to the provision of personnel, logistics, and other support. Redeployment is essentially a deployment to peacetime locations or to another operations area. In this regard, deployment and sustainment planning are of key concern in a wartime supply chain.

Kress (2005) argues that wartime supply chain is different from a business/peacetime supply chain. He points out the following discrepancies: The operations are generally routine, long-term, and small-scale in the peacetime while the operations are rare, short-term, and (extra) large-scale in the wartime. As a result of this, the flow through the network is sparse, e.g., single trucks, in the peacetime while it is massive, e.g., convoys of trucks, in the wartime. The operating network in the wartime changes depending on the movements of units in the operations areas while it is stationary in the peacetime. In the peacetime, there are uncertainties in the demands, costs, and lead times. In the wartime, the operations are carried out in a hostile environment and hence there are uncertainties in the survivability and success of the operations in addition to uncertainties that are prevailing in the peacetime. In the peacetime, economical solutions are preferable and hence cost is the main planning consideration. Planners have a chance to choose which demand to meet, to meet a demand at a later time, or not to meet a demand at all. In the wartime, however, operational success is the main planning consideration and cost is of secondary concern. Planners do not have a chance to choose which demand to meet as any unmet demand may cause failure of a mission. As a result of these issues, the modeling approach is microscopic and service level measures are relaxed in the peacetime. In the wartime, on the other hand, the modeling approach is macroscopic and service level measures are strict.

The aforementioned discrepancies between the two chains do not allow analytical planning tools developed for peacetime supply chain to be used for wartime supply chain; specialized analytical tools are needed for wartime supply chain.

Analytical tools are needed for both peacetime and wartime decisions regarding wartime supply chain. Peacetime decisions regarding wartime supply chain are essentially strategic decisions, e.g., national supply levels or transportation capabilities determined as a function of threat and national capabilities. These decisions are actually related to investment decisions. Wartime decisions regarding wartime supply chain are operational, e.g., theater-level deployment and employment, and tactical, e.g., combat unit's logistics support.

In this dissertation, we develop optimization models that are of use for both wartime and peacetime decisions regarding the deployment planning of military units.

2.2. A Classification of Transportation Systems

Crainic (2003) classifies transportation systems into customized/door-to-door transportation and consolidation transportation.

2.2.1. Customized/Door-to-Door Transportation

In *customized/door-to-door transportation*, transportation services are tailored to the specific needs of the customer. *Truckload Trucking* (*TL*) is one example of door-to-door transportation. It arises in distributing goods over long distances. In the TL, a truck is usually dedicated to each customer. When the customer calls, a truck with a driver or a driving team is assigned to it. The truck is moved to the customer-designated location, and it is loaded. It then moves to the specified destination. At destination, the truck is unloaded, and the driver calls the dispatcher to give its position and requests a new assignment. The dispatcher may indicate a new load, ask the driver to move empty to a new location where demand should appear in the near future, or have the driver wait and call later.

In this regard, the truckload carrier operates in a highly dynamic environment. There is little information regarding future demands, travel times, waiting delays at customer locations, precise positions of loaded and empty vehicles at later moments in time. Thus, there is certainly a need to respond to customer requests in a timely fashion and predict the effects of today's decisions on future decisions.

2.2.2. Consolidation/Service Transportation

In *consolidation* transportation, demands of several customers are served simultaneously by using the same vehicle or convoy. Transportation services are not tailored to specific needs of the customers. Regular transportation services are established with certain operating characteristics, e.g., routes and schedules, to satisfy the expectations of the largest number of customers. For example, origin, destination, intermediary stops, departure time from origin, arrival time at destination, departure/arrival times from/at intermediary stops, capacity, and speed of a container ship moving from port A to port B are determined and proposed to the customers. Less-than-truckload trucking companies, railways, shipping lines, and postal and express shipment services may offer this type of transportation. Consolidation transportation is characterized by the existence of terminals where cargo and vehicles are consolidated, grouped, or simply moved from one service to another.

The operating infrastructure in consolidation transportation consists of a rather complex network of terminals connected by physical or conceptual links. Air and sea lines correspond to the latter while road, highways, and rail tracks are examples of the former.

In consolidation transportation, a transportation demand is defined between given points of the transportation network, i.e., origin and destination, together with commodity-related physical characteristics, e.g., weight and volume. Particular service requirements, e.g., delivery conditions and type of vehicle, may also be requested. The transportation service provider moves commodity/freight by a large number of vehicles: rail cars, trailers, containers, ships, etc. Vehicles usually move on pre-specified routes and sometimes follow a given schedule. Vehicles may move individually or in convoys such as rail or barge trains. Convoys are formed and dismantled at terminals. Other operations at terminals include freight sorting and consolidation, its loading onto or unloading from vehicles as well as vehicle sorting, grouping, and transferring from one convoy to another.

Terminals may be in different designs and sizes and specialized for certain operations. Major consolidation centers/terminals are referred to as *hubs*. The hubs are linked by high frequency and capacity services, e.g., planes and ships. There are also terminals where freight and vehicles are consolidated at the beginning and end of freight's journey. These terminals are linked to hubs by *feeder services*, i.e., *spoke* links. It is possible that a terminal be linked to more than one hub. *Local delivery and pick-up* operations are usually arranged by these terminals.

To clarify the notion of consolidation transportation, we now focus on specific transportation modes.

A railway transportation system is composed of single and/or double track lines that link many large and small classification yards, where rail cars are grouped and trains are formed, pick-up and delivery stations, junction points, etc. The process begins when the customer issues an order for a number of empty cars or when freight is brought into the loading facility following a pick-up operation. At the appropriate yard, rail cars are selected, inspected, and then delivered to the loading point. Once loaded, cars are moved to the origin yard (possibly the same) where they are sorted, classified, and assembled into *blocks*. A block is a group of cars, with possibly different final destinations, arbitrarily considered as a single unit for handling purposes from the yard where it is put together to its destination yard where its component cars are separated. Rail companies use blocks to take advantage of some of the economies of scale related to full train loads and the handling of longer car strings in yards. The block is eventually put on a train and this signals the beginning of the journey. During the long-haul (intercity) part of the journey, the train may overtake other trains or may be overtaken by trains with different speeds and priorities. When the train travels on single-track lines, it may also meet trains traveling in the opposite direction. Then, the train with the lowest priority has to give way and wait on a side track for the train with the higher priority to pass by. At yards where train stops, cars and engines are regularly inspected. Also, blocks of cars may be transferred. When a block finally arrives at destination, it is separated from the train, its cars are sorted, and those having reached their final destination are directed to unloading station. Once empty, the cars are prepared for a new assignment, which may be either a loaded trip or an empty repositioning movement.

Similar to rail transportation, Less-than-truckload-trucking (LTL) networks may encompass different types of terminals. Local traffic is picked up by small trucks and is delivered to *end-of-line terminals* where it is consolidated into larger shipments before long-haul movements. Symmetrically, loads from other parts of the network may arrive at end-of-lines to be unloaded and moved into delivery trucks for final delivery. *Breakbulks* are terminals where traffic from many end-of-line terminals is unloaded, sorted, and consolidated for the next portion of the journey. Breakbulks are the hubs of LTL networks as major yards are the hubs of rail transportation systems.

LTL transportation follows the same basic operational structure described for rail but on a simpler scale and with significantly more flexibility. In addition, a truck is only formed of a tractor and one or several trailers (when more than one trailer is used, these are smaller and are called *pups*). Consequently, terminal operations are generally simpler; freight is handled to consolidate outbound movements but there are no convoy-related operations. However, LTL transportation may become rather complex when the option to use rail (the *trailer-on-flat-car – TOFC*) for long distances.

Intermodal container transportation may be viewed as either customized/door-to-door or consolidation transportation. For the customer, it is door-to-door transportation. When requested, containers are delivered, loaded, moved through a series of terminals and vehicles (of which the customer has little knowledge even when the exact position of the shipment is available), and are delivered to final destination where the goods are unloaded. For the service provider, i.e., shipping company, it is a consolidation transportation system. Containers from many customers must be moved to a port by truck, barge, or rail, or a combination of these. At the port, containers are grouped and loaded on a ship. The ship follows a prespecified route and a tight schedule and delivers the containers at the destination port. From the destination port, a land transportation system delivers the containers to the final destination by using a variety of modes and terminals. A hub system may be operated between major ports. Container transportation systems that operate solely on land may also be encountered. In this case, rail trains and inland terminals usually play the role of ships and ports.

A similar argument may be made for *express letter* and *small package services*. For customers, it is obviously a door-to-door transportation. For the company, it is a consolidation transportation system that usually makes use of various air, truck, and rail services. The company implements a *Vehicle Routing Problem*-type of service to interact with its customers and collect and distribute letters and packages. The collection and distribution centers where mail is sorted and consolidated play a role similar to that of end-of-line terminal in LTL transportation. To reach its destination, a letter or package usually passes through at least one major hub. These terminals are similar to breakbulks in LTL. To link its national hubs and major collection and distribution centers, the company may operate its own planes or may use scheduled passenger flights or train services. When distances are not long, trucks may also be used.

It is useful to differentiate between moving people and freight. The above classification is essentially for freight transportation; however, the aforementioned operating characteristics can be applied to moving people. For example, selecting some airport terminals as hubs is also an operating strategy in operating airlines. What is important in the context of moving people is that airlines, passenger trains and bus companies typically run fixed schedules over fixed routes that are planned months in advance. This allows people to arrange their travel plans around a fixed schedule. Freight transportation, however, is operated in a dynamic environment which may require operational plans to be modified frequently, e.g., on a daily basis. This does not mean that no long-term or medium-term planning is made in freight transportation. For example, freight companies have to determine locations of terminals, which is a long-term and strategic decision.

Planning levels in transportation are classified into *strategic*, *tactical*, and *operational*. *Strategic planning* is concerned with long-term planning. Decisions at the firm/service level include the design of physical network, and, the location of major facilities, e.g., terminals, the acquisition of major resources, e.g., locomotives. Strategic planning at the international, national, and regional levels deals with transportation networks or services of several carriers simultaneously. *Tactical planning* is concerned with medium-term planning. It aims to determine an efficient allocation of resources. Decisions at this level include the *design of the service network* that may consist of the determination of the routes and types of service to operate, service schedules, vehicle and traffic routing, repositioning of the fleet for use in the next planning period. *Operational planning* is concerned with short-term planning. It is made by local management, yard masters, and dispatchers in a highly dynamic environment where the time factor plays an important role and detailed representations of vehicles, facilities, and activities are needed. Decisions at this level include scheduling crews, services, maintenance activities; routing and dispatching of vehicles and crews; dynamic allocation of scarce resources.

Transportation planning problems that are studied in the literature and that fall into one or more of the aforementioned planning levels can be classified into *Network Design*, *Vehicle Routing*, *Driver Assignment*, *Crew Scheduling*, *Dynamic Fleet Management*, *Empty Vehicle Distribution*, and *Intermodal Container Operations Problems*. Of these problems, network design problem is considered as strategic and tactical while others are considered as operational. It is common to consider all problems except network design and vehicle routing problems as *Dynamic Resource Allocation Problems*, e.g., Powell (2002).

All of these problems have been studied extensively in the literature and hence the literature is too rich to discuss all of them. We give in the following a selective review of studies that are notably more important.

2.3. Literature Review

The DPP is related to the transportation planning problems in the literature. However, before giving literature related to these problems, we discuss *dynamic network flow problem* because many transportation planning models turn out to be *dynamic network flow models*. In addition, time component, i.e., dynamic aspect, inherent in the DPP requires using a time-dynamic network in the modeling of the problem.

2.3.1. Dynamic Network Flow Problem (DNFP)

The DPP is related to the *Dynamic Network Flow Problem* (DNFP) because routing occurs over time. Ford and Fulkerson (1958, 1962) generalize standard definition of a network by introducing an element of time, ending up with a dynamic network. The purpose in the DNFP is to model decision problems over a time horizon T. The common characteristics are networks with transit times τ_{ij} and capacities u_{ij} on an arc from node i to node j. The transit time of an arc specifies the amount of time it takes for flow to travel through a particular arc and the capacity of an arc specifies a flow rate entering an arc for each point in time.

The research on DNFP has two main directions with respect to the modeling of time, namely, discrete and continuous-time (Fleischer and Skutella, 2002). In *discrete-time DNFP*, time is discretized into steps of unit length. In each step, flow can be sent from node i to node j through an arc (i, j) where flow arrives at node j τ_{ij} steps later. Ford and Fulkerson (1958, 1962) introduce *time-expanded networks* in which dynamic flows can be described and computed. A time-expanded network contains a copy of the node set of the underlying static network for each discrete time step. Furthermore, for every arc in the static network with integral transit time τ_{ij}, there is a copy between all pairs of time layers with distance τ_{ij} in the time-expanded network. Thus, a discrete dynamic flow in the given network can be interpreted as a static flow in the corresponding time-expanded network. This allows applying optimization techniques developed for static flows in solving dynamic flow problems. However, the drawback of this approach is that the size of the underlying time-expanded network may be enormously large due to the linear dependency of the size of the time-expanded network on the number of time steps.

In the case of *continuous-time DNFP*, the flow on an arc (i,j) is a function $f_{ij}:R^{+}\to R^{+}$. However, there is a strong connection between the two types. Many results and algorithms developed for the discrete time DNFP can be carried over to the continuous-time DNFP. The most commonly used approach is to consider a sequence of discrete-time intervals in which the data is kept constant. Obviously, this approach implies a certain level of error. The smaller the time intervals are, the smaller this error becomes, i.e. the more accurately the model represents the current flow's evolution, but at the expense of blowing up the size of the network.

Aronson (1989) and Powell et al. (1995) give a comprehensive survey of dynamic network flows. Below are some results from the literature.

Maximum Dynamic Flows. In the *Maximum Dynamic Network Flow Problem*, the problem is to send the maximal possible amount of flow from a source node s to a sink node t within time horizon T. Ford and Fulkerson (1958, 1962) show that a solution obtained for a static network flow problem in the given network can efficiently be used to find a dynamic flow by decomposing it into flows on paths. Their method starts to send flow on each path at time zero and repeats at each time period as long as there is enough time left in the time horizon T for the flow to arrive at the sink. A dynamic flow obtained using this structure is called *temporally repeated*.

Earliest Arrival and Latest Departure Flow. In the *Earliest Arrival Flow Problem*, the purpose is to find a single feasible dynamic flow from a source node s to a sink node t within a specified time horizon T that maximizes the total amount of flow reaching the sink by every time step up to and including

T. Gale (1959) proves that such flows always exist but does not develop any algorithms to find such flows. Wilkonson (1971) and Minieka (1979) give pseudo-polynomial time algorithms to compute such flows. Minieka (1979) also studies *the Latest Departure Flow Problem,* in which the purpose is to find a single feasible dynamic flow from a source node s to a sink node t within a specified time horizon T that maximizes the total amount of flow departing from the source after every time step (subject to the constraint that the flow is finished by time T). The flow that occurs when these two types of problems are solved simultaneously is called a *Universally Maximum Dynamic Flow*.

Quickest Flows. In the *Quickest Flow Problem,* the problem is to send a given amount of flow f from a source node s to a sink node t in the shortest possible time. This problem can be solved in polynomial time by incorporating the algorithm of Ford and Fulkerson (1958, 1962) for the maximum dynamic problem. Burkard et al. (1993) develop a faster algorithm that solves the quickest $s-t$ flow problem in strongly polynomial time.

In the *Quickest Path Problem*, a quickest flow that uses only a single path is sought. Chen and Chin (1990), Rosen et al. (1991), and Hung and Chen (1991) show that the problem can be solved in polynomial time.

The *Evacuation Problem* is a multi-source single-sink version of the quickest flow problem. Given a vector of supplies, the problem is to find a feasible dynamic flow that satisfies all supplies in the minimum overall time, if such a flow exists. Berlin (1979) and Chalmet et al. (1982) study this problem as a means of modeling emergency evacuation from the buildings. Jarvis and Ratliff (1982) show that three optimality criteria may be achieved simultaneously: (1) an earliest arrival schedule that maximizes the total flow into the sink by every time step, (2) overall minimization of the time required to evacuate the network, and (3) minimization of the average time for all flow to reach the sink.

Quickest Transshipments. The *quickest transshipment problem* is a multi-source multi-sink version of the quickest flow problem. Given a vector of supplies and demands at the nodes, the purpose is to find a dynamic flow with the minimum possible time horizon that satisfies all supplies and demands. Unlike standard network flows, this multiple-source, multiple sink, single commodity flow over time is not equivalent to an $s-t$ maximum flow over time. Hoppe (1995) and Hoppe and Tardos (2000) describe the first polynomial time algorithm to solve this problem. They use the chain decomposable flows that generalize the class of temporally repeated flows. However, their algorithm is not practical because a submodular function minimization is required for a subroutine.

The quickest transshipment problem is closely related to the *Dynamic Transshipment Problem* in which the goal is to move the appropriate amount of flow through the network within the pre-specified time horizon T, if possible. Hoppe (1995) and Hoppe and Tardos (2000) develop first polynomial time algorithm for this problem as well.

Minimum-Cost Dynamic Flows. The quickest flow problem and dynamic maximum-flow problem can be generalized by defining additional costs on the arcs. The problem may be to find either a quickest flow within a given cost budget or a minimum-cost flow with a given time horizon. Klinz and

Woeginger (1995) prove that the minimum-cost dynamic flow problem is NP-hard even for the special case of series parallel graphs. They also show that the problem of computing a maximum temporally repeated flow with minimum cost is strongly NP-hard.

Multi-Commodity Dynamic Flows. Single commodity dynamic flow problems can be extended to include multiple commodities. However, although there is substantial literature on the static multi-commodity flow problem, there exists hardly any result on dynamic multi-commodity flows. Only recently, Hall, Hippler, and Skutella (2003) prove that *multi-commodity dynamic flow problem without costs and without storage of flows at intermediate nodes* is NP-hard. (Storage of flows at intermediate nodes is an issue that arises in the dynamic network flow setting.) For single commodity flow problems, Fleischer and Skutella (2003) show that storage of flow at intermediate nodes is unnecessary even in the NP-hard setting with costs. However, Fleischer and Skutella (2002) prove that, for the quickest multi-commodity flow problem, there exist cases where the time horizon of an optimal solution increases when storage of flow is prohibited.

In the studies mentioned so far, the assumption is that dynamic networks do not change with time, i.e., edge capacities and transit times are deterministic. However, there may be cases in which a network changes with respect to time-varying characteristics, e.g., a *dynamic (stochastic) dynamic network*. There are also some results for these networks, e.g., Minieka (1974) and Halpern (1979). Minieka (1974) studies the maximum dynamic flow problem where an arc is associated with time intervals when the arc is deleted (added) from (to) the network. Halpern (1979) considers the maximum dynamic flow problem where edge capacities change over time.

Despite the fact that the DPP is defined on a dynamic network where prespecified flow is to be moved between certain origin-destination pairs, it cannot be characterized simply as a DNFP with side constraints. The main difference arises from the fact that the problem structures are different. In the DNFP, a single type of entity, i.e., commodities/items, moves through the network whose arc capacities and arc transit times are given as deterministic or stochastic parameters. In the DPP, however, two types of entities, commodities/items and transportation assets, move through the network. For a flow of items to occur on an arc at a certain time period, the arc must be *activated* at that specific time by allocating transportation assets with sufficient capacity and appropriate loadability characteristics to deliver those items. However, this allocation of transportation assets is also a problem to be solved optimally, e.g., to minimize cost such that units arrive at their destinations at their required times. Thus, arc capacities are determined endogenously, i.e., in the model, and hence are not parameters but variables. Similarly, arc transit times are not predefined parameters as they are determined by the type and speed of transportation assets assigned to arcs at a certain time. In addition to "variable" arc capacities limiting the flow of items, parametric arc capacities limiting the flow of transportation assets and items through a node and/or arc may also be defined.

With this problem structure, the DPP can be regarded, disregarding multimodality, unsplittability, and precedence issues, as a hybrid multicommodity dynamic network flow-vehicle routing problem. (Note that the problem is a multicommodity flow problem even when there is a single unit because it comprises several items.) However, as will be clear in the coming paragraphs, vehicle routing problem in the context of the DPP is different from usual vehicle routing problems. One may argue that an optimal flow for transportation assets (items) is determined and then an optimal flow for items (transportation assets) is obtained given optimal flow of transportation assets (items), i.e., sequential planning may be proposed. This approach may be appropriate in the context of business transportation planning problems where several firms/organizations are responsible for managing several parts of the transportation system. This is also commensurate with the planning levels discussed in the previous paragraphs. However, in the context of the DPP, there is only one optimizer that must coordinate all activities that will take place on the transportation system. Hence, the movements of both items and transportation assets must be solved simultaneously taking into account interactions between flow of items and that of transportation assets, which adds a new level of complication not dealt with in the theory of DNFP.

Literature regarding DNFP shows that dynamic network flow problem with unsplittable flow property and dynamic network flow problem with time windows have not been studied. We think that these properties of the DPP may contribute to the theory of the DNFP. For example, to send a given amount of flow from a source to a sink in the shortest possible time with unsplittable flow requirement can be regarded as an extension of the quickest path problem. Similarly, maximizing flow through a network by allocating different time windows for each origin-destination pairs can be regarded as an extension of the maximum-dynamic network flow problem. Thus, several research problems regarding DNFP theory can be derived from given properties of the DPP.

2.3.2. Transportation Planning Problems

2.3.2.1. Strategic and Tactical Level Problems

In this section, we consider network design problems at the strategic and tactical levels.

2.3.2.1.1. Network Design Problem (NDP)

Network Design Problem (NDP) at international, national, or regional level deals with the movements of several commodities through a multimodal transportation network and services of several carriers simultaneously. The main purpose of this planning is to adapt a given transportation system to modifications in its environment. Some factors that affect the transportation system are changes to existing infrastructure, construction of new facilities, changes in the volume of production, trade, and consumption, introduction of new technological advances, and changes to environmental conditions. These issues are often part of a cost-benefit analysis and comparative studies of investment alternatives.

Crainic (2003) notes that the study of multicommodity freight flows over a multimodal network is not mature in contrast to passenger transportation where car and transit flows over multimodal networks has been studied extensively and put into practice, e.g., Florian and Hearn (1995) and Cascetta (2001).

Network optimization models are seen as appropriate models to address planning issues at this level. Guelat, Florian and Crainic (1990) and Crainic et al. (1990) give a review of these studies. Crainic (2003) presents a modeling framework based on Guelat, Florian and Crainic (1990). The modeling framework includes a multimodal network comprised of modes, nodes, links, and intermodal transfers. Multiple commodities (people or freight) are moved between origin-destination pairs by specific vehicles and convoys through the network. The model allows a detailed representation of transportation infrastructure, facilities, and services and the simultaneous assignment of multiple commodities to multiple modes. A mode is a means of transportation with its own characteristics such as vehicle type, capacity, and cost measures. A mode may represent a particular transportation service, an aggregation of several carrier networks, and transportation network infrastructures. Intermodal transfers at a node are modeled as link to link transfers. The decision variables in the proposed model are only the flows of commodities. Vehicle and convoy traffic on the network is deduced from the values of the decision variables. Applications of this modeling framework can be found in Crainic, Florian, Leal (1990), Guelat, Florian and Crainic (1990), Crainic et al. (1990), Crainic, Florian, and Larin (1994), and Crainic et al. (2002).

The NDP at this level aims to establish a transportation system taking several transportation modes and carriers into account. Because this suggests a single optimizer coordinating various functions in the system, there is a degree of similarity between the DPP and the NDP. However, the proposed model of the NDP is essentially a multicommodity flow model where only the flow of commodities is considered. The purpose is essentially to allocate the flow of commodities to several carriers or modes to minimize cost. The flow of vehicle traffic is determined given the flow of commodities, i.e., a sequential planning approach is used while in the DPP these decisions are simultaneous. Moreover, time aspect, i.e., scheduling, is not dealt with in compliance with the strategic planning level. Thus, the proposed model of the NDP is not appropriate for the DPP.

When the NDP is considered at the company level, e.g., by a freight carrier, some questions to address are where to locate facilities including, for example, loading and unloading terminals, consolidation centers, rail yards, or intermodal platforms as well as what type of equipment to install in each facility, on which lines to add capacity, and what types of lines or capacity to add. These issues are the subject of *location* and *logistics network design* models.

Location models can be classified into covering models, center models, and median models (Crainic and Laporte, 1997). Covering models locate facilities such as health clinics, post offices, libraries, and schools at the vertices of a network so that demand vertices are covered by a facility. Center models locate p facilities such as fire or ambulance stations at the vertices of a network in order to minimize the maximum distance between demand points and their closest facilities. Median models locate p facilities at vertices on the network and allocate demand between facilities in order to minimize the total weighted distance between demand points and their closest facilities. Two problem classes, namely, *production-distribution* and *hub location* problems, extend modeling features of these problems by taking into account the potential economies of scale associated with the consolidation of cargo and passengers. Although these issues are certainly a part of the problem of designing a distribution system, a part of sustainment planning, they are not considered in the context of the DPP. We refer the reader to Miirchandani and Francis (1990), Daskin (1995), Drezner (1995), Labbe, Peeters, and Thisse (1995), Labbe and Louveaux (1997), Laporte (1988), Federgruen and Simchi-Levi (1995), O'Kelly (1987), Campbell et al. (2002) in Chapter 12 of Drezner and Hamacher (2002) for additional information on location models.

Logistics network design models simply aim to determine what links and with what capacity to open in the network to satisfy demand for transportation at the lowest possible cost. The cost is calculated as the sum of the total fixed cost of links opened and the total variable cost of using the links. Transportation demand is defined between origin-destination pairs. Main decision variables are whether a link is opened or not and the amount of flow of all commodities on the links.

Clearly, logistics network design issues are also a part of sustainment planning. However, they are not considered in the context of the DPP. Additional information on logistics network design models can be found in Magnanti and Wong (1986), Minoux (1986), Ahuja et al. (1995), Nemhauser and Wolsey (1988), Salkin and Mathur (1989), and Balakrishan, Magnanti, and Mirchandani (1997).

Network design problems that have been covered up to now are strategic level problems. Now, we focus on the *tactical level network design problems*. These problems are of interest to the companies that operate consolidation transportation systems and are related to the planning of operations. Hence, they are also referred to as operational. Crainic (2003) refers to these problems as *service network design* problems and classifies the main decisions at this level into four groups:

(1) *Service selection*: The routes – origin and destination terminals, physical route and intermediate stops – on which services will be offered and the characteristics of each service. *Frequency* or *scheduling* decisions are part of this process.

(2) *Traffic distribution*: The itineraries (routes) used to move the flow of each demand: services used, terminals passed through, operations performed in these terminals, etc.

(3) *Terminal policies*: General rules that specify for each terminal the consolidation activities to perform. For rail applications, these rules would specify, for example, the blocks into which cars should be classified (the *blocking* policies) as well as the trains that are to be formed and the blocks that should be put on each train (the *make up* rules). An efficient allocation of work among terminals is an important policy objective.

(4) General *empty balancing* strategies, indicating how to reposition empty vehicles to meet the forecast needs of the next planning period.

Assad (1980), Crainic (1988), Delorme, Roy, and Rousseau (1988), and Cordeau, Toth, and Vigo (1998) give reviews of tactical level models. Network optimization models seem to be the ones mostly used and we are going to discuss some of them in this section. Crainic (2003) classifies them into *frequency* and *dynamic service network design* models.

Frequency network design models address questions such as: What type of service to offer? How often over the planning horizon to offer it? Which traffic itineraries to operate? What are the appropriate terminal workloads? There are two approaches used to formulate service frequencies. In the first one, service frequencies are explicit integer variables. In the second one, "operate or not" binary decision variables are used and service frequencies are derived from traffic flows subject to minimum service levels. The output of these models is the transportation or load plan used to determine daily operating policies. These models can also be used to answer what-if questions in strategic planning. Dynamic service design models aim to plan schedules and to support decisions related to if and when services depart. These models are also considered as operational.

The network optimization model offered by Crainic and Rousseau (1986) uses explicit decision variables to determine how often each selected service will be run during the planning period. The resulting model is a multimodal multicommodity model that integrates service selection and traffic distribution problems with general terminal and blocking policies. The network represents a physical network over which the carrier operates. Nodes in the network represent terminals where particular operations are carried out. Each service is defined by its route through the network, i.e., by its origin, destination, intermediary terminals where the service stops and work may be performed on its vehicles and cargo, capacity on each link of the route, and service class that indicates characteristics such as the mode, preferred traffic or restrictions, speed and priority of the service, etc. Transportation demand is defined in terms of volume, e.g., the number of vehicles, of a certain commodity to be moved between two terminals in the network. Empty vehicles may be included as commodities to be moved between given origin-destination pairs. Traffic moves according to itineraries. An itinerary for a commodity/product specifies the service path used to move some or all of the corresponding demand, i.e., origin, destination, and intermediary terminals where operations are to be performed, the sequence of services between each pair of consecutive terminals where work is performed, the commodity class that indicates characteristics such as priority, minimum service level, preferred transportation mode, etc. Service frequencies define how often each service is run during the planning period. To design the service network means to determine the frequency of each service in the planning period such that the demand is satisfied. Main decision variables are service frequencies and flow of commodities using itineraries. Workloads and general consolidation strategies for each terminal in the system are derived from these variables. The objective is to minimize the sum of the fixed cost of operating a service and the variable cost of moving commodities using itineraries. The delays incurred by vehicles, convoys, and freight due congestion and operational policies at terminals and on the roads are incorporated into the objective function by defining appropriate costs. Penalties may also be defined when service quality standards are announced for not meeting standards.

Rail transportation applications of the proposed modeling framework can be found in Crainic (1984), and Crainic, Ferland, and Rousseau (1984) while LTL applications can be found in Delorme, Roy, and Rousseau (1989).

The transportation planning model for LTL motor carriers introduced by Powell and Sheffi (1983) is an example of service network design model in which service frequencies are not explicitly formulated as integer variables. In this model, the network is composed of nodes and links where nodes refer to terminals and links refer to potential direct services between two terminals. Two types of terminals are considered: end-of-lines where freight originates and terminates and breakbulks where freight is consolidated. The network design decisions are to determine services between end-of-lines and breakbulks and between breakbulk terminals. The flow between end-of-lines is disregarded as it is rare in LTL service.

The main decision variables are binary service design decisions that show whether the carrier offers a service on a link, i.e., between two terminals, the volume of traffic on a link with a certain destination, the volume of traffic handled at a breakbulk terminal, and flow of empty trailers on a link. Service frequency – the number of trailers dispatched from a terminal to another over the planning period – is defined as a function of the volume of LTL traffic between two terminals. The objective is to minimize the total cost of dispatching trailers, moving the loaded and empty trailers, and handling freight in terminals while satisfying demand and ensuring that freight itineraries obey routing restrictions.

Authors implement a heuristic procedure based on a hierarchical decomposition of the problem into a master problem and several subproblems. The master problem is a simple network design problem where the total cost is computed for given selected services. The design is modified by adding or dropping one arc at a time. Each time the design is modified, the subproblems are solved and the objective function is evaluated. The first subproblem routes the loaded LTL trailers. The second problem is an empty balancing problem where supply and demand is adjusted to account for timing conditions not included in the original formulation.

Other references regarding this modeling framework can be found in Powell and Scheffi (1986, 1989), Powell (1986a), Lamar, Sheffi, and Powell (1990), and Braklow et al. (1992).

In *deterministic dynamic service network design problem*, time dimension is introduced into the formulation. This is usually achieved by using a time-expanded network. In the network representation, a service starting from its origin in a given period arrives (and leaves in the case of intermediary stops) later at other terminals. Services thus generate temporal service links between different terminals at different time periods. Temporal links that connect two representations of the same terminal at two different time periods may represent the time required by terminal activities of the freight waiting for the next departure. Additional arcs may be used to capture penalties for arriving too early or too late.

There are two main decision variables. The first ones are the integer decision variables associated with each service. When restricted to binary, these variables indicate whether the service leaves at the specified time. When several departures are allowed in the same period, general integer variables must be used. The second decision variables are continuous flow variables that represent the volume of freight through the network.

The resulting formulations are network design models similar to the ones in the case of static network. However, time dimension significantly increases the size of the network. The size of the network and additional constraints required by time dimension make these problems harder to solve than static ones. The pioneering effort of Morlok and Peterson (1970) that integrates blocking, train formation, and train scheduling results in a very large mixed integer model. As a result, no solution method or application has been offered for the model and only heuristics methods have been used so far.

Farvolden and Powell (1994) present a dynamic service network design model for LTL transportation. The formulation allows for several departures in the same period. An efficient primal-partitioning with column generation algorithm is used to solve the freight routing problem for a given service configuration.

Haghani (1989) attempts to combine the empty car distribution problem with train make-up and routing problems. The dynamic network includes normal and express modes for each service route for each time period, but traffic on each link is prespecified and access to express links is restricted to given markets. Travel times are fixed. Linear functions are used for costs and delays, except for classification, which makes use of a convex congestion function. The dynamic service network design has continuous empty and loaded car flows and integer engine flows. A heuristic decomposition approach is used to solve simpler problems. The study shows that performance, in terms of operating costs, obtained by using an integrated formulation is better than performance obtained by using traditional hierarchical approach.

Gorman (1998) also attempts to integrate the various network design aspects into a scheduled operating plan that minimizes operating costs, meets the customer's service requirements, and obeys operation rules of a particular railroad. Model simplifications must be introduced in order to achieve a comprehensive mathematical network design formulation. The solution method includes generating candidate train schedules using a tabu-enhanced genetic search and evaluating their economic, service, and operational performances. A major US railroad has used this model for strategic scenario analysis of their operations (Gorman 1998a).

Several other network design models make use of binary mixed integer network flow formulations to address railroad operations, e.g., Keaton (1989, 1991, and 1992), Newton, Barnhart, and Vance (1998), and Barnhart, Jin, and Vance (2000).

Kuby and Gray (1993) develop a model for the design of the network of an express package delivery firm. It is a path-based network design model where multistop, aircraft routes (restricted to at most one stop) are selected in and out of a given hub. Paths are generated a priori and the model is solved by a standard mixed integer package. Analyses illustrate the cost-effectiveness of a design with multiple stops over a pure hub-and-spoke network. Kim, Barnhart, and Ware (1999) propose more comprehensive models for the design of the multimodal version of the problem. In the model, several hubs and aircraft types are considered while trucks perform pickup and delivery activities as well as transportation over limited distances. The problem is further complicated by time window restrictions on pickup and delivery times at major collection centers as well as on the sorting periods at hubs. One product is considered in the application. The authors combine heuristics to reduce the size of the problem, cut-set inequalities, and column generation. Branch and bound is then used to obtain an integer solution.

The design of postal networks and services forms a class of problems very close the ones just mentioned. The reorganization of the German postal services belongs to the same problem class but on a more comprehensive scale. To bring the problem down to a manageable size, Grünert and Sebastian (2000) decompose it into several subproblems: the optimization of night airmail network, the design of the groundfeeding and delivery transportation system, the scheduling operations. Vehicle routing models and techniques, which we are going to cover shortly, are used for routing and scheduling tasks. A discrete dynamic network design formulation is also proposed. The air network design formulation is further decomposed into a direct flight problem and a hub system problem, both of which result in fixed cost, multicommodity, capacitated network design formulations. The authors propose using combinations of tabu search and branch and bound to solve the models.

The main decisions at the tactical level are certainly related to the decisions in the DPP. However, both deterministic and dynamic service network design models fall short of solving the DPP. A service actually refers to a transportation asset with a certain route (and schedule). Thus, each transportation asset with a different route (and schedule) can be considered as a different service. As a service is characterized by its route and capacity and transportation demand is defined in terms of the service capacity offered, the problem in these models essentially turn out to be sharing transportation demand among services (transportation assets). Each time a service is selected, transportation demand is reduced in the amount equal to the service capacity. Thus, only the movement of one type of entity, i.e., transportation assets, is considered without dealing with the content of the load on a transportation asset, which is not sufficient in the DPP.

2.3.2.2. Operational Level Planning Models

Strategic and tactical plans can be drawn up to guide operations, but the operational capabilities of a firm/organization ultimately determine its performance.

In operational level planning, two factors, *time* and *stochasticity* are emphasized. Time factor is important because customer requests must be met in real time, time restrictions must be obeyed, and the impact of today's decisions on future decisions must be taken into account. Stochasticity is of concern because there are many uncertainties in real life, e.g., travel time between two points, the volume of transportation demand, etc. These two characteristics must be reflected in the models and methods aimed at operational planning and management issues.

2.3.2.2.1. Vehicle Routing Problem (VRP)

The *Vehicle Routing Problem (VRP)* arises especially in distributing and collecting letters and packages. In the VRP, there is a set of demand points dispersed in a geographic region and a daily demand of a specific commodity has to be delivered daily to each of the demand points. The daily demand at each demand point is known and usually different at different demand points. The deliveries to demand points are to be made from a central depot and sufficient supply is always available at the depot. A fleet of capacitated vehicles is to serve the demand points and direct travel distances between the demand points and the depot are known. It is assumed that a vehicle's load capacity exceeds the demand of each demand point and that a vehicle is allowed to visit each demand point exactly once (sometimes multiple visits to a demand point are allowed, e.g., split delivery). The objective is to find a set of routes for the vehicles, where each route begins and ends at the depot, serves a subset of the customers without violating the capacity constraints, while minimizing the total length of the routes. Bodin et al. (1983) and Golden and Assad (1988) give comprehensive surveys on the VRP.

Several variants of the VRP have been studied depending on the features incorporated. One variant that has gained considerable interest is the *VRP with Time* Windows (VRPTW) or the *Vehicle Routing and Scheduling Problem (VRSP)*. In this problem, the requirement is to make deliveries to the demand points within their pre-determined *time windows*, i.e., earliest delivery time and delivery deadline are imposed. In this problem, both spatial and temporal aspects of vehicle movements are considered. Golden and Assad

(1986) and Solomon (1987) give a considerable amount of the analysis on the VRSP.

Desrochers et al. (1990) propose a classification scheme for the VRSPs to cover variants of the problem that have been studied in the literature. They propose to classify a VRSP based on several characteristics: (1) addresses (number of depots, type of demand, e.g., whether the customers are located at the nodes and/or edges or the customers correspond to an origin-destination pair, address scheduling constraints, e.g., whether there are no scheduling constraints or time windows are allocated, address selection constraints, e.g., whether all/some addresses are visited, etc., (2) vehicles (homogeneous/heterogeneous fleet, fixed or variable fleet size, physical characteristics of the vehicles, e.g., whether they have compartments or not, scheduling constraints), (3) problem characteristics (the network underlying the problem, e.g., directed and undirected, or mixed, service strategy, e.g., whether splitting of the customer demand is allowed, vehicles are allowed to start a route at a depot and finish at another depot, or one or more routes per period are allowed, relations between addresses and vehicles, e.g., whether precedence relations and depot-address, address-address, depot-vehicle, address-vehicle, vehicle-vehicle restrictions exist or not, and (4) objectives (minimize route duration, vehicle costs, the penalty implying the deviation from preferred service level.). Desrochers et al. (1998) propose a model base and algorithm selection system based on Desrochers et al. (1988). Bodin (1990) survey practical VRSPs and discuss how information technology is used in solving these problems.

In the VRSP literature, many of the characteristics outlined above are treated individually. Most studies focus on the basic VRSP formulation of Desrochers et al. (1988) to deal with different characteristics of the problem. Dror and Trudeau (1990) study the split delivery routing and propose heuristics based on splitting and merging routes. Fisher et al. (1995) discuss the pick-up and delivery routing problem with a homogeneous fleet in which loads are in integer truckloads. Their objective is to minimize total travel cost with the restriction that a vehicle cannot carry more than one order at a time. With these assumptions, the problem is solved as a network flow problem. Ribeiro and Soumis (1994) study the multi-depot version of the problem as an integer multi-commodity flow problem assuming non-split delivery. Although a homogeneous fleet is used, the problem is treated as a multi-commodity flow problem because vehicles leaving different depots represent different commodities as each vehicle is tracked with its origin depot. The problem turns out to be the assignment of trips to vehicles such that each trip is carried out by one vehicle and the number of vehicles at each depot is not exceeded. Ribeiro and Soumis show that linear programming relaxation of the integer multi-commodity flow formulation gives a good lower bound. Malandraki and Daskin (1992) study the vehicle routing problem with time-dependent arc distances. In this problem, the travel time between two demand points or between a demand point and the depot depends on the distance between the points and the time of day, e.g., rush hour, etc. They develop mixed integer programming formulations treating travel times as step functions and give several heuristics.

Because the vehicle routing problem is NP-hard, heuristics and optimization algorithms are of special interest. In this regard, extensive discussions on heuristics algorithms can be found in the literature. See, for example, Laporte (1992) and Fisher (1995). Recent examples of global search heuristics (e.g., simulated annealing, tabu search, etc.) designed for solving these problems can be found in Rodriguez et al. (1998) and Gendreau et al. (1999).

The usual setting of the VRSP focuses on planning local delivery and pick-up operations such as those of UPS and other delivery firms. Although VRSP is relevant to the DPP, this setting is too restricted to be used in the planning of operations in the context of deployment planning. In the VRSP, the assignment of individual vehicles to demands is of concern and the purpose is to develop a single route for each vehicle to satisfy demand. A second route for a vehicle is determined only after it finishes its first route, which means actually solving another VRSP. In the DPP, a set of vehicles may be assigned to the same job, i.e., convoy formation is of concern, and an itinerary consisting of successive assignments to different jobs (as opposed to a single job/route) is to be determined simultaneously. Because a demand in the VRSP is defined in terms of vehicle capacity, the capacity of transportation asset is depleted at each point the vehicle visits. The content of a load/package is not important to the carrier/planner because the load is not *reusable* from the planner's point of view in the sense that it disappears from the system as soon as a vehicle is assigned to it. This makes it unnecessary to model a load explicitly in these models because there is no need to assign other vehicles to the same load. In the DPP, on the other hand, a load is reusable and may have to be carried on different types of vehicles at different time intervals. Similarly, transportation assets to deploy in the DPP are reusable as they can be allocated to different loads at different time intervals.

Multimodality in the DPP actually complicates the definition of a customer. In the VRSP, a vehicle is expected to visit a demand point in the given time window. In the DPP, time window is assigned to the final destination of the unit. Thus, unless there is a single mode where the origins, destinations, and time windows are known exactly, the definition of a customer is not clear. For example, the planner cannot know to what transfer point a unit moves after it leaves its origin and what time it arrives there. In the VRSP, there is no such thing as the movement of empty vehicles, which is one of the main planning issues in the DPP. To sum up, the DPP is different from the VRSP in many respects and cannot be used to plan operations in the DPP. However, it can be used in the planning of sustainment operations.

2.3.2.2.2. Dynamic Resource Allocation Problem (DRAP)

In the DRAP, tasks arriving over time are realized by a set of *reusable* resources of different types. For example, empty vehicles, trailers and rail cars are allocated to the appropriate terminals; motive power tractors and locomotives to services; crews to vehicles or services; loads to driver-truck combinations; empty containers from depots to customers and returning containers from customers to depots; and so on.

Crainic (2003) lists the following common characteristics of these problems:

(1) Some future demands are known, but most can only be forecasted, and unpredictable requests may happen.
(2) Many requests materialize in real or quasi-real time and must be acted upon in relatively short time.

(3) Once a resource is allocated to an activity, it is no longer available for a certain duration (whose length may be subject to variations as well).

(4) Once a resource becomes available again, it is often in a different location than its initial one.

(5) The value of an additional unit of a given resource at a location greatly depends on the total quantity of resources available (which are determined from previous decisions at potentially all terminals in previous periods) and the current demand.

Powell (1996) studies the problem of assigning drivers for a truckload motor carrier to handle loads that arise over time. He classifies the elements of the *fleet management problem* into *supply management* and *demand management*, which are in general valid for other application areas as well. The *supply management* includes (1) what driver (resource) to assign to a load (demand), (2) repositioning empty drivers (excess capacity), (3) routing and scheduling of the driver (resource) while moving a load and the *demand management* includes (1) load (demand) acceptance/rejection, i.e., carrier may accept or reject certain loads based on capacity or system balance considerations and (2) load (demand) solicitation, i.e., the carrier may wish to aggressively solicit freight out of specific regions or in specific lanes to correct short-term balance problems. The assignment of a resource to a task produces a profit, removes the task from the system, and modifies the state of the resource.

Brown and Graves (1981) develop an integer programming formulation for the real-time routing and scheduling problem for petroleum tank trucks. The purpose of the model is to determine routes for trucks to meet known (deterministic) customer demands. Bell et al. (1983) develop a set partitioning formulation for real-time routing and scheduling of tanker trucks in the distribution of industrial gases.

Hane et al. (1995) study the fleet assignment problem, which is to determine which type of aircraft should fly each flight segment given a flight schedule and a set of aircraft. They develop a large-scale multicommodity flow problem with side constraints defined on a time expanded network and explain the methods they use to improve the solution time of the model.

In the crew scheduling problem, crews are assigned to vehicles and convoys in order to support the planned operations. It finds applications especially in airline industry, e.g., Ball and Roberts (1985), Crainic and Rousseau (1987), and Marsten and Shepardson (1981). Given a fixed set of flights, the purpose is to develop an itinerary for each crew so that all flights are covered at least cost. There are also numerous other issues related to manpower management such as the scheduling of reserve crews, terminal employees (e.g., Nobert and Roy 1998), maintenance crews, etc. The resulting mixed-integer formulation is usually very large and addressed by column generation and branch-and-price techniques. See, for example, Barnhart and Talluri (1997), Desrosiers et al. (1995), Desaulniers et al. (1998). Crew scheduling issues in the freight transportation industry have rarely been studied, e.g., Crainic and Roy (1992).

One major problem in transportation planning is the *empty vehicle distribution problem*. That there are geographic differences in demand and supply of each commodity often results in an accumulation of empty vehicles in regions where they are not needed and in deficits of vehicles in other regions that require them. Then, vehicles must be moved empty, or additional loads must be found, in order to bring them where they will be needed to satisfy known and forecasted demand in the following planning periods. This operation is known as *repositioning or empty balancing* and is a major component of what is known as *fleet management.* In its most general form, fleet management covers the whole range of planning and management issues from procurement of power units and vehicles to vehicle dispatch and scheduling of crews and maintenance operations. However, the term designates a somewhat restricted set of activities: allocation of vehicles to customer requests and repositioning of empty vehicles.

Empty balancing is a major objective of dispatchers and a central component of planning and operations of most transportation firms. Although it is considered as operational, it must also be considered at the tactical level. For example, in rail transportation, empty rail cars are put on the same trains as loaded ones and thus contribute to an increase in the number of trains, in the volume of vehicles handled in terminals and, ultimately, in system costs and delays. For planning purposes, the demand for empty cars may be approximated and introduced in tactical model by viewing empties as another commodity to be transported, e.g., Crainic, Ferland, and Rousseau (1984). A similar approach may also be used for the planning of multimodal regional or national systems, e.g., Crainic, Florian, and Leal (1990). The issue is also relevant in LTL trucking where empty balancing is an integral part of a transportation plan. In this case, a load plan is first obtained for the actual traffic demands, and an empty balancing model is then solved to reposition the empties, e.g., Roy and Delorme (1989) and Braklow et al. (1992).

Dejax and Crainic (1987) give a review studies in this area spanning the whole spectrum of modeling approaches from simple static transport models to formulations that integrate the dynamic and stochastic characteristics of the problem.

The first empty vehicle allocation models are straightforward transportation formulations, e.g., Leddon and Wrathall (1967), Misra (1972), and Baker (1977). In these models, the distribution of empty cars is optimized to minimize the total cost using given estimations of future supply and demand of empty cars of a homogeneous fleet at the yards of the network, and the cost in car-hours usually, for each pair of yards.

The second modeling approach considers the time aspect explicitly. Starting with contributions of White (1968) and White and Bomberault (1969) for rail car distribution, and of White (1972) for container allocation, many models that deals with the distribution of empty vehicles are in the form of a dynamic transshipment network optimization model, e.g., Herren (1973, 1977) and McGaughey, Gohring, and McBrayer (1973). Linear programming and network flow algorithms are usually applied to solve the model. This line of research is still very active today; however, the formulations are more complex. Multiple commodities, substitutions, and integer flows are some of the characteristics that add realism to these formulations, e.g., Shan (1985), Chih (1986), Turnquist and Markowicz (1989), Markowicz and Turnquist (1990), and Turnquist (1994). There are also studies that impose additional conditions on empty vehicle distribution such as limited hauling capacity for empties and predefined itineraries, e.g., Joborn (1995), Holmberg, Joborn, and Lundgren (1998), Joborn et al. (2001).

Shan (1985) and Chih (1986) present multicommodity network flow models for empty freight car distribution where each commodity represents freight cars of one specific type owner. The purpose of the model is to determine which cars should be used to meet the demands of the customers. Joborn (1995) and Holmberg et al. (1998) develop a multicommodity network flow model that considers the capacity restrictions on trains for repositioning of empty freight cars and the arrival and departure times of the trains with the modeling assumptions.

Another modeling approach in empty vehicle distribution considers uncertainties explicitly. The first comprehensive effort in this direction is by Jordan and Turnquist (1983) for rail transportation. The formulation aims to maximize the profits of the firm and integrates revenues from performing the service as well as various costs from moving cars between yards, holding them at yards, or from not filling orders due to stockouts. The model structure is again a multicommodity dynamic network. Stochasticity of supply, demand, and travel times is explicitly considered. The resulting model is a nonlinear optimization formulation.

Powell (1986) develops dynamic network models for the dynamic vehicle allocation problem, in which a fleet of vehicles are to be managed over space and time, with random arc capacities.

A similar approach is proposed by Beaujon and Turnquist (1991) for a model that simultaneously considers vehicle inventories at terminals and their allocation in order to answer fleet-sizing issues. The whole research area addressing the dynamic allocation of limited resources in uncertain environments naturally continues with these important developments.

Powell et al. (1995a) propose a new modeling approach that addresses dynamic resource allocation problem as a logistics queuing network. This approach views the system as a network of double-ended queues, comprised of a queue of vehicles waiting to serve customers, and a queue of customers waiting to be served by a vehicle. This modeling approach provides flexibility in modeling complex operations by decomposing large dynamic problems into sequences of small problems that deal with one location at a time, one time period at a time.

Powell and Carvalho (1997) extend this approach to multi-commodity problems, e.g., heterogeneous fleet. Powell and Carvalho (1998) test this approach in the management of a fleet of flatcars for a railroad. Their study results in two interconnected dynamic resource allocation models, one to optimize the flows of trailers and containers owned by the railroad, and the second to optimize the flows of the flatcars. They solve the models sequentially: They first optimize the movements of trailers and containers and then add the planned movement of empty equipment to the customer-driven demands to move loads.

Other studies regarding this approach can be found in Carvalho (1996), Carvalho and Powell (2000), Powell and Carvalho (1998a), Powell (2002).

In the DPP, we do not deal with driver assignment and crew scheduling problems. We assume that sufficient number of drivers and crews will be available. Hence, these two problems are not relevant to the DPP. The dynamic fleet management problem is related to the DPP. However, the issues pointed out for the VRSP all apply to fleet management problem and hence the models addressing the fleet management problem are not sufficient to solve the DPP.

Empty vehicle allocation problem, also a part of the fleet management problem, is related to the DPP; however, models addressing the empty vehicle allocation problem are not sufficient for the DPP. In these models, only a single type of entity, i.e., empty vehicles, is moved through a network to meet demands for empty vehicles. Loaded vehicles are not modeled explicitly as they are assumed to be predetermined. Routing is essentially simple because vehicle movements occur between terminals rather than on a physical network. The purpose is not to develop itineraries for empty vehicles. Once empty vehicles are assigned to a task, they are considered out of the system or assumed to enter the system after a certain time. Hence, the movements of vehicles are not tracked.

Although the movement of empty and/or loaded cars on the trains resembles the movement of items on transportation assets, the existing models for this problem have the following features, which are similar to the ones mentioned in the previous paragraphs, that make them insufficient to solve the DPP: (1) In most models, the routing of freight cars is not considered explicitly; the flow requirements are generally defined from one terminal to the other. (2) Although train routing, train makeup, and car distribution should be integrated, either the train routing or car flows are accepted as given and one decision is based on the other given one. (3) Timetable for trains, e.g., arrival/departure times of trains and available train capacities, are generally accepted as given. Because the loading, unloading, and other operations at yards are included in the timetable, such issues are disregarded in the modeling.

It is appropriate to differentiate between convoy formation in the context of rail transportation and convoy formation in the context of the DPP. In rail transportation, the point is to form blocks of cars to be able to handle them as a single entity; cars are formed into blocks taking their final destinations into account. This problem is generally called the *blocking problem* for which specialized models are developed. In the DPP, a convoy is formed due to the need to move a deployment component as a whole. In this regard, a "natural block" is automatically created. It is still possible that a blocking problem be solved in the DPP when rail transportation is used. In this case, the idea would be essentially to bring natural blocks to form new ones. However, it is likely that there will not be a need for such a blocking operation as a natural block will usually form a train. Note that several types of cars, i.e., ones that carry personnel and ones that carry cargo, may comprise a block as ground vehicles of different types, e.g., trucks, buses, and tank carriers, may comprise a convoy for ground transport.

In all, transportation planning models in the context of the SCM do not address routing, scheduling, and resource allocation issues in the DPP simultaneously. Models in the literature deal with only certain parts of transportation systems, e.g., local delivery and rail transportation. This is a natural result of the fact that several firms/organizations own several parts of the transportation system and that they only concentrate on planning issues relevant to them. Thus, there does not seem to be a comprehensive model that studies the movement of commodities/items through several transportation systems, i.e., transportation modes and different types of services, using a system approach. As the responsibilities are shared among several organizations and hence a hierarchical planning approach is adopted, the models, in addition to addressing only certain parts of the transportation system, are directed at routing, scheduling, or resource allocation depending on the planning level. For example, when scheduling issues are studied, routing issues are assumed given or other simplifying assumptions are adopted.

Due to inadequacy of existing transportation planning models to address the DPP, the need for specialized models are foreseen/realized by some researchers. In the next section, we give literature regarding those studies.

2.3.3. Mobility Analysis Problem (MAP)

The models that directly address the DPP are grouped under the name of military *mobility models*. Although there is a concerted effort to develop models in this area for more than 20 years, the literature review shows that the attempts to solve the problem are generally simulation based and that the existing simulation and optimization based studies address only certain parts of the problem.

Schank et al. (1991) review and analyze a number of strategic mobility models. They evaluate the attributes and limitations of the major existing models up to that time, and determine whether another computer model is necessary. The study concentrates on resource planning, which is typically long-range force planning and programming. The study indicates that the strategic mobility models examined have the following shortcomings: they do not optimize the usage of transportation assets; they all work in one direction only; they have limited credibility outside the organizations that use them; they do not sufficiently recognize uncertainty; they have narrow, rigid objective functions; and their output measures do not adequately serve analysts' needs. They recommend that those broader-based trade-off analyses be addressed using new formulations of traditional mathematical programming procedures and off-the-shelf software.

McKinzie and Barnes (2003) review current, legacy, and supporting military mobility models. They focus on four major ones in current use: Global Deployment Analysis System (GDAS), Joint Flow and Analysis System for Transportation (JFAST), Model for Intertheater Deployment by Air and Sea (MIDAS), and Mobility Simulation Model (MobSim©).

Table 1 compares the models with respect to task coverage. When a model addresses a task completely within its own framework, an X is placed in the table for the associated task. If the model obtains a task or part of it from another model, then an O is placed in the table. If the task is not available when using the model, then the cell for the associated task is left blank.

Table 2 shows how the models do calculations for each of the stages within the deployment process. Calculations are performed either by simulation or by a simplified mathematical calculation. An (S) denotes that the model simulates the stage and a (C) indicates that the model provides a simplified calculation for the stage.

Table 3 compares the models with respect to several features such as operating platform (UNIX or PC and software requirements), ease of use by analysts and planners, level of tracking detail for both cargo and pax and transportation assets, the ability to model multiple port pick up and drop off locations for each airplane/ship modeled, and set up and run times.

Table 1. Comparison of the models with respect to task coverage. (X : the task is covered completely within the model's framework, O : the model obtains the task or part of it from another model, Blank : task is not available within the model) (McKinzie and Barnes, 2003)

	GDAS	JFAST	MIDAS	MobSim©
Merge Files	X	X	X	X
Aggregate Records	X	X	X	X
Prioritize Records	X	X	X	X
Select Modes	X		X	X
Schedule Cargos	X	X	X	X
Simulate Movements	X	X	O	X
Multi port PU & DO				
Airlift POE & POD	X			X
Sealift POE	X			X
Sealift POD	X	X	X	X
# Transshipment Points	Unlimited	1	Unlimited	4
VISA Modeling	X		X	X
Prepare Textual Output	X	X	X	X
Prepare Graphical Output	X	X	X	X
Check and Correct	X	X	X	X

Table 2. Comparison of the models with respect to type of calculations used for each stage of the deployment process (S : simulation, C : simple calculation) (McKinzie and Barnes, 2003)

	GDAS	JFAST	MIDAS	MobSim©
Home Station to POE	S	S	C	S
Sealift				
POE Berthing	S	S	S	S
POE Loading	S	S	C	S
Movement to Convoy Point	S	S	S	S
Convoy Assembly	S	S	S	S
Movement to POC	S	S	S	S
POD Berthing	S	S	S	S
POD Unloading	S	S	C	S
Airlift				
POE Berthing/Ferry	S	S	C	S
POE Loading	S	S	C	S
En Route Stop Operations			S	S
Movement to POD	S	S	C	S
POD Ground Operations	S	S	C	S
POD Unloading	S	S	C	S
Movement to Recovery Area	S		S	S
Recovery Ground Operations	S		S	S
POD to Destination				
Movement to Marshalling Area	S	S	C	S
Marshalling	S	S	C	S
Movement to Destination	S	S	C	S

In scheduling transportation assets and assigning cargo and pax, GDAS uses route insertion techniques over a rolling time horizon. Once the schedule and assignments are complete, it performs a deterministic simulation to determine the actual arrival, departure, loading, unloading and queuing events at each facility. In JFAST, schedules are generated by using only a simple greedy heuristic. JFAST assesses cargo and pax based on priorities and schedules the cargo and pax starting from the highest priority to the lowest. So, there is no optimization process considered in the system. MIDAS uses a greedy feasible solution obtained by a one-pass greedy algorithm. MobSim does not find the optimal assets, but it attempts to a find good transportation asset mix within MobSims scheduling algorithm. When cargo and PAX are ready to move, MobSim randomly looks for vehicles by type by first looking at existing idle vehicles. If no vehicle exists, it creates a new vehicle subject to constraints on total number of vehicles to be used.

Taking into account all of the above issues, McKinzie and Barnes (2003) conclude that "the major aspect that is lacking in the models today is the use of advanced optimization techniques for estimating force closure. Each mobility model described in this paper either uses cumbersome ineffective classical optimization algorithms or simplistic and ineffective greedy approaches to find solutions. This aspect was addressed as a shortcoming in the models eleven years ago (Schank et al., 1991) and remains the major shortcoming today."

Table 3. Comparison of models (X : ability exists, A : air, S : sea, TPFDD : time-phased force deployment data – See Section 3.2 for TPFDD) (McKinzie and Barnes, 2003)

	GDAS	**JFAST**	**MIDAS/AMP**	**MobSim**
Operating Platform	PC	PC	Unix	PC
Tail Number Aircraft Scheduling	X	X	X	X
Line Item Number Tracking	X	X	X	X
Multiple Pick Up Locations by Mode	A & S	S	S	A & S
Multiple Drop Off Locations by Mode	A & S	S	S	A & S
Government owned software	X	X	X	**NO**
Avg. Model Set Up Time (large problem) 85,000 TPFDD records	30 days	7 days	14 days	14 days
Avg. Model Run Time (large problem) 85,000 TPFDD records	4 hours	2 hours	2 hours	30 min
Avg. Model Set Up Time (small problem) 45 TPFDD records	3 hours	10 min	15 min	15 min
Avg. Model Run Time (small problem) 45 TPFDD records	1 min	1 min	1 min	5 min
Data Entry / Complexity of Use	Hard	Easy	Hard	Easy

Following is a review of analytical studies that aim to use mathematical programming formulations for mobility analysis.

Wing et al. (1991) describe the Mobility Optimization Model (MOM) developed at the Naval Postgraduate School (NPS). It is a time-dynamic model that includes both airlift and sealift assets, but has a single-channel topology and hence is not designed to capture the airlift system's transportation network. Yost (1994) describes THRUPUT developed at the US Air Force Studies and Analysis Agency (AFSAA). It is a time-static strategic airlift model on a general routing network. Weng (1994) describes THRUPUT II also developed at NPS. It is a time-dynamic model with the ability to route aircraft through a general network and combines the features of MOM and THRUPUT. THRUPUT II is extended in Morton, Rosenthal, and Weng (1996). However, the model does not consider aerial refueling, crew scheduling and transshipment options. Other studies to improve the model include several theses that examine stochastic airlift models (Goggins, 1995), route generation techniques (Turker, 1995), route prioritization (Toy, 1996), and aggregation schemes (Fuller, 1996). THRUPUT II is reported to serve as a real-world test problem for the development of a solution methodology for large-scale staircase linear programs, described in Baker (1997) and Baker and Rosenthal (1998).

In parallel with the THRUPUT modeling efforts at NPS, a group at RAND has developed a similar model called CONOP (CONcept of OPerations). It captures many details not incorporated in THRUPUT II: aerial refueling, flow-balance and utilization constraints for crews, options for direct delivery versus delivering cargo that is subsequently transshipped by in theater aircraft, and optional in-theater recovery bases where aircraft may receive services and crew changes. On the other hand CONOP does not offer sufficient resolution with respect to ownership (the associated military unit) of the cargo being delivered. Killingsworth et al. (1994) uses CONOPT to conduct an investigation of the utility of aerial refueling tanker aircraft within the strategic air mobility system.

Baker et al. (1999, 2002) describe NRMO (NPS/RAND Mobility Optimizer), which is a large-scale linear programming model and merges CONOP's ability to examine alternative delivery strategies and THRUPUT II's ability to track cargo ownership for optimizing strategic airlift capability. The NRMO routes cargo and troops through a specified network with a given fleet of aircraft subject to many physical and political constraints. The model captures various aspects of an airlift system in a deployment, including aerial refueling, tactical aircraft shuttles, and constraints based on crew availability. The authors state that the model is designed to provide insight into issues associated with designing and operating an airlift system but not to provide operational flight schedule recommendations. Some example usages of the model that the authors list are allocating resources that govern the processing capacity of airfields, assessing the relative performance of different mixes of aircraft types, evaluating investment (or divestment) decisions in airfields, and studying roles for aerial refueling aircraft.

Of the military mobility models, the closest one to ours is the NRMO. However, NRMO and our model are different from each other with respect to several aspects. The underlying network in our model is a physical transportation network consisting of highways, railways, flight routes, and shipping lanes while in NRMO the network of interest is a simpler one defined by airports and arcs corresponding to direct flights between them. One major difference between our model and NRMO is that moving a deployable unit from its home base to its destination requires determining a route of movement on the physical transportation network in our problem whereas routing decisions are absent in NRMO since direct flights between on-load and off-load bases (airports) (with a stopover for refueling as necessary) predetermines the routing structure. Another major difference is the requirement of a convoy formation for movement in our model while no convoy formation is required in NRMO (each flight is a convoy by itself). The scheduling issues encountered in our model is substantially more complicated than in NRMO due to carrier changes at transfer points as well as the presence of possible synchronization and precedence requirements that must be obeyed during movement. Finally, the fact that the underlying network in our model is multi-modal causes additional complications in resource allocation at transfer points that arise from the need to handle transfer of items between different transportation assets of different modes. A comparison of the NRMO and our model in more detail is given in Chapter 3 (the end of Section 3.2).

Niemi (2000) modifies and extends the existing deterministic NRMO model to include stochastic parameters. He introduces stochastic ground times into the model to bring desired flexibility and hedging against uncertainty into the airlift system.

Özdamar and Ekinci (2002) develop an optimization model to provide decision support in dispatching commodities (e.g., medical materials and personnel, food, specialized equipment, etc.) to distribution centers in affected areas during military crises and natural disasters complying with the time-dependent supply and demand. With respect to a military operation, this problem addresses the sustainment of military forces after they are deployed to their areas of operations. In their model, they assume a multimodal network composed of supply/demand locations and direct routes (arcs) belonging to one or more transport modes between them. A prespecified traversal time is associated with each arc depending on its mode, i.e., the traversal time of an arc does not vary with respect to type of transportation asset. They do neglect delay times due to loading, unloading, and mode switching operations at the nodes. Hence, the model does not actually address multimodality. Notice that this network structure is similar to but simpler than that of the NRMO and hence there are no routing decisions. Remarks made for scheduling, unsplittability, precedence/synchronization issues in the NRMO are valid for this model as well. The problem turns out to be assigning available transportation assets at a node to arcs emanating from that node such that sufficient capacity to transport commodities is provided on the arcs. The model is essentially a combination of two multicommodity network flow models, one for commodities and one for transportation assets, where each node is a holdover node in the sense that an inventory of transportation assets and commodities is allowed.

In all respects, the DPP appears to stand out as a unique and multi-faceted problem for which existing models in the literature fall short of. In this dissertation, we break away from the existing literature and give an all-encompassing optimization model that deal with all aspects of the DPP simultaneously.

CHAPTER 3

COST MINIMIZATION DEPLOYMENT PLANNING MODEL

In this chapter, we give the abstraction of the DPP, the formulation of the *cost minimization deployment planning model* in which the objective is to plan the deployment of units with minimum transportation cost, a solution methodology to solve the model, and computational results using the solution methodology.

3.1. Abstraction of the Problem

We may view the DPP as posed on a network $\tilde{G}=(\tilde{N},\tilde{A})$ defined by the union of five sub-networks $\tilde{G}_i=(\tilde{N}_i,\tilde{A}_i),\ i=1,\ldots,5$, corresponding, respectively, to ground, rail, air, sea, and inland-water transportation. We assume each $\tilde{G}_i=(\tilde{N}_i,\tilde{A}_i)$ is connected and directed. The node and arc sets are defined by $\tilde{N}=\bigcup_{i=1}^{5}\tilde{N}_i$ and $\tilde{A}=\bigcup_{i=1}^{5}\tilde{A}_i$ where $\tilde{A}_1,\ldots,\tilde{A}_5$ are assumed to be disjoint. Multiple arcs belonging to the same or different transportation modes between nodes i and j are allowed and differentiated by assigning a different arc number to each arc in the network. Nodes that are common to at least two of the node sets $\tilde{N}_i$ are transfer nodes where a switch occurs in movement from one transportation mode to another.

Let $U=\{1,\ldots,\bar{u}\}$ be the active set of units that need to be deployed. For each $u \in U$, a source-destination pair (s_u, t_u) is specified with $s_u \in \tilde{N}$ denoting the home base and $t_u \in \tilde{N}$ denoting the designated destination. For convenience, let $N_S = \{i \in \tilde{N}: \text{node } i \text{ is an } s_u \text{ for some } u \in U\}$ and $N_D = \{i \in \tilde{N}_i: \text{node } i \text{ is a } t_u \text{ for some } u \in U\}$. Some nodes may be both in N_S and in N_D. Let N_{TR} be the set of nodes that are transfer points (harbors, airports, rail stations). We refer to all remaining nodes as transshipment nodes, i.e., $N_T = \tilde{N} - N_S \cup N_D \cup N_{TR}$, generally used as control points to check the movement of a unit. Define also AF_i and AB_i to be the *forward* and *backward stars* of node i, respectively, where AF_i (AB_i) consists of arcs whose tails (heads) are at node i.

An item list I_u is given for each unit $u \in U$ that specifies the set of items (personnel, equipment, and supplies) to be moved for that unit. We assume I_u is partitioned into $q(u)$ subsets $I_u^1,\ldots,I_u^{q(u)}$ where each subset defines a component that must be moved as a whole. In the current practice, $q(u)=3$, corresponding to advance, pax, and cargo parties. For each unit $u \in U$, three parameters e_u, a_u, and b_u are given specifying, respectively, the earliest time to depart from s_u and the earliest and latest times to arrive at t_u. The same earliest and latest times are valid for all components and hence for all items of unit u.

For modeling purposes, we assume that all data regarding the to-be-deployed items of all units are arranged in a list as exemplified in Table 4. Each line on the list refers to a particular item that belongs to a particular deployment component of a particular unit. A line specifies the associated item's quantity (in number of units) (column 5), the earliest time to depart from its origin (column 6), the earliest and latest times to arrive at its destination (columns 7 and 8), the dimensions for one unit of it, transportation assets and parts of the transportation network it can use, and any other related data.

As an indexing convention, each deployment component of each military unit is assigned a distinct index $g \in \{1, 2, ..., q^*\}$ where q^* is the total number of components (i.e., $q^* = \sum_{u \in U} q(u)$). Similarly, each line, i.e., a particular item in possession of a particular component g, is assigned a distinct line index c. The indexing is done in such a way that components that belong to the same military unit and items that belong to the same deployment component are consecutively numbered. Table 4 illustrates both indexing conventions (columns 1 and 9) on a typical item list. Table 4 gives the item list for two military units A and B. The last column in Table 4 indicates that unit A has two components, numbered 1 and 2, while unit B also has two components, numbered 3 and 4. The list, if expanded, goes on with unit C, D, E, etc., where the components belonging to these units receive consecutively increasing numbers. In Table 4, components 1 and 2, i.e., $g = 1$ and $g = 2$, have 1 and components 3 and 4, i.e., $g = 3$ and $g = 4$, have 6 different items. Thus, there is a total of 14 items, each of which is given a different number to refer to a particular item in a particular component. Notice that the indexing of items is not to number each individual item of a certain type. For example, $c=1$ in the list (of Table 4) refers to "200 troops of component 1" while $c=2$ refers to "50 troops of component 2," not an individual troop.

Most data regarding items is obtained from data regarding units, e.g., the parameters e_c, a_c, and b_c are derived from e_u, a_u, and b_u, respectively, where $c \in I_u$. τ is the reference time at which the whole deployment activity begins.

The first indexed item in each deployment component I_u^i is designated as the <u>leader item</u> for that component with the understanding that all other items in that component follow the same route and the schedule as does the leader.

For a given unit u, there may be certain precedence requirements between components of u (such as, an advance party must arrive at t_u before a pax or cargo party). There may also be a synchronization requirement between two components, say g and g', if components g and g' must arrive at a node simultaneously. Precedence and synchronization requirements are given in any convenient form (e.g., as a list) and incorporated into the model as side constraints. It is also possible to have precedence or synchronization requirements between components that belong to different units whose home bases or destinations coincide.

Table 4. A portion of an item list (τ+h means that item is ready for movement or to be at its destination h time periods after the day the movement is announced to start).

Line Index (c)	Unit	Deploy-ment compo-nent	Item	Qty.	e_c	a_c	b_c	Compo-nent Index (g)
1	Unit A	Pax	Troop	200	τ +3	τ +13	τ +15	1
2	Unit A	Cargo	Troop	50	τ +3	τ +13	τ +15	2
3	Unit A	Cargo	M-60 Tank	5	τ +3	τ +13	τ +15	2
4	Unit A	Cargo	Pax Carrier	14	τ +3	τ +13	τ +15	2
5	Unit A	Cargo	MRC Truck	5	τ +3	τ +13	τ +15	2
6	Unit A	Cargo	2 m^3 Box	5	τ +3	τ +13	τ +15	2
7	Unit A	Cargo	1.5m^3 Box	10	τ +3	τ +13	τ +15	2
8	Unit B	Pax	Troop	200	τ +4	τ +10	τ +14	3
9	Unit B	Cargo	Troop	50	τ +4	τ +10	τ +14	4
10	Unit B	Cargo	M-60 Tank	5	τ +4	τ +10	τ +14	4
11	Unit B	Cargo	Pax Carrier	14	τ +4	τ +10	τ +14	4
12	Unit B	Cargo	MRC Truck	5	τ +4	τ +10	τ +14	4
13	Unit B	Cargo	2 m^3 Box	5	τ +4	τ +10	τ +14	4
14	Unit B	Cargo	1.5m^3 Box	10	τ +4	τ +10	τ +14	4

We now focus on transportation assets. We assume that there is a list of transportation assets where each line on the list is indexed by v and refers to transportation assets of a type (truck, tank, armored vehicle, cargo plane, etc.), from a source (organic, common use, civilian), and at a location (home base if organic, the most recent location at the time a call for deployment is issued if common use or civilian), i.e., the set of transportation asset types is partitioned into subsets based on location and source type. Additionally, each line v on the list specifies the quantity (available number) of transportation asset v, the particular transport mode(s) on which transportation asset v can move, the loadability feature (pax, cargo, separate pax and cargo, mixed pax and cargo), weight, volume, and lanemeter capacities (if applicable) of transportation asset v, the indices of items that can be carried by transportation asset v (e.g., a truck can carry only boxes or personnel whereas a plane can carry boxes, armored vehicles, and personnel), the ready, loading, unloading, and travel times, fixed and variable costs associated with transportation asset v. Travel times of transportation assets are determined based on regular speeds of transportation assets or based on predetermined speeds, e.g., a convoy speed, on all arcs on which they can move. Additional columns may also be added to the list as necessary to identify other relevant attributes of transportation assets. For example, planes may be grouped as large-body and small-body planes. Such groupings are especially useful in defining capacities on the transportation network.

Table 5. A portion of a list of transportation assets. (TA: transportation asset)

Line Index (v)	Source	Loca-tion	TA Type	Qty	Load Type	Weight Cap. (ton)	Fixed Cost ($) (x1000)	Load/ Unload Time (hour)	Travel Time per km (min)	TA Group (w)
1	Unit A	X	Tank Carrier	10	Cargo	75	125	1	1.2	1
2	Unit A	X	Truck A	10	Cargo	5	50	1	1.2	2
3	Unit A	X	Truck B	5	Cargo	10	75	1	1.2	2
4	Unit B	X	Tank Carrier	10	Cargo	75	125	1	1.2	1
5	Unit B	X	Truck A	10	Cargo	5	55	1	1.2	2
6	Unit B	X	Truck B	5	Cargo	10	80	1	1.2	2
7	Firm A	Y	Truck A	5	Cargo	5	150	1	1.2	2
8	Firm A	Y	Truck B	5	Cargo	10	65	1	1.2	2
9	Firm A	Y	Truck C	5	Cargo	10	90	1	1.2	2

A portion of a list of transportation assets is given in Table 5. The first column in the table gives the index v that is assigned to each line on the list. By convention, v refers to a particular source, location, and type. Lines are indexed by v in such a way that all transportation assets in possession of a military unit or a civilian company are consecutively numbered to form a block. Transportation assets in a given block share a common location (the home base of the military unit or the location of the civilian company that identifies that block). For example, $v=2$ in Table 5 refers to trucks of type A in possession of Unit A at location X while $v=7$ refers to trucks of type A in possession of Firm A at location Y. The last column in Table 5 differentiates transportation assets according to their types only regardless of their location or to which military unit (civilian company) they belong to. For example, $w=1$ refers to tank carriers while $w=2$ refers to trucks. Such groupings are especially useful in defining capacities on the transportation network.

We define $V=\bigcup_{m=1}^{5} V_m$ as the set of transportation asset indices where $V_m, m=1,...,5$ contains the indices of transportation assets belonging to transportation mode m. In the model, we differentiate transportation modes implicitly by allowing transportation assets of different transportation modes to move on the appropriate arcs. For example, if the transportation asset is a ship, then this transportation asset is allowed to move only on the arcs of the sea transportation network.

Transportation assets are classified into four groups depending on their loadability features: *Pax*, transportation assets that can carry only personnel (e.g., buses) (V_{pax}); *Cargo*, transportation assets that can carry only cargo (e.g., trucks, tank carriers, cargo planes) (V_{cargo}); *Pax and Cargo*, transportation assets that carry cargo and personnel in separate compartments (e.g., ships) (V_{both}); and *Mixed Pax and Cargo*, transportation assets that carry cargo and personnel in a single compartment (e.g., trucks, some types of planes) (V_{mix}). While loading cargo and/or personnel onto a transportation asset, depending on the group that transportation asset belongs to, weight, volume, and/or lanemeter capacities of the transportation asset are taken into account such that the total number, weight, volume, and/or length of the loaded cargo and/or personnel onto a transportation asset do not exceed the transportation asset's seat, weight, volume, and/or lanemeter capacities. Here, we focus on transportation assets that are in class V_{mix} as issues related to transportation assets in other classes are clear. (Please see Section 1.2 for more details).

Remember that for a transportation asset in class V_{mix}, the same space is shared by both cargo and personnel and one displaces the other in discrete blocks that can be characterized by a step function. This is due to the fact that seats are built onto the transportation assets in blocks of different sizes, not one by one. Figure 2 illustrates a sample situation. In the figure, the change in volume capacity of a transportation asset against the number of passengers is shown. It is assumed that the volume capacity of the transportation asset is 500 cubic units, that there are 18 seats in a seat block, and maximum passenger capacity of the transportation asset is 90. Thus, each block mounted on the transportation asset consumes 100 cubic units of the volume capacity of the transportation asset. Thus, the remaining volume capacity of the transportation asset can be expressed as an equation as follows: $RemainingVolumeCapacity = 500 - \left\lceil \frac{P}{18} \right\rceil \times 100$ where P is the number of passengers and $\lceil . \rceil$ is the ceiling function.

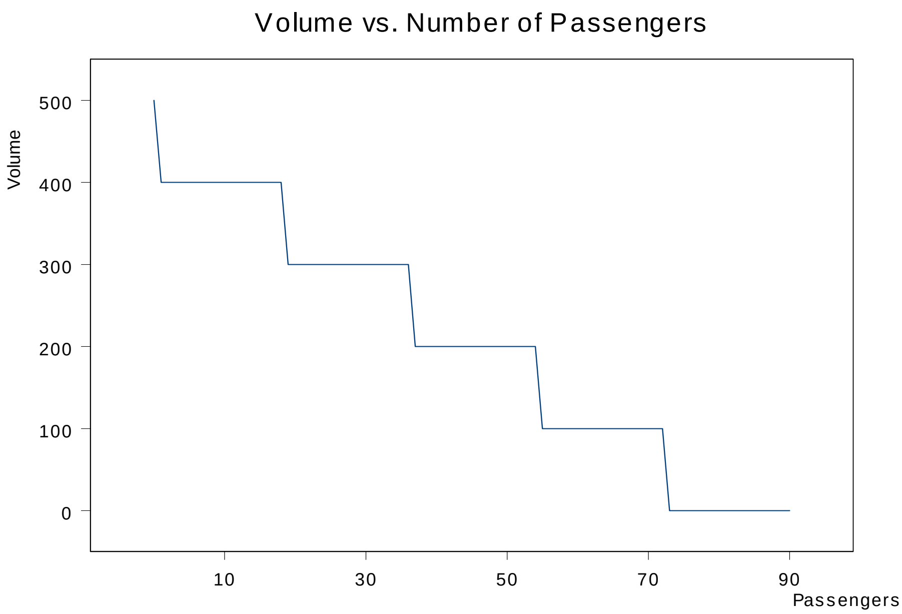

Figure 2. Volume Capacity versus Number of Passengers

To generalize the above situation, we define the parameters below.

$nseatbl_v$ number of seats in a block for a unit of transportation asset v

$CapP_v$ maximum passenger carrying capacity of a unit of transportation asset v

$CapVol_v$ volume capacity of a unit of transportation asset v

For a transportation asset in class V_{mix}, we can define a volume reduction factor. This factor is used in calculating the reduction in volume capacity of the transportation asset depending on the number of passengers (thus the number of blocks of seats added) mounted on it. Let $volredfac_v$ be the volume reduction factor for a transportation asset v when a block of seat is installed onto it. That is, $volredfac_v = \dfrac{nseatbl_v}{CapP_v} \times CapVol_v$.

In the above example, this value is 100. That is, adding one block of seats consumes 100 cubic units of volume capacity. So,

$$RemainingVolumeCapacity_v = CapVol_v - \left\lceil \frac{P}{nseatbl} \right\rceil \times volredfac_v .$$

The exact solution of this problem may be handled in several ways. However, using an approximate solution is also possible. The reason is twofold. First, we need additional variables and constraints to handle the problem that will make the model more complex. Second, such details are disregarded in practice, i.e., if five troops are to be carried on a transportation asset, it is highly likely that one block of seats is not installed and the troops travel mostly seated on other items. That is, gains obtained by adding more complexity to the model may not be significant enough to justify the added complexity of the model.

One approximation is to define a volume for each troop. Note that this value is not fixed but changes depending on the transportation asset, its pax and volume capacities. Using the above parameters, the volume of a passenger for a unit of transportation asset v is computed as

$$paxVolume_v = \frac{CapVol_v}{CapP_v} .$$

This value is multiplied with the number of troops mounted on the transportation asset to find the capacity consumed by the passengers. Then,

$$RemainingVolumeCapacity_v = CapVol_v - paxVolume_v \times P .$$

This logic is used in formulating volume capacity constraints for transportation assets in class V_{mix}. As to the weight capacity constraints for these transportation assets, it is easy because the change in weight capacity is linear. Thus, decreasing weight capacity for each mounting passenger by a standard passenger weight, i.e., $paxweight$, is sufficient.

The issue of *self-deployable* items, i.e., items that are to be treated as transportation assets on some parts of the transportation network, (see Section 1.1 for details) is solved by defining dummy transportation assets that are allowed to carry only self-deployable items and setting weight and volume capacities of the artificially-defined transportation assets depending on whether they are capable of carrying other items as well. If a self-deployable item cannot transport other items (e.g., a tank), then weight and volume capacities of the corresponding artificially-defined transportation asset are set to the weight and volume of the self-deployable item. If a self-deployable item of a unit can carry other items (e.g., a truck), then the items of the unit that can be carried by the self-deployable of the unit are firstly planned to be moved on the self-deployable (taking weight and volume capacities of the self-deployable into account). That is, the load of a self-deployable item, if any, is predetermined. In developing deployment plans, a self-deployable and its load is regarded as a single entity with a certain weight and volume. The weight and volume of the corresponding artificially-defined transportation asset is set accordingly.

To take into account loading and unloading times of transportation assets in the formulation, we define different travel times for empty and loaded transportation assets. We assume that the travel time of a unit of transportation asset v on arc l when it is loaded, trv_{lv}^{loaded}, includes the travel time of a unit of transportation asset v on arc l when it is empty, trv_{lv}^{empty}, plus loading and/or unloading times of transportation asset v if the tail and/or head node of the arc is a source, destination, or transfer point. For example, if a unit of loaded transportation asset v leaves a source node i on an arc $l \in AF_i$, then trv_{lv}^{loaded} includes trv_{lv}^{empty} plus loading time of v. Similarly, if a unit of loaded transportation asset v arrives at a demand node i on an arc $l \in AB_i$, then trv_{lv}^{loaded} includes trv_{lv}^{empty} plus unloading time of v. If both head and tail nodes are transshipment nodes, then trv_{lv}^{loaded} is the same as trv_{lv}^{empty}. This issue will be made clearer when the second modification to the network is explained in the coming paragraphs.

We assume that the transportation network $\tilde{G}=(\tilde{N},\tilde{A})$ is node-wise capacitated. Node capacities are expressed in terms of the number of transportation assets that can pass through the node per unit time and are generally defined by working capacities of the handling and loading/unloading equipment and personnel available at that node. They can be taken as infinity for most nodes, but finite capacities are generally assigned to source, demand, and transfer points as well as to critical nodes such as major intersections and bridges.

The model proposed here is arc-wise uncapacitated. One reason for this is that arc capacities, if present, can easily be accommodated by introducing artificial dummy nodes on arcs as necessary and assigning appropriate node capacities to the artificial nodes. The other reason for leaving out arc capacities is that arcs of the sea and air transport networks essentially have unlimited capacity except for those rare cases where an air corridor or a shipping lane might be expected to carry a heavy traffic load leading to possible congestion. This may be the case, for example, if intense sea traffic is expected through the Bosphorous that connects Marmara Sea to Black Sea in İstanbul. In such a case, redefining the node capacities at both ends of the Bosphorous will correctly impose an upper limit on the sea traffic allowed to pass through the Bosphorous. As for surface transportation, the imposed convoy speeds in ground and rail transportation naturally regulate the traffic in these parts of the network in such a way as not to lead to any congestion when roads or railways are temporarily closed to civilian traffic to allow for free passage of military convoys. These considerations well justify the absence of arc capacities from the proposed model. Nevertheless, arc capacities can easily be handled by the model if needed.

For modeling purposes, we make two modifications on the network $\tilde{G}=(\tilde{N},\tilde{A})$. The first modification is to add a single super node n_d, and a set A_{dum} of directed arcs of the form (n_d,i) for each node $i \in \tilde{N}$ that houses at least one transportation asset. The super node is a dummy node that represents a virtual pool of transportation assets available anywhere in the network. A given transportation asset is drawn from n_d for its first time usage in the system. Note that the dummy arc (n_d,i) is used for initial activation of transportation assets whose initial location is node i. To provide the activation of transportation assets on the correct dummy arc, they are allowed to move only on the dummy arc (n_d,i). This is possible because our indexing convention distinguishes transportation assets based on their initial locations.

Let $V_{init\,i}$ be the set of transportation asset indices v for which the initial location is node i. Let $ready_v \geq 0$ be the ready time of transportation asset v. For example, $ready_v > 0$ represents the contracted time to make transportation asset v available from a civilian company. Let $fixcost_v$ be the fixed cost associated with the initial activation of transportation asset v. For dummy arc (n_d, i), we associate two vectors of size $|V_{init\,i}|$, one representing the ready times and the other representing the fixed costs. Whenever there is a demand for transportation asset v from a node k, this (empty) transportation asset is either directed to node k from the super node n_d (if it has not already been used in the system) or from some node k' at which the last usage of it has terminated. In the former case, the transportation asset traverses the dummy arc (n_d, i) at a cost of $fixcost_v$ and with an arc traversal time of $ready_v$. The empty transportation asset is then routed from i to k.

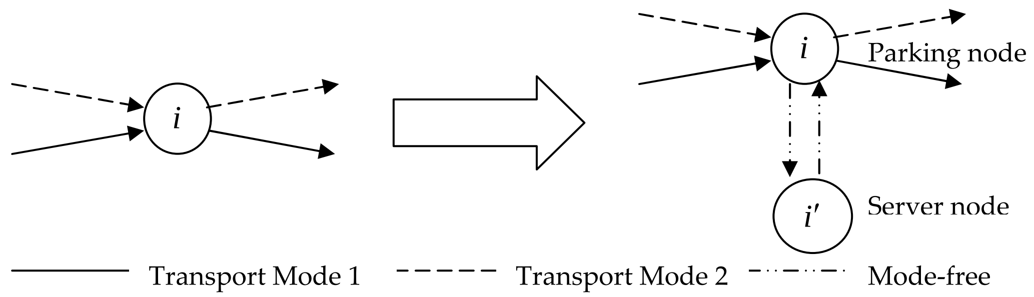

Figure 3. Second modification of the network.

The second modification we make on the network is to add a replica i' for nodes i at which loading/unloading operations are expected to take place, i.e., transfer nodes, sources, and destinations (as necessary). Mode-free directed arcs of the form (i, i') and (i', i) are also added for each replicated node. The modification is depicted in Figure 3. It is convenient to view node i' as a “server” node, where the actual loading/unloading operations take place, and i as a “parking/waiting lot” for a transportation asset. We define N_P as the set of parking/waiting nodes and refer to server nodes with their original names, i.e., source, demand, and transfer nodes. A transportation asset v which is to get service at node i' goes through node i before and after getting service. Define $G=(N,A)$ to be the network obtained from $\tilde{G}=(\tilde{N},\tilde{A})$ after the dummy nodes and arcs are added to it.

This modification of the network helps to define node capacities and handle timing issues appropriately. If a transportation asset is not to get service at node i', it just passes through node i without stopping or spends an idle time, e.g., time to comply with a time-wise constraint, at the node. If a transportation asset is to get service at node i', it must move from node i to node i' before getting service and from node i' to node i after getting service. To incorporate an appropriate service time for a transportation asset at a server node, the associated loading and/or unloading times are taken as travel times of the transportation asset on the arcs (i, i') and (i', i) depending on whether the transportation asset arrives (leaves) at (from) node i' empty or loaded. For example, if a transportation asset leaves loaded from node i' to node i, a non-zero loading time is assigned to the transportation asset whereas a zero load time is assigned if the transportation asset is empty. If node i' is not available for service at any time, the transportation asset spends an idle time at node i until it becomes available. A transportation asset may also spend an idle time at a server node (e.g., waiting for an item to arrive). Such idle times at the nodes are handled by inventory variables appropriately.

From a modeling viewpoint, there are two types of flows on the network: those of items and of transportation assets. The transportation assets are the active agents in that the items cannot flow by themselves unless engaged with transportation assets. In this sense, flows of transportation assets are coupled for a length of time with items and then disengaged again upon arrival at a destination. We may view everything taking place on a time-expanded network to keep track of scheduling and time-dependent issues associated with movements. We define T to be the set of all time periods and TD_c to be the subset of T consisting of time periods at which item type c is allowed to be at a destination. That is, $TD_c = \{t \in T : a_c \leq t \leq b_c\}$ where a_c and b_c are the earliest and latest time at which items of type c are allowed to be at destination.

Given the network settings as described above, together with a list of transportation assets and item lists for active units, the deployment planning problem involves decisions on (1) the route each deployment component is to follow, (2) the schedule of the movement on this route (departure times from home bases, pass times through transshipment and transfer points, arrival times at destinations, load/unload times at origin, destination, and transfer nodes), (3) the transportation assets and the transportation network each component uses on its route, (4) the load compositions of transportation assets allocated to each deployment component, (5) the routings of empty transportation assets subsequent to unloading at destinations, (6) the schedule of the movements of transportation assets while loaded and empty, and (7) a sourcing strategy of transportation assets for a successful deployment of units.

3.2. Model formulation

Based on the modeling conventions mentioned in the foregoing section, we now give a mixed integer programming model, *Cost Minimization Deployment Planning Model* (CMDPM), for planning and executing force deployment. The model handles the deployment process from home bases to destinations addressing issues regarding the scheduling, routing, and use of transportation assets in moving a unit's troops, weapon systems, equipment, and supplies.

The objective function used in the model is to minimize the sum of fixed and variable transportation costs. In this respect, this model is of use, in the first place, for investment decisions regarding transportation assets and transportation infrastructure. The model is also of use for cases in which there is enough time to create deliberate deployment plans that take costs into account.

The constraints in the model can be grouped into

- Flow-balance constraints for transportation assets,
- Node capacity constraints,
- Constraints for coupling transportation assets and items,
- Flow-balance constraints for items ,
- Prioritization/precedence constraints,
- Group/Convoy formation constraints.

Major assumptions of the model are listed below.

A1. Transportation assets, once they leave their origins, are allowed to wait at their final destinations and not required to return to their origins until they are assigned to another unit. This is reasonable as transportation assets are stationed to forward-support units after a deployment. We note that it may also be possible to allocate pooling areas close to final destinations for transportation assets.

A2. Ready, travel, loading, and unloading times are deterministic and discrete.

A3. Deployment plans are made so as to deploy all personnel and cargo. Non-delivery of items is not allowed. This is reasonable, as otherwise some missions in the operational plan cannot be achieved.

Set restrictions that ensure compatibility between transportation assets and items or transportation assets and the transportation infrastructure are omitted in the formulation to avoid notational clutter. Such restrictions are assumed present implicitly.

3.2.1. Cost Minimization Deployment Planning Model (CMDPM)

3.2.1.1. Sets

Sets related to nodes:

N All nodes in the network $G=(N,A)$ $(i, j \in N)$ (Modified network)

N_S Set of nodes that are home bases of the units

N_D Set of nodes that are destinations of units

N_T Set of nodes that are transshipmet nodes for the units

N_{TR} Set of nodes that are possible transfer points (terminals) for units

N_P Set of nodes that represent waiting/parking places at source, demand, and transfer nodes

n_d the super dummy node assumed to hold all transportation assets

Sets related to arcs:

A Set of arcs in the network $G=(N,A)$ $(l, l' \in A)$

A_{dum} Set of dummy arcs in the network

AF_i Set of arcs whose tails are at node i (the forward star of node i)

AB_i Set of arcs whose heads are node i (the backward star of node i)

A_v Set of arcs on which transportation assets of index v are allowed to move

A_c Set of arcs on which items of index c are allowed to be transported

Sets related to deployment components and items:

G	Set of all deployment-component indices ($g, g' \in G$)
C	Set of all item indices ($c, c' \in C$)
C_g	Set of indices of items that are in deployment component g
$CFIRST$	Set of indices of leader items in all deployment components
C_{pax}	Set of indices of items that are troops (personnel)
C_{cargo}	Set of indices of items other than troops ($C_{cargo} = C - C_{pax}$)
C_{lane}	Set of indices of items for which lanemeter capacity is to be taken into account while loading onto a transportation asset
CS_i	Set of indices of items for which node i is a source
CD_i	Set of indices of items for which node i is a destination
CT_i	Set of indices of items for which node i is allowed to be a transshipment point
CTR_i	Set of indices of items for which node i is allowed to be a transfer point
CP_{ic}	Set of indices of items that have a lower priority in arriving at a node i than items of index c (This is derived from the precedence relations between deployment components.)
$C_{node\ i}$	Set of indices of items that are allowed to use node i
$C_{arc\ l}$	Set of indices of items that are allowed to use arc l
$C_{TA\ v}$	Set of indices of items that are allowed to be transported on transportation assets of index v

Sets related to transportation assets:

W Set of indices of transportation asset groupings, e.g., large-body, small-body planes ($w \in W$)

V Set of indices of all transportation assets ($v \in V$)

$V_{gr\,w}$ Set of indices of transportation assets in a group w of transportation assets (e.g., trucks)

V_{pax} Set of indices of transportation assets that can carry only personnel

V_{cargo} Set of indices of transportation assets that can carry only cargo

V_{both} Set of indices of transportation assets that can carry cargo and personnel in separate compartments

V_{mix} Set of indices of transportation assets that can carry cargo and personnel in a single compartment

V_{lane} Set of indices of transportation assets for which lanemeter capacity is to be taken into account (e.g., a ship)

$V_{node\,i}$ Set of indices of transportation assets that are allowed to use node i

$V_{arc\,l}$ Set of indices of transportation assets that are allowed to use arc l

$V_{item\,c}$ Set of indices of transportation assets that can carry items of index c

Sets regarding time periods:

T Set of time periods $(t,\ t' \in T)$

TD_c Subset of time periods at which items of index c are allowed to be at (destination) node i, i.e., $TD_c = [a_c, b_c]$ where a_c and b_c are the earliest and latest times at which items of index c are allowed to be at destination

3.2.1.2. Data

Data related to items:

$weight_c$ weight of one unit of item c

$volume_c$ volume of one unit of item c

$length_c$ length of one unit of item c

$demand_c$ the number of units of item c to be deployed

$stdpaxw$ predetermined, standard weight for a person

e_c the earliest time at which items of index c are allowed to leave their origin, i.e., home base

a_c the earliest time at which items of index c are allowed to be at their destination

b_c the latest time at which items of index c are allowed to be at their destination

Data related to transportation assets:

$fixcost_v$ the cost of activating a unit of transportation asset v for the first time, i.e., a fixed cost for activating a unit of transportation asset v

$trvcostf_{lv}$ travel cost when a unit of transportation asset v moves loaded on arc l

$trvcoste_{lv}$ travel cost when a unit of transportation asset v moves empty (not loaded) on arc l

$availVeh_v$ number of units of transportation asset v initially available at the dummy super node n_d

$WCap_v$ weight capacity of a unit of transportation asset v

$VCap_v$ volume capacity of a unit of transportation asset v

$PCap_v$ pax capacity of a unit of transportation asset v

$LCap_v$ lanemeter capacity of a unit of transportation asset v

$PaxVol_v$ the volume consumed by a passenger on a unit of transportation asset v (this parameter is defined appropriately to handle the issue regarding $v \in V_{mix}$ with a good approximation; $PaxVol_v = VCap_v / PCap_v$)

trv_{lv}^{loaded} travel time of a unit of transportation asset v when it is loaded

trv_{lv}^{empty} travel time of a unit of transportation asset v when it is empty

$ready_v$ ready time of a unit of transportation asset v

Data related to transportation network:

$ParkC_{iw}$ the parking capacity of node i at a time for transportation assets in group w

$SerC_{iw}$ the service capacity of node i at a time for transportation assets in group w

3.2.1.3. Decision Variables

TF_{lvt} number of units of transportation asset v that start moving loaded on arc l at time t

TE_{lvt} number of units of transportation asset v that start moving empty on arc l at time t

CT_{lcvt} number of units of item c that start moving on arc l via a unit(s) of transportation asset v at time t

IV_{ivt} the number of units of transportation asset v remaining at node i at time t

IC_{ict} the number of units of item c remaining at node i at time t

Y_{lct} zero/one variable which is 1, if a unit of item c is assigned to start moving on arc l at time t and 0, otherwise

3.2.1.4. Model

Objective Function

$$\underset{TF,TE,CT,IV,IC,Y}{\text{Minimize}} \sum_{l\in A_{dum},v,t} fixcost_v \times TE_{lvt} + \sum_{l\in A,v,t} trvcostf_{lv} \times TF_{lvt} + \sum_{l\in A,v,t} trvcoste_{lv} \times TE_{lvt} \qquad (1)$$

Constraints

Flow-Balance Constraints for Transportation Assets

$$IV_{ivt} + \sum_{l\in A_{dum}} TE_{lvt} - IV_{iv,t-1} = 0 \qquad i = n_d, v, t \qquad (2)$$

$$IV_{ivt} \equiv availVeh_v \qquad i = n_d, v, t = ready_v \qquad (3)$$

$$IV_{ivt} + \sum_{l\in AF_i} (TF_{lvt} + TE_{lvt}) - IV_{iv,t-1} - \sum_{l\in AB_i} \left(TF_{lv,t-trv_{lv}^{loaded}} + TE_{lv,t-trv_{lv}^{empty}} \right) = 0 \qquad i \in (N - n_d), v, t \qquad (4)$$

Node Capacity Constraints

$$\sum_{v\in V_{grw}} IV_{ivt} + \sum_{l\in AB_i, v\in V_{grw}} \left(TF_{lv,t-trv_{lv}^{loaded}} + TE_{lv,t-trv_{lv}^{empty}} \right) + \sum_{l\in AF_i, v\in V_{grw}} (TF_{lvt} + TE_{lvt}) \leq SerC_{iw} \qquad i \in (N_S \cup N_D \cup N_{TR}), w, t \qquad (5)$$

$$\sum_{v \in V_{grw}} IV_{ivt} \leq ParkC_{iw} \qquad i \in N_P, w, t \qquad (6)$$

Constraints for Coupling Transportation Assets and Items

$$\sum_{c \in C_{cargo}} weight_c \times CT_{lcvt} \leq WCap_v \times TF_{lvt} \qquad l \in (A - A_{dum}), v \in (V_{cargo} \cup V_{both}), t \qquad (7)$$

$$\sum_{c \in C_{cargo}} volume_c \times CT_{lcvt} \leq VCap_v \times TF_{lvt} \qquad l \in (A - A_{dum}), v \in (V_{cargo} \cup V_{both}), t \qquad (8)$$

$$\sum_{c \in C_{pax}} CT_{lcvt} \leq PCap_v \times TF_{lvt} \qquad l \in (A - A_{dum}), v \in (V_{pax} \cup V_{both}), t \qquad (9)$$

$$\sum_{c \in C_{lane}} length_c \times CT_{lcvt} \leq LCap_v \times TF_{lvt} \qquad l \in (A - A_{dum}), v \in V_{lane}, t \qquad (10)$$

$$\sum_{c \in C_{cargo}} weight_c \times CT_{lcvt} + \sum_{c \in C_{pax}} paxweight \times CT_{lcvt} \leq WCap_v \times TF_{lvt} \qquad l \in (A - A_{dum}), v \in V_{mix}, t \qquad (11)$$

$$\sum_{c \in C_{cargo}} volume_c \times CT_{lcvt} + \sum_{c \in C_{pax}} paxVol_v \times CT_{lcvt} \leq VCap_v \times TF_{lvt} \qquad l \in (A - A_{dum}), v \in V_{mix}, t \qquad (12)$$

Flow-Balance Constraints for Items

$$IC_{ict} - IC_{ic,t-1} - \sum_{l \in AB_i, v} CT_{lcv,t-trv_{lv}^{loaded}} = 0 \qquad i \in N_D, c \in CD_i, c \in CFIRST, t \in TD_c \qquad (13)$$

$$IC_{ict} \equiv demand_c \qquad i \in N_D, c \in CD_i, c \in CFIRST, t = b_c \qquad (14)$$

$$IC_{ict} + \sum_{l \in AF_i, v} CT_{lcvt} - IC_{ic,t-1} = 0 \qquad i \in N_S, c \in CS_i, c \in CFIRST, t \geq e_c \qquad (15)$$

$$IC_{ict} \equiv demand_c \qquad i \in N_S, c \in CS_i, c \in CFIRST, t = e_c \qquad (16)$$

$$IC_{ict} + \sum_{l \in AF_i, v} CT_{lcvt} - IC_{ic,t-1} - \sum_{l \in AB_i, v} CT_{lcv,t-trv_{lv}^{loaded}} = 0 \qquad i \in N_{TR}, c \in CTR_i, c \in CFIRST, t \qquad (17)$$

$$\sum_{l \in AF_i} CT_{lcvt} - \sum_{l \in AB_i} CT_{lcv,t-trv_{lv}^{loaded}} = 0 \qquad i \in N_T, c \in CT_i, v, t \qquad (18)$$

$$IC_{ict} + \sum_{l \in AF_i} CT_{lcvt} - IC_{ic,t-1} - \sum_{l \in AB_i} CT_{lcv,t-trv_{lv}^{loaded}} = 0 \qquad i \in (N_P \cup N_S \cup N_D),\ c \in (C - CS_i - CD_i), v, t \qquad (19)$$

Constraints for Component Unity

$$\left(\sum_v CT_{lcvt} \Big/ demand_c \right) = Y_{lct} \qquad i \in (N - n_d),\ l \in AF_i,\ c \in CFIRST,\ t \qquad (20)$$

$$\left(\sum_v CT_{lc'vt} \Big/ demand_{c'} \right) = Y_{lct} \qquad i \in (N - n_d), l \in AF_i, g \in G,\ c \in CFIRST, c \in C_g, c' \in C_g, c' > c, t \qquad (21)$$

Precedence Constraints

$$\left(\sum_v CT_{lc'v,t-trv_{lv}^{loaded}} \Big/ demand_{c'} \right) \le \left(\sum_{l' \in AB_i, v, e_c \le t' < t} CT_{l'cv,t'-trv_{l'v}^{loaded}} \Big/ demand_c \right) \qquad i \in N_D,\ l \in AB_i,\ c \in CFIRST,\ c' \in CFIRST,\ c' \in CP_{ic},\ t \qquad (22)$$

Non-negativity Constraints

$$\begin{aligned} TF_{lvt}, TE_{lvt} &\ge 0 && l,v,t \\ CT_{lcvt} &\ge 0 && l,c,v,t \\ IV_{ivt} &\ge 0 && i,v,t \\ IC_{ict} &\ge 0 && i,c,t \\ Y_{lct} &\ge 0 && \text{binary}, l, c, t \end{aligned} \qquad (23)$$

Initial Conditions

$$IT_{ivt} \equiv 0 \qquad i \in N_T, v, t \qquad (24)$$

Objective function (1) minimizes the sum of fixed and variable transportation costs. A fixed cost is incurred when a transportation asset is drawn into the system for the first time from the pool of transportation assets and a variable cost is incurred when it circulates through the network either empty or loaded.

Constraints (2), (3), and (4) ensure the flow-balance of each transportation asset at the super node (the pool of transportation assets) and at the remaining nodes for each time period. The initial inventory levels of transportation assets at the super node are set to their available numbers at their ready times. Hence, the number of units of transportation asset v drawn from the pool is at most equal to the available number of v. Constraints (4) provide the flow-balance of transportation assets after they are drawn from the super node. As an inventory of transportation assets is not allowed at transshipment nodes, the initial condition (24) is stated accordingly.

Constraints (5) and (6) ensure that node capacities are observed at server and parking nodes at each time period. Node capacities are defined by the maximum number of transportation assets in each transportation asset group w (ground vehicles, large-body planes, etc.) that can get service and wait/park at a node for each time period. Recall that each source, demand, and transfer node is split into a server node and a waiting/parking node and that a transportation asset goes through a waiting/parking node before and after receiving service.

Constraints (7) -(12), which are of identical form but expressed differently as dependent on the loadability group that the transportation assets belong to, are used to couple transportation assets and items by taking into account number, weight, volume, and lanemeter capacities on each arc at each time point.

Constraints (13)-(19) ensure the flow-balance of each item at each node for each time period. Constraints (13)-(14) and (15)-(16) are demand and supply constraints, respectively. Notice that items are allowed to arrive at (leave from) their destinations (home bases) within their allowable time windows (after their earliest departure times). Constraints (13) and (17) are needed for each item to allow a coupling of that item with a transportation asset at the source and at transfer nodes while constraints (18)-(19) are needed for each item and transportation asset to disallow a new coupling at the transshipment and parking nodes as well as at the source and demand nodes of other items. Note that an inventory of items is not allowed at transshipment nodes either.

Constraints (20) and (21) require that all items in a deployment component move as a whole. Constraints (20) are the usual all-or-nothing constraints that ensure that the amount of items on the move at a time be either zero or equal to the whole quantity available of that item. Constraints (21) ensure that when an item of a component starts moving at a time period, so do the other items of the component.

Constraints (22) provide precedence relations between items in arriving at a node (and hence between deployment components). These constraints establish time-wise dependency relations between different items. For an item to arrive at a node through a specific arc at a given time, all items with a higher priority than that of the item under consideration must arrive at that node through one of the arcs incoming to that node prior to that time. This must be checked on all incoming arcs to the node for each time period by taking into account different travel times on the arcs. Constraints (22) can also be used to require a certain time span between different items, i.e., use $e_c \le t' < t - t_{span}$ instead of $e_c \le t' < t$ in the summation on the right-hand side of the constraint where t_{span} is the time span required between two items. Constraints (21) and (22) can also be expressed using only the binary variables Y_{lct} because $\left(\sum_v CT_{lcvt} \Big/ demand_c \right) = Y_{lct}, \forall l,c,t$. However, we do not use this form as it is computationally more expensive to solve the model with this form than with the form presented above.

Notice that constraints (13)-(17) and (20)-(22) are expressed only for leader items instead of all items. Constraints (21) establish the dependency between leader and non-leader items in each deployment component. Flow-balance constraints (18) and (19) are needed for all items to ensure that the set of transportation assets assigned to a deployable unit remains intact during the unit's journey except possibly at transfer points.

The decision variables regarding both items and transportation assets are allowed to take on fractional values. Among these variables, item inventory variables IC_{ict} always take on integral values as items are required to move as a whole. Item flow variables CT_{lcvt} also take on integral values if item c is allowed to be coupled with only transportation asset v due to constraints (20). If item c is coupled with two or more transportation assets, then CT_{lcvt} may take on fractional values. In this case, the solution can easily be modified so that item c is allocated to its assigned transportation assets in integral values. This is possible because the total delivery capacity of transportation assets assigned to carry item c suffices to transport item c (due to unsplittable flow requirement of an item). The decision variables regarding transportation assets may take on fractional values. In this case, the fractional values are rounded up. This is reasonable in the context of the DPP. A unit is required (1) to move as a whole and (2) to be ready at its destination between its earliest and latest arrival times. To satisfy these two requirements, it may be necessary to deploy a unit without fully using the capacities of transportation asset(s) assigned to it. In a sense, being economical is secondary to getting the job done (i.e., delivering the items). Increasing the time windows at which units are to be ready at their destinations and/or reducing the sizes of the deployment components is likely to create more economical solutions. In such a case, the model's solution has a fractional number of transportation assets because its objective is to minimize the sum of the fixed and variable transportation costs. Hence, the degree of error resulting from rounding up is of small scale. Furthermore, our computational results show that the average number of fractional variables in the solutions is about 0.15% of total number of fractional variables in the problems.

The number of nonnegative (fractional) variables in the CMDPM can be expressed as $(2+|C|)\times(|A|\times|V|\times|T|)+(|C|+|V|)\times(|N|\times|T|)$ where $|\,.\,|$ is the cardinality of the associated set. Similarly, the number of binary variables is equal to $|A|\times|G|\times|T|$. Notice that the number of binary variables is not equal to $|A|\times|C|\times|T|$ because the variables are defined only for leader items.

The number of constraints for each constraint group is as follows: $|T| \times (|V| + |N - n_d|) + |V|$ for flow-balance constraints of transportation assets, $(|W| \times |T|) \times (|N_S \cup N_D \cup N_{TR}| + |N_P|)$ for node capacity constraints, $(|A - A_{dum}| \times |T|) \times \lfloor 2 \times (|V_{cargo} \cup V_{both}| + |V_{mix}|) + |V_{pax} \cup V_{both}| \rfloor$ for coupling constraints, $\sum_{c \in CFIRST} [|TD_c| + (|T| - e_c) + 2] + \sum_{i \in N_{TR},\, c \in (CTR_i \cap CFIRST)} |T| + \sum_{i \in N_T,\, c \in CT_i} |V| \times |T| + \sum_{i \in (N_P \cup N_S \cup N_D),\, c \in (C - CS_i - CD_i)} |V| \times |T|$ for flow-balance constraints of items, $\sum_{i \in (N - n_d),\, l \in AF_i} |T| \times (|CFIRST| + |C - CFIRST|)$ for component-unity constraints, and $\sum_{i \in N_D,\, l \in AB_i,\, c \in CFIRST,\, c' \in (CP_{ic} \cap CFIRST)} |T|$ for precedence constraints.

Having given the formulation of the CMDPM, we now compare the NRMO model developed by Baker et al. (1999, 2002) and the CMDPM in more detail. For this purpose, we first give the problem structure in the NRMO and then consider the formulation.

In the NRMO, the single-mode network represents an air transportation network. The node set is composed of aerial ports of embarkation (APOEs), aerial ports of debarkation (APODs), and enroute, forward operating, shuttle bed down, and aerial refueling bases. A deployable unit's pax and cargo are moved from a prespecified APOE to a prespecified APOD or to a forward operating base (FOB). A direct delivery from an APOE to an FOB is possible but cargo destined for an FOB can be dropped at an APOD and then transshipped to the FOB by shuttle aircraft between the APOD and the FOB. An aircraft leaving from an APOE can either go through an enroute base or fly nonstop to the APOD. An aircraft may move through a refueling base and to a shuttle beddown base as needed. With this problem structure, the node set N in the NRMO is composed of source, demand, and transshipment nodes, i.e., $N = N_S \cup N_D \cup N \cup N_T$, while $N = N_S \cup N_D \cup N_T \cup N_{TR} \cup N_P \cup \{n_d\}$ in the CMDPM. Each node in the NRMO is defined such that it is only in one node group, which predetermines how transportation assets and pax/cargo move through the network.

Arcs connecting the nodes in the NRMO are direct flight routes. Routes are distinguished from each other depending on whether it is a direct delivery, transshipment, refueling, empty-return, etc. Thus, an assignment of an aircraft to an arc predetermines its route, the operations it will undergo (e.g., unloading), and the times needed for travel, loading, unloading, refueling, etc. Similarly, an assignment of a deployable unit's pax and cargo to an arc predetermines its movement structure. This greatly simplifies the coordination required between aircraft and pax/cargo (because both aircraft and pax/cargo move through the network according to the predetermined structure). It becomes sufficient to assign transportation assets to arcs and pax/cargo to aircraft. With respect to our set definitions in the CMDPM, the arc set A in the NRMO represents all routes between onload, offload, and transshipment bases. The arc set is classified into subsets such as $A_{dir}, A_{tr,}$ and A_{ret} to represent direct delivery, transshipment, and return arcs. As a result, the underlying network in the NRMO is much simpler than the one in the CMDPM.

In the NRMO, data regarding deployable units are given in the time-phased force deployment data (TPFDD) form. The TPFDD gives, POE, POD, and FOB (if any), available-to-load and required delivery dates, the number of shorts tons of bulk, oversize and outsize cargo, and the number of troops for each deployable unit. With respect to our representation, G denotes the set of deployable units, each of which has four types of items (pax, bulk, outsize, and oversize) to be moved. Thus, a single index c may be associated with each cargo class in each deployable unit G. This allows us to define $C_{pax}, C_{\text{cargo}}, CS_i, CD_i, CT_i, C_{arc\,l}$, and C_{TA} to refer to the same sets in the CMDPM.

In the NRMO, the fleet of transportation assets consists of a fleet of planes. Planes are differentiated based only on their types. With respect to our representation, V represents the set of types of planes. $V_{pax}, V_{\text{cargo}}, V_{mix}$, $V_{node\,i}, V_{arc\,l}$, and $V_{item\,c}$ are modified accordingly.

All decision variables in the NRMO are allowed to take on fractional values. The primary decision variables specify the number of aircraft missions for each deployable unit, for each aircraft type, via each eligible route, in each time period. Another set of decision variables tracks the delivery of short tons of each deployable unit for each cargo class. Additional variables account for empty movements of aircraft, role changes of aircraft, and crew availability. The decision variables in the NRMO and CMDPM are similar with the exception one main difference. In the CMDPM, there is a need to track the movement of pax/cargo through arcs explicitly. In the NRMO, however, this is not required and hence there are no decision variables tracking the amount of pax and cargo moved through an arc. In the NRMO, only delivery of short tons of each deployable unit by an aircraft type is tracked. In a sense, the index l is removed from the definition of CT_{lcvt}, i.e., CT_{cvt}, and index c is added to the definition of TF_{lvt}, i.e., TF_{lcvt} such that

CT_{cvt} represents the amount of c arriving at destination via v in time period t and TF_{lcvt} represents the number of missions of aircraft v to move c through route l in time period t.

The objective in the NRMO is to minimize a weighted sum of penalties for late delivery and nondelivery plus some secondary terms. This can be achieved in the CMDPM by defining appropriate penalties and associating zero ready times for transportation assets, and adding upper bounds on the number of aircraft drawn through the dummy arc that links the super node and the original location of a given type and number of aircraft. Because new aircraft can be added to the fleet at any time, constraints (3) may be modified to take this into account such that the inventory from the previous period is added to the right-hand side and the availability of aircraft is time dependent.

The flow-balance of transportation assets in the NRMO is much simpler than that in the CMDPM. In the CMPDM, constraints (4), which are valid with the second modification, are expressed in such a way as to take into account the fact that the movement patterns of transportation assets at the source, demand, and transfer nodes are not known in advance. In the NRMO, however, the movement patterns of transportation assets are predetermined. Hence, constraints (4) can be split into much simpler flow-balance constraints for source, demand, and transshipment nodes. Specifically, constraints are expressed as $IV_{ivt} + \sum_{l \in AF_i} TF_{lvt} - IV_{iv,t-1} - \sum_{l \in AB_i} TE_{lv,t-trv_{lv}^{empty}} = 0$ for $i \in N_s, v, t$ source nodes and as $IV_{ivt} + \sum_{l \in AF_i} TF_{lv,lv,t-trv_{lv}^{loaded}} - IV_{iv,t-1} - \sum_{l \in AB_i} TE_{lvt} = 0$ for $i \in N_D, v, t$. Note that transshipment nodes act as source and demand nodes for different types of aircraft. Hence, flow-balance constraints at transshipment nodes are the same as above except that they are expressed for appropriate types of aircraft. Flow-balance constraints at nodes such as shuttle beddown and refueling are defined as $IV_{ivt} + \sum_{l \in AF_i} TF_{lvt} - IV_{iv,t-1} - \sum_{l \in AF_i} TF_{lv,t-trv_{lv}^{loaded}} = 0$ for $i \in N_s, v, t$ and as $IV_{ivt} + \sum_{l \in AF_i} TE_{lvt} - IV_{iv,t-1} - \sum_{l \in AF_i} TE_{lv,t-trv_{lv}^{empty}} = 0$ for $i \in N_s, v, t$.

Node capacity constraints (5) and (6) are also valid with the second network modification. In the CMDPM, a node capacity is defined in terms of the number of transportation assets that can get service and park at a time. In the NRMO, however, a node capacity is defined in terms of time periods. This is reasonable because each route through which an aircraft arrives at a node predetermines the aircraft's ground, loading/unloading, and refueling times. Thus, the consumed capacity of a node in the NRMO is obtained by multiplying the number of aircraft arriving at a node with appropriate service times and summing all of them.

Constraints (7) and (12) for coupling transportation assets and items are similar in nature. However, because there are no explicit decision variables tracking the movement of pax and cargo through an arc in the NRMO, the left-hand side of constrainsts is the amount of pax and cargo of a deployable unit moved by a specific aircraft type, i.e., CT_{cvt} for c, v, t, and the right-hand side is the capaciy of the number of missions of that specific aircraft type through all arcs to move associated unit's pax and cargo, i.e., $\sum_l capacity \times TF_{lcvt}$ for c, v, t.

Demand satisfaction constraints (13) and (14) are similar to the ones in the NRMO with the exception that no inventory variables are defined for pax/cargo and non-delivery of items is allowed. Thus, the left-hand side is the sum of the amount of cargo moved by all possible arrival options of pax/cargo over allowable time periods and the amount of non-delivered pax and cargo. Flow-balance constraints at transshipment nodes in the NRMO are essentially the same as constraints (18) in the NRMO except that they are defined only for items and not for item and vehicle pairs. Notice that constraints (13)-(14) and (18) are defined for all items in the NRMO and for only leader items in the CMDPM.

In the NRMO, there are no constraints resembling (15)-(17), (19), and (20)-(22). There are constraints regarding the crew availability, aircraft utilization and aircraft-hours consumption which we do not address in the CMDPM in compliance with the purpose of the model.

3.3. Computational results for CMDPM

Table 6 summarizes the characteristics of the problems generated to test the performance of the CMDPM. In the test problems, three networks of different sizes are used. The numbers of nodes and arcs are, respectively, 13, 18, 25 and 48, 77, and 109. Five problems are generated for each network by setting the number of item indices to 4, 8, 16, 32, and 64. Four deployment components consisting of equal number of item indices are assumed in all problems. The home bases of the components are different for all problems, i.e., four source nodes. There are three destinations for problems 1 through 10 and four destinations for the remaining ones. A precedence relationship is established between two components in arriving at the destination for problems 1 through 10. In all problems, the deployment components can use six different indices of transportation assets that are located at three different locations. A time window of 20 time periods is allocated to units to arrive at their destinations for the smallest-size network and a time window of 40 time periods for the remaining two networks. The time span of the planning is 100 time periods in all problems. The number of fractional and binary variables and the number of constraints in the test problems are given in the table. The GAMS codes of all test problems are given in the CD on the back cover of the dissertation.

Table 6. Characteristics of the generated test problems. (FV : fractional variables, BV : binary variables, C : constraints)

Pr.Id.	$\|N\|$	$\|A\|$	$\|N_S\|$	$\|N_D\|$	$\|G\|$	$\|C\|$	$\|V\|$	$\|TD\|$	$\|T\|$	Number of FV	Number of BV	Number of C
1	13	49	4	3	4	4	6	20	100	189400	19600	60742
2	13	49	4	3	4	8	6	20	100	312200	19600	68308
3	13	49	4	3	4	16	6	20	100	557800	19600	93352
4	13	49	4	3	4	32	6	20	100	1049000	19600	189772
5	13	49	4	3	4	64	6	20	100	2031400	19600	334335
6	19	77	4	3	4	4	6	40	100	296200	30800	78076
7	19	77	4	3	4	8	6	40	100	488600	30800	185041
8	19	77	4	3	4	16	6	40	100	873400	30800	264068
9	19	77	4	3	4	32	6	40	100	1643000	30800	626026
10	19	77	4	3	4	64	6	40	100	3182200	30800	1482804
11	25	109	4	4	4	4	6	40	100	417400	43600	123811
12	25	109	4	4	4	8	6	40	100	689000	43600	279813
13	25	109	4	4	4	16	6	40	100	1232200	43600	607781
14	25	109	4	4	4	32	6	40	100	2318600	43600	1478113
15	25	109	4	4	4	64	6	40	100	4491400	43600	2975420

The computational tests are implemented on a 1.5 GHz PIV PC with 1.5 GB RAM by using ILOG CPLEX 9.0 and by letting the models run until the desired optimality criterion is attained or for eight hours (28800 CPU seconds) at maximum. The solution times at which LP relaxation (at the root node) is solved and 10%, 5%, and 0% deviation from optimality are achieved are recorded in all computational studies.

Table 7 gives the solution times of the CMDPM for problems 1 through 5 where constraints (2) and (3) are not used, i.e., an infinite number of transportation assets of each index is assumed. Table 7 shows that the solution times of CMDPM based on a direct use of CPLEX 9.0 are not good enough to be used in a real-world application. The optimal solutions of the CMDPM are obtained only for problems 1 and 2 in around 5,000 CPU seconds. However, even the root solutions cannot be obtained for problems 4 and 5 in the allocated time. Hence, a solution methodology to improve the solution times of the CMDPM is needed for real-world applications. In what follows, we give the proposed methodology and the solution times based on it.

Table 7. Solution times of CMDPM (CPU seconds). A "-" for a corresponding optimality criterion shows that the branch and bound jumps to a solution with a lower optimality criterion.

Pr.Id.	CMDPM			
	Root	10%	5%	0%
1	10.42	-	-	**48.67**
2	267.22	3623.99	-	**4536.65**
3	2511.03	-	-	**5182.31**
4	11980.34	**NO INTEGER SOLUTION**		
5	**NO ROOT SOLUTION**			

3.4. Proposed solution methodology

The proposed solution methodology includes finding a relaxation and restriction of the CMDPM and then using their solutions in solving the CMDPM. Let z^*_{REL}, $\underline{z_{CMDPM}}$, and x^*_{RES} denote the optimal objective function value of the CMDPM-REL, the lower bound for the optimal objective function value of the CMDPM, and the optimal solution of the CMDPM-RES, respectively. The procedure works as follows: (1) Solve the CMDPM-REL using CPLEX and set $\underline{z_{CMDPM}}$ to z^*_{REL}. (2) Solve the CMDPM-RES using CPLEX and set its solution x^*_{RES} as a starting solution of the CMDPM. The objective function value corresponding to x^*_{RES} is an upper bound on the optimal objective function value of the CMDPM. (3) Solve the CMDPM using CPLEX with the given initial starting solution and lower bound. The details of the solution methodology are depicted in Figure 4.

If the time available to solve the CMDPM is restricted, time limits to solve the CMDPM-REL, CMDPM-RES, and/or CMDPM may be set by the user. Thus, it is possible that some models are not solved to optimality. In such a case, the solution methodology is modified as follows. If the CMDPM-REL is not solved to optimality within the given time limit, the lower bound $\underline{z_{REL}}$ for the optimal objective function value z^*_{REL} of CMDPM-REL obtained at the end of the time limit is set as the $\underline{z_{CMDPM}}$. If the CMDPM-RES is not solved to optimality, then the solution x_{RES} of the CMDPM obtained at the end of the time limit is set as the initial starting solution for the CMDPM. The CMDPM is then solved with the given lower bound and the given initial starting solution. If the CMDPM is not solved to optimality within the set time, then the solution reached at the end of the time limit is used as a feasible solution. Such an approach may be of use when the user wants to have simply as good a solution as is practicable when the time limit is reached.

The CMDPM-REL is obtained by relaxing some requirements in the CMDPM. In the CMDPM, a deployment component is required to move as a whole from its origin to its destination. In the CMDPM-REL, however, this requirement is observed only on the arcs outgoing from (incoming to) source, demand, and transfer nodes, i.e., a deployment component is not required to move as a whole on the intermediate arcs. In the CMDPM, items can move only when coupled with transportation assets. In the CMDPM-REL, this is required only on the arcs outgoing from (incoming to) source, demand, and transfer nodes. Thus, items can move by themselves without being assigned to transportation assets on the intermediate arcs. In the CMDPM, flow-balance of all items is observed at all nodes and coupling of an item and a transportation asset is not allowed to change in the same transport mode. In the CMDPM-REL, however, flow-balance at all nodes is required only for leader items. Non-leader items appear only on the arcs outgoing from (incoming to) source, demand, and transfer nodes. Thus, coupling of a non-leader item and a transportation asset is allowed to change on the same transport mode.

The CMDPM-REL is formulated by restricting some of the constraint-defining sets to their subsets, thereby deleting the related constraints. Specifically, we redefine constraints (7) -(12) and (20)-(21) for $l \in (A_{out} \cup A_{in})$ and constraints (18) and (19) for $c \in CFIRST$. Here, $A_{out} = \bigcup_{i \in (N_S \cup N_D \cup N_{TR})} AF_i$ and

$A_{in} = \bigcup_{i \in (N_S \cup N_D \cup N_{TR})} AB_i$ are the sets of arcs outgoing from source, demand, and transfer nodes and incoming to source, demand, and transfer nodes, respectively.

The CMDPM-RES is obtained by fixing the time periods at which deployment components arrive at destinations and transfer points. The CMDPM-RES is formulated by fixing the values of the decision variables Y_{lct} that correspond to $l \in A_{in}$ and IC_{ict} in the CMDPM. In deciding the values of the decision variables to fix, the solution of the CMDPM-REL is used. The values of the decision variables Y_{lct}, $l \in A_{in}$, and IC_{ict} in the solution x_{REL} of CMDPM-REL are set to the values of the corresponding decision variables in the CMDPM. The CMDPM-RES determines the routes and schedules of the deployment components and the allocation of transportation assets and transportation infrastructure given the arrival times of the deployment components at destinations. Hence, the CMDPM-RES is solved to optimality in a very short time.

One main question is whether the CMDPM-RES may be infeasible or not. There is a possibility of infeasibility when the number of transportation assets is not sufficient to make the deployment components ready at their fixed times. However, this is a strong indication, based on our empirical studies, that the CMDPM itself is infeasible because the availability of transportation assets is checked in the CMDPM-REL. In such a case, our suggestion is to change data, e.g., increase the number of available number of transportation assets or change the time windows at which units are to be at their destinations, and restart the solution methodology.

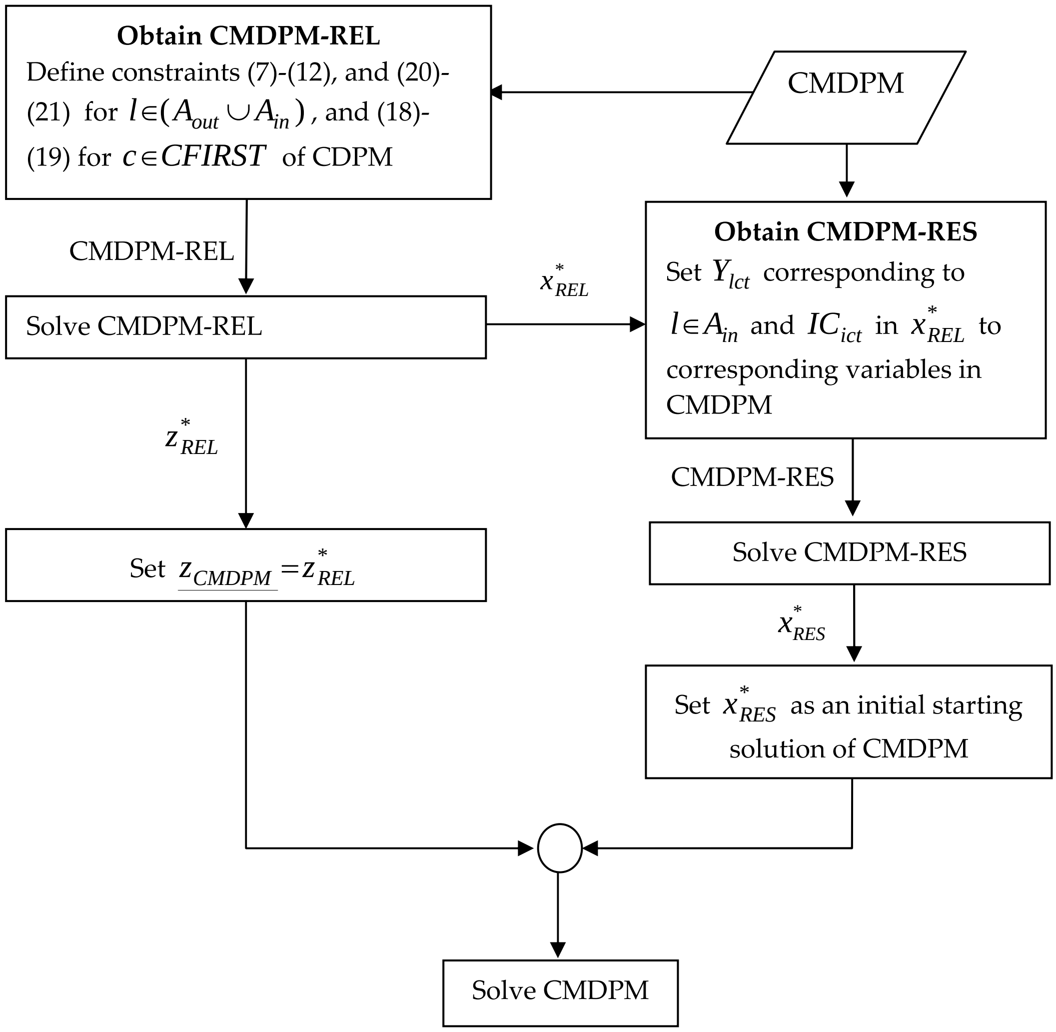

Figure 4. The solution methodology

3.5. Computational results for the proposed solution methodology

We test the performance of the proposed solution methodology using the test problems 1 through 15 defined in Table 6. The model is tested for both an unlimited and limited supply of transportation assets.

Table 8 and Table 9 give the solution times in CPU seconds for unlimited and limited fleet sizes, respectively, for CMDPM-REL and CMDPM-RES. The solution times for CMDPM are not presented because the optimal objective function values of the CMDPM-RES are either equal to or slightly greater than (the difference is less than 0.01%) those of the CMDPM-REL and the optimal solutions of CMDPM-RES are feasible for CMDPM for all problems. That is, the optimal solutions of the CMDPM-RES are also optimal for the CMDPM.

Table 8. Solution times of CMDPM-REL and CMDPM-RES (CPU seconds). A "-" for a corresponding optimality criterion shows that the branch and bound jumps to a solution with a lower optimality criterion.

Pr.Id.	CMDPM-REL				CMDPM-RES	**Total Solution Time**
	Root	10%	5%	0%	0%	
1	10.64	348.21	373.37	405.91	7.69	**413.60**
2	67.94	-	-	792.12	8.56	**800.68**
3	47.25	-	54.40	197.17	4.64	**201.81**
4	92.48	-	97.88	553.61	5.63	**559.24**
5	112.92	-	115.437	574.10	4.28	**578.38**
6	80.67	-	-	115.60	3.91	**119.51**
7	314.23	-	-	362.34	5.08	**367.42**
8	456.25	-	-	514.73	6.97	**521.70**
9	344.76	-	-	457.54	12.90	**470.44**
10	609.78	-	-	612.36	23.94	**636.30**
11	176.60	-	-	472.06	6.68	**478.74**
12	567.06	-	-	854.23	5.78	**860.01**
13	562.66	-	-	565.47	12.07	**577.54**
14	554.52	-	-	557.93	22.35	**580.28**
15	805.44	-	-	1035.90	45.27	**1081.17**

The results in Table 8 and Table 9 can be considered in two groups, one for the results of problems 1-5 and the other for the results of problems 6-15. The total solution times for the latter group range from 119.51 (Pr. 6) to 1081.17 (Pr. 15) CPU seconds in Table 8 and from 193.17 (Pr. 6) to 4005.64 (Pr. 13) CPU seconds in Table 9. On the average, the solution times of the CMDPM-REL account for more than 98% of the total solution times. The times at which the root solution and the optimal solution of the CMDPM-REL are reached are very close to each other because the optimal solutions are obtained at the root node or after fathoming a small number of nodes in the branch and bound algorithm. No solution times of the CMDPM-REL are presented in the fields corresponding to the optimality criteria of 10% and 5% because the branch and bound jumps directly to the optimal solution from the first integer solution with a 20-30% gap between the lower and upper bounds.

Table 9. Solution times of CMDPM-REL and CMDPM-RES (CPU seconds). A "-" for a corresponding optimality criterion shows that the branch and bound jumps to a solution with a lower optimality criterion.

Pr.Id.	CMDPM-REL				CMDPM-RES	**Total Solution Time**
	Root	10%	5%	0%	0%	
1	12.834	961.50	1098.66	1404.79	5.76	**1410.55**
2	96.55	-	-	1270.85	7.21	**1278.06**
3	109.75	115.42	-	499.82	3.45	**503.27**
4	86.27	98.36	-	572.47	4.43	**576.90**
5	153.95	156.39	-	965.26	6.87	**972.13**
6	130.17	-	-	182.10	11.07	**193.17**
7	498.36	-	-	519.33	16.02	**535.35**
8	918.48	-	-	935.56	28.07	**963.63**
9	836.21	-	-	881.91	46.44	**928.35**
10	1169.78	-	-	1203.05	98.48	**1301.53**
11	253.70	-	-	2983.39	21.99	**3005.38**
12	1597.69	-	-	3755.53	31.44	**3786.97**
13	2881.34	-	-	3959.07	46.57	**4005.64**
14	2238.60	-	-	2310.08	84.79	**2394.87**
15	2276.52	-	-	2375.29	191.33	**2566.62**

The total solution times for problems 1-5 range from 201.81 (Pr. 3) to 800.68 (Pr. 2) CPU seconds in Table 8 and from 503.27 (Pr. 3) to 1410.55 (Pr. 1) CPU seconds in Table 9. As in problems 6-15, the solution times of the CMDPM-REL account for more than 98% of the total solution times. However, unlike in the problems 6-15, the times at which the root solutions and the optimal solutions of the CMDPM-REL are reached are not close. The average ratios of the root solution times to total solution times of the CMDPM-REL for problems 1-5, 6-10, and 11-15 are about 12%, 84%, and 76% in Table 8 and 12%, 90%, and 80% in Table 9. The branch and bound algorithm finds a first integer solution in a very short time (in seconds) after the root solution. The gap between the lower and upper bounds is about 20% for problems 1-2 and between 5% and 10% for problems 3-5. The upper bounds, i.e., the first integer solutions, are very close to the optimal solutions and the branch and bound algorithm finds a first integer solution in a very short time (in seconds) after the root solution. However, the branch and bound spends around 90% of the total solution times to increase the lower bounds. For example, for problem 1 in Table 8 (Table 9), the difference between the integer solution obtained at 38.55 (22.26) seconds and the optimal solution obtained at 405.91 (1404.79) seconds is less than 0.001%. This combined with the results for problems 6-15 leads one to think that the optimal solutions are close to the upper bounds.

The solution times in Table 9 are worse than those in Table 8. This is expected. The worst case is for problem 13 where the solution time in Table 8 is about one seventh of its corresponding solution time in Table 9 while the best case is for problem 4 where the solution time in Table 8 is about 97% of its corresponding solution time in Table 9.

There are no preceding results to compare with ours. However, solution times in Table 8 and Table 9 are really encouraging for a real world application as it is known by experience that it may take a planner about one week to come up with a feasible, not optimal, detailed deployment plan for the size of the problem that we deal with.

The solution times in all tables are obtained under certain parameter settings of the CPLEX. Our experience shows that using the bound strengthening has the effect of worsening the solution times almost ten times. Using cuts also has an adverse effect on the solution times. The aggressive scaling parameter improves the solution times of the CMDPM-REL but worsens those of the CMDPM-RES and the CMDPM. The algorithm used to solve the LP relaxations at the nodes is primal simplex for the CMDPM-REL and CMDPM and dual simplex for the CMDPM-RES. The steepest-edge pricing is good for the CMDPM-REL while the devex pricing is good for the CMDPM-RES and the CMDPM. In addition, best bound node strategy in selecting a node, and pseudo costs in selecting branching variable improves the solution times of the models.

Computational results show that the CMDPM-REL and the CMDPM-RES obtained from the solution of the CMDPM-REL provides so good a lower bound and an upper bound, respectively, that an optimal (near-optimal) solution to the CMDPM can be obtained without having to solve the CMDPM. Why the CMDPM-REL provides so good a bound is explained as follows. In the CMDPM, the movement of a component as a whole, the coupled movements of items of a component and transportation assets, and tracking the movements of all items of a component are required on all arcs from the origin to the destination of a component. In the CMDPM-REL, these are required only on the arcs adjacent to the origin, the destination, and possible transfer points of a component. In this regard, the number of transportation assets used to deploy a component at the origin and transfer points of a component is essentially the same in both CMDPM and CMDPM-REL; however, the routes they follow in CMDPM and CMDPM-REL may differ. With respect to cost, this means that fixed costs incurred in both CMDPM and CMDPM-REL are essentially the same; however, variable costs may differ. Thus, the objective function values of CMDPM and CMDPM-REL are close to each other as their objective functions are the same. What is done in the CMDPM-RES is actually to correct the relaxed requirements. In this regard, we can mainly focus on improving the solvability of the CMDPM-REL.

3.6. Conclusion

In this chapter, we have abstracted the DPP and developed the CMDPM, *Cost Minimization Deployment Planning Model*, to solve it. The solution times for CMDPM based on a direct use of CPLEX 9.0 are not good enough to be used in a real-world application. In this regard, we have proposed a solution methodology that involves an effective use of a relaxation and restriction strategy that significantly speeds up a CPLEX-based branch and bound. The solution times for intermediate sized problems are around one hour at maximum, whereas it takes about a week in the Turkish Armed Forces to produce a suboptimal feasible solution based on trial-and-error methods for a problem of the same size.

The CMDPM aims to develop deployment plans with minimum cost. The model can be used for evaluating and assessing investment decisions in transportation infrastructure and transportation assets as well as for planning and execution of cost-effective deployment operations. This model is of use in cases where carrying out a deployment operation in a short time is not important, i.e., the operation is not imminent, and cost is of primary concern.

In the next chapter, we give two models that will be of use in cases where carrying out deployment operation in a short time is of utmost concern and cost is of secondary concern.

CHAPTER 4

LATENESS AND TARDINESS MINIMIZATION DEPLOYMENT PLANNING MODELS

In Chapter 3, we have introduced the CMDPM, Cost Minimization Deployment Planning Model. The CMDPM is expected to be of use in cases where there is a need to create cost-effective deployment plans. Such planning needs may occur when the operation is not imminent and hence there is enough time to create deliberate deployment plans taking cost into account. Examples of such planning needs occur when updating operations plans for potential contingencies on a routine basis or when deploying units for exercises, international operations, e.g., peace support operations, or other purposes during peacetime. The CMDPM is also expected to be of use when evaluating and assessing investment decisions regarding transportation infrastructure and transportation assets.

Although plans are made for possible contingencies from peacetime, events at an actual crisis typically do not evolve as predicted in contingency plans. In fact, there may arise a contingency which is not at all considered in the peacetime. In the former case, a refinement or a complete change of operations plans may be required, which leads to changes in deployment plans. In the case of a new contingency not considered at peacetime, a new operations plan must be prepared from scratch. In both cases, the approach used in planning is different from the one used at peacetime.

During a crisis, the success of the operation is the main planning consideration. Because the timely movement of units is the primary measure for success, time is of primary concern while cost is of secondary concern. In contrast, because cost is the main planning consideration at peacetime, the deployment plans offered by the CMDPM are expected to target at minimum amount of resources used. As a result, the CMDPM will suggest deployment plans that use the whole range of time windows of deployable units to the extent possible, i.e., arrival times of many deployable units at their destinations may be close to the upper bounds of their allowable time windows. Hence, such plans leave little or no time to the units to compensate for possible delays caused by unexpected events (such as malfunctions of equipment/transportation assets or enemy attacks on some parts of the transportation network). Such delays are likely to cause some deployable units not to arrive at their destinations within their time windows. In this regard, deployment plans that draw the arrival times at destinations towards the lower bounds of time windows may be more valuable during wartime to protect against possible delays caused by unexpected events. Why plans from peacetime are not prepared with this in mind is due to the fact that several possible contingencies are considered simultaneously during peacetime and that resources must be allocated to several contingencies such that a balance is established. However, in actuality only one or two contingencies are realized at the same time. Thus, all national resources can be directed towards thwarting off the imminent, realized threat(s). In addition, the nature and extent of the threat may directly dictate that units be deployed to their destinations in the shortest possible time within their time windows. Such situations may usually occur after units' initial deployments to their destinations, e.g., a redeployment of a unit to another operations area may be needed. In this regard, deploying units to their destinations as soon as possible may be of primary concern to the planners during a crisis. We refer to this criterion as that of *minimizing the maximum lateness*. This can be achieved by minimizing the maximum of the differences between the units' arrival times at their destinations and their earliest allowable arrival times.

Although plans are made to deploy all units to their destinations within their allowable time windows, there may be cases that prevent this goal from being realized. For example, national resources may not be sufficient or some events may cause delays in the deployment of units. This means that some deployable units arrive at their destinations after their allowable latest arrival times. In such cases, the logical objective for operations planners is to draw as much as possible the arrival times of units to their latest allowable arrival times whenever such arrivals occur later than the latest permissible arrival time. We refer to this objective as the *minimization of tardiness*. This can be achieved by minimizing the maximum of the differences between units' arrival times at their destinations and their latest allowable arrival times.

In this chapter, we propose two min-max models, one minimizes lateness and the other minimizes tardiness, to address the above cases, and report computational results based on the solution methodology given in Chapter 3.

4.1. Modeling Time

The formulations of both models are based on the modeling structure of the CMDPM. Hence, decision variables and constraint sets in the CMDPM with minor changes are used in the formulations. This allows us to use the solution methodology developed for the CMDPM.

In the CMDPM, there are no decision variables tracking the departure/arrival times of units from/at the nodes of transportation network. Hence, there is a need to extract arrival times without changing the modeling structure of the CMDPM.

Different formulations are possible to extract arrival times using the modeling structure in the CMDPM. Actually, it is possible to extract arrival times without defining any additional variables. For example, the arrival time of an item c (items of index c) at node i, AT_{ic}, may be defined as $AT_{ic} = \left(\sum_{l \in AB_i, v, t} t \times CT_{lcv,t-trv_{lv}^{loaded}} \middle/ demand_c \right)$. To justify this expression, observe first that item c is required to move as a whole by constraints (20) so that item c activates only one of the arcs incoming to i, i.e., $l \in AB_i$, only once. Thus, $\left(CT_{lcv,t-trv_{lv}^{loaded}} \middle/ demand_c \right)$ takes the value of 1 only once. This is why the summation is over all time periods and all incoming arcs at i, $l \in AB_i$. Although it is possible to use this given formulation or other formulations using only the decision variables of the CMDPM to extract arrival times, our experimentation with different formulations has led us to use the following formulation for better computational results. The formulation is closely related to the one given above and is as follows:

$$\left(\sum_{l \in AB_i, v} CT_{lcv, t-trv_{lv}^{loaded}} \Big/ demand_c \right) = X_{ict} \qquad i, c, t \tag{25}$$

$$AT_{ic} = \max_t \; t \times X_{ict} \qquad i, c \tag{26a}$$

$$AT_{ic} = \sum_t t \times X_{ict} \qquad i, c \tag{26b}$$

$$AT_{ic} \geq t \times X_{ict} \qquad i, c, t \tag{26c}$$

In this formulation, X_{ict} is a binary variable that takes on the value of 1 if item c arrives at node i at time t and 0 otherwise. Due to the unsplittable flow requirement provided by constraints (20), all terms in the summation on the left-hand-side of (25) are zero except one term which is 1. Consequently, X_{ict} takes on the value of 1 for exactly one t value while it is 0 for all other values of t. Then, there remains only to multiply X_{ict} with the time index to extract the correct arrival time for item c at node i. This is done by either constraint (26a), (26b), or (26c).

4.2. Lateness Minimization Deployment Planning Model (LMDPM)

In this section, we introduce the *Lateness Minimization Deployment Planning Model* (LMDPM). The objective in the LMDPM is to minimize the maximum lateness and hence is of min-max type. The purpose is to minimize the maximum of the differences obtained by subtracting the units' earliest allowable arrival times at their destinations from the arrival times of units determined in the model. The model aims to develop a deployment plan for deployable units in which the arrival times of units at their destinations are as much close to the lower bounds of their allowable time windows as possible.

All the modeling artifacts used in the development of CMDPM are valid in the development of LMDPM. Thus, all sets, data, and decision variables regarding the network, deployable items, transportation assets, and time periods are used as defined in Chapter 3. Now, we give the formulation of the LMDPM.

4.2.1. Model (LMDPM)

Objective Function

$$\underset{TF,TE,CT,IV,IC,Y,X}{\text{Minimize}} \quad \underset{i\in N_D,\, c\in(CFIRST\cap CD_i),\, t\geq a_c}{\text{Max}} \left(t\times X_{ict} - a_c\right) \tag{27}$$

Constraints

In addition to Constraints (2)-(24),

$$\left(\sum_{l\in AB_i, v} CT_{lcv, t-trv_{lv}^{loaded}} \Big/ demand_c \right) = X_{ict} \qquad i\in N_D,\, c\in(CFIRST\cap CD_i),\, t\geq a_c \tag{28}$$

$$X_{ict} \in \{0,1\} \qquad i, c, t \tag{29}$$

The LMDPM is obtained by using the first option (26a) for modeling time. The objective function (27) minimizes the maximum of the differences between arrival times of leader items at their destinations and their earliest allowable arrival times. This is equivalent to minimizing the maximum lateness of deployment components as well as that of units because a deployment component can be tracked only by its leader item and the arrival time of a unit at its destination is equal to the largest of the arrival times of its components. Notice that the term $t \times X_{ict}$ gives the arrival time of the leader item c of a component as explained in constraints (26a) through (26c). Subtracting earliest allowable arrival time a_c of the leader item c at its destination from its earliest arrival time, $t \times X_{ict}$, defines the deviation of the arrival time of the component from its earliest allowable time. Hence, minimizing the maximum of these differences requires that the arrival times of deployment components at their destinations be as close to their earliest allowable times as possible. Constraints (28) are the same as constraints (25) except that they are defined for restricted sets in compliance with the objective function.

Because the objective function is nonlinear, the resulting model is a nonlinear mixed integer program. However, it can easily be linearized by the usual linearization method used for min-max objective functions.

The term $\underset{i \in N_D, c \in (CFIRST \cap CD_i), t \geq a_c}{\text{Max}} \left(t \times X_{ict} - a_c \right)$ in the objective function (27) is equal to the smallest number z that satisfies $z \geq \text{Max}\left(t \times X_{ict} - a_c \right)$ for $i \in N_D, c \in (CFIRST \cap CD_i), t \geq a_c$. For this reason, LMDPM is equivalent to the *Lin-LMDPM*, *Linearized* LMDPM, given below.

Linearized LMDPM (Lin-LMDPM)

Objective Function

$$\underset{TF,TE,CT,IV,IC,Y,X,z}{\text{Minimize}} \quad z \tag{30}$$

Constraints

In addition to constraints (2)-(24) and (28)-(29),

$$z \geq t \times X_{ict} - a_c \qquad i \in N_D, c \in (CFIRST \cap CD_i), t \geq a_c \tag{31}$$

$$z \geq \sum_{t \geq a_c} (t \times X_{ict}) - a_c \qquad i \in N_D, c \in (CFIRST \cap CD_i) \tag{32}$$

In the Lin-LMDPM, constraints (32) are not actually the result of the linearization of the LMDPM. Constraints (32) are obtained by summing over time component on the right-hand-side of constraints (31) and, accordingly, dropping time component from the sets for which constraints (31) are defined. Constraints (32) do not change the feasible space due to the unsplittable flow requirement that ensure that the variable X_{ict} take on the value of 1 only once for a leader item c for all time periods. Notice that if the LMDPM is modeled using the second option (26b) for modeling time, the linearization process ends up only with constraints (32). In this regard, constraints (31) and (32) can be used stand-alone as well as together.

The preference to use both constraints (31) and (32) in the Lin-LMDPM is for computational improvement. Our experience has shown that Lin-LMDPM with only one of the constraints (31) and (32) is computationally bad. When constraints (31) are used alone, the objective function value of the LP relaxation of the Lin-LMDPM, i.e., the lower bound, is too low. Hence, it takes a lot of time to increase the lower bound. When constraints (32) are used alone, the lower bound increases dramatically. However, the summation over all time periods creates too many fractional values in the solution. When constraints (31) and (32) are used together, solution times decrease significantly.

To see how constraints (32) increase the lower bound, assume that the earliest time for a leader item c (unit) to arrive at its destination is 60. Assume also that in the solution of the LP relaxation, $X_{ict}=0.5$ for $t=65$ and $t=70$, i.e., unsplittable flow requirement is violated. Thus, without constraints (32), the right-hand-side of constraints (31) is $65\times0.5-60=-27.5$ and $70\times0.5-60=-25$ for $t=65$ and $t=70$, respectively. On the other hand, the right-hand-side of constraints (32) is $(65\times0.5+70\times0.5)-60=7.5$. It is clear that constraints (32) dramatically increase the lower bound.

4.2.2. Computational Results

We test the performance of the Lin-LMDPM by using the same test problems, computer, solver, stopping criteria, and solution methodology used for testing the performance of the CMDPM.

The relaxation Lin-LMDPM-REL and the restriction Lin-LMDPM-RES of the Lin-LMDPM correspond to the relaxation CMDPM-REL and the restriction CMDPM-RES of the CMDPM, respectively. The Lin-LMDPM-REL is obtained in the same way the CMDPM-REL is obtained, i.e., redefine constraints (7) -(12) and (20)-(21) for $l\in(A_{out}\cup A_{in})$ and constraints (18) and

(19) for $c \in CFIRST$ and keep the objective function and all other constraints in Lin-LMDPM as is. The formulation of the Lin-LMDPM-RES is the same as that of the CMDPM-RES with the exception that the values of the decision variables X_{ict}, $i \in N_D$, are also fixed in addition to the values of the decision variables Y_{lct} that correspond to $l \in A_{in}$ and IC_{ict} in the Lin-LMDPM.

When the fleet size is not restricted, it is clear that the objective function is zero because the model fixes the arrival times of deployable units to their earliest allowable times by drawing as many transportation assets as needed from the pool of transportation assets. The solution times for this usage of the model are in seconds and hence not presented.

Table 10 gives the solution times in CPU seconds for limited fleet sizes for the Lin-LMDPM-REL and the Lin-LMDPM-RES. The solution times for the LMDPM are not presented because the optimal objective function values of the Lin-LMDPM-RES are equal to those of the Lin-LMDPM-REL and the optimal solutions of the Lin-LMDPM-RES are feasible for the Lin-LMDPM for all problems. That is, the optimal solutions of the Lin-LMDPM-RES are also optimal for the Lin-LMDPM.

The total solution times in Table 10 range from 37.70 (Pr. 1) to 2104.85 (Pr. 15) CPU seconds where the solution times of the Lin-LMDPM-REL account for more than 98% of the total solution times of the Lin-LMDPM as in the case of the CMDPM. The root solutions of the Lin-LMDPM-REL are obtained in relatively short times. However, a first integer solution is not found in a very short time after the root solution unlike in the case of the CMDPM. A first integer solution is generally found in later phases of the branch and bound algorithm. The gaps between the first integer solutions and the lower bounds range from 60% to 90%. It turns out that that the optimal solutions are generally close to the first integer solutions and a significant amount of time is spent to increase the lower bound.

Table 10. Solution times of Lin-LMDPM-REL and Lin-LMDPM-RES (CPU seconds). A "-" for a corresponding optimality criterion shows that the branch and bound jumps to a solution with a lower optimality criterion.

Pr.Id.	Lin-LMDPM-REL				Lin-LMDPM-RES	**Total Solution Time**
	Root	10%	5%	0%	0%	
1	1.66	-	-	36.58	1.12	**37.70**
2	6.14	-	-	37.16	1.17	**38.33**
3	4.30	-	-	55.75	3.67	**59.42**
4	16.56	-	-	319.13	4.94	**324.07**
5	43.36	-	-	491.11	5.64	**496.75**
6	9.77	-	155.34	306.00	2.98	**308.98**
7	58.52	-	-	178.00	4.66	**182.66**
8	121.84	-	-	973.20	5.12	**978.32**
9	85.59	-	-	651.97	8.49	**660.46**
10	99.63	-	837.56	903.45	17.21	**920.66**
11	15.83	-	-	357.78	4.32	**362.10**
12	82.48	-	-	342.08	4.59	**346.67**
13	96.19	-	-	1616.11	7.43	**1623.54**
14	54.94	-	-	636.91	17.25	**654.16**
15	58.83	-	-	2062.13	42.72	**2104.85**

The solution times in Table 10 are obtained under certain parameter settings of the CPLEX. Our experience shows that using the bound strengthening and cuts has an adverse effect on the solution times as in the case of CMDPM. Although the aggressive scaling parameter is good for the CMDPM, it is not for the Lin-LMDPM-REL and hence only standard scaling is used. The algorithm used to solve the LP relaxations at the nodes is primal simplex for the Lin-LMDPM-REL and the dual simplex for the Lin-LMDPM-RES. The pricing, branching, and node-selecting strategies used are devex pricing, strong branching, and best bound node, respectively.

Computational results show that the Lin-LMDPM-REL and the Lin-LMDPM-RES obtained from the solution of the Lin-LMDPM-REL provides so good a lower bound and an upper bound, respectively, as in the case of the CMDPM that an optimal (near-optimal) solution to the Lin-LMDPM can be obtained without having to solve the Lin-LMDPM. In this regard, we can again mainly focus on improving the solvability of the relaxation Lin-LMDPM-REL.

4.3. Tardiness Minimization Deployment Planning Model (TMDPM)

In the LMDPM, it is implicitly assumed that it is possible to deploy all units to their destinations within their time windows, i.e., the problem is feasible. However, there may be cases where the problem is infeasible. Although there may be long-term decisions (such as increasing the processing capacity of an airfield) to get rid infeasibility, short-term and implementable decisions are of high value during a crisis. One possible short-term decision, relative to increasing the capacity of an airfield, is to increase the fleet size at the expense of increased cost and coordination requirements, e.g., to communicate with civilian companies for determining appropriate transportation assets and getting them to sign a lease. Another decision is to change time windows the units are assigned to. Then, there comes decisions on determining which ones and how much to change, which are not easy to make. Another option is to give priorities to units and allow units with low priorities to arrive at their destinations later than required, i.e., after their latest arrival times. In this case, how many low-priority units and how late they will be are arising questions. One viable decision during a crisis is to allow late deliveries but require that *lateness after latest arrival times*, which we call *tardiness*, be minimized. The purpose of this section is to introduce the model developed for this purpose.

We call the second model *Tardiness Minimization Deployment Planning Model* (TMDPM). The objective in the TMDPM is to minimize maximum tardiness and hence is of min-max type. The purpose is to minimize the maximum of the differences obtained by subtracting the units' latest allowable arrival times at their destinations from the arrival times of units determined in the model. The model aims to develop a deployment plan in which the arrival times of units that cannot be deployed within their time windows at their destinations are drawn to their latest allowable arrival times.

The development of the TMDPM is similar to that of the LMDPM. The decision variables, sets, and data used in the TMDPM are the same as those in the LMDPM.

4.3.1. Model (TMDPM)

Objective Function

$$\underset{TF,TE,CT,IV,IC,Y,X}{\text{Minimize}} \quad \underset{i\in N_D,c\in(CFIRST\cap CD_i),t\ge b_c}{\text{Max}} \left(t\times X_{ict}-b_c\right) \tag{33}$$

Constraints

In addition to Constraints (2) -(12), (15)-(24), and (28)-(29),

$$IC_{ict}-IC_{ic,t-1}-\sum_{l\in AB_i,v} CT_{lcv,t-trv_{lv}^{loaded}}=0 \qquad i\in N_D,c\in CD_i,c\in CFIRST,t\ge a_c \tag{34}$$

$$IC_{ict}\equiv demand_c \qquad i\in N_D,c\in CD_i,c\in CFIRST,t=|T| \tag{35}$$

The objective function (33) is the same as objective function (27) except that a_c in the inner term is replaced by b_c, i.e., it minimizes the maximum of the differences between arrival times of leader items at their destinations and their latest allowable arrival times.

Constraints (34) and (35) are the same as constraints (13) and (14), respectively, except that the set of time periods for which constraints (13) and (14) are defined are expressed for $t\ge a_c$ and $t=|T|$ instead of $t\in TD_c$ and $t=b_c$, respectively. These changes in constraints (13) and (14) are needed because units are allowed to arrive at their destinations at any time after their earliest allowable times, not only within their time windows.

The TMDPM is a nonlinear mixed integer program that can be linearized in the same way LMDPM is linearized.

The term $\underset{i \in N_D, c \in (CFIRST \cap CD_i), t \geq b_c}{\text{Max}} (t \times X_{ict} - b_c)$ in the objective function (33) is equal to the smallest number z that satisfies $z \geq \text{Max}(t \times X_{ict} - b_c)$ for $i \in N_D, c \in (CFIRST \cap CD_i), t \geq b_c$. For this reason, TMDPM is equivalent to the *Linearized TMDPM, Lin-TMDPM*, given below.

Linearized TMDPM (Lin-TMDPM)

Objective Function

$$\underset{TF, TE, CT, IV, IC, Y, X, z}{\text{Minimize}} \quad z \tag{36}$$

Constraints

In addition to constraints (2) -(12), (15)-(24), (28)-(29), and (34)-(35),

$$z \geq t \times X_{ict} - b_c \qquad i \in N_D, c \in (CFIRST \cap CD_i), t \geq b_c \tag{37}$$

$$z \geq \sum_{t \geq b_c} (t \times X_{ict}) - b_c \qquad i \in N_D, c \in (CFIRST \cap CD_i) \tag{38}$$

Note the resemblance between constraints (31)-(32) and constraints (37)-(38), respectively.

4.3.2. Computational Results

We test the performance of the Lin-TMDPM in the same way we have tested the Lin-LMDPM. The relaxation Lin-TMDPM-REL and the restriction Lin-TMDPM-RES of the TMDPM are obtained in the same way the relaxation Lin-LMDPM-REL and the restriction Lin-LMDPM-RES are obtained, respectively.

Table 11 gives the solution times in CPU seconds for limited fleet sizes for the Lin-TMDPM-REL and the Lin-TMDPM-RES. The solution times for the TMDPM are not presented because the optimal objective function values of the Lin-TMDPM-RES are equal to those of the Lin-TMDPM-REL and the optimal solutions of the Lin-TMDPM-RES are feasible for the Lin-TMDPM for all problems. That is, the optimal solutions of the Lin-TMDPM-RES are also optimal for the Lin-TMDPM.

The solution times in Table 11 range from 112.33 (Pr. 1) to 6747.23 (Pr. 5) CPU seconds. As in the cases of CMDPM and Lin-LMDPM, the solution times of the relaxation Lin-TMDPM-REL account for more than 98% of the total solution times of the Lin-TMDPM. The root solutions of the Lin-TMDPM-REL are obtained in relatively short times. However, a first integer solution is in later phases of the branch and bound algorithm. The gaps between the first integer solutions and the lower bounds range from 80% to 90% because the lower bounds are close to zero. However, the branch and bound generally jumps to the optimal solutions after finding a first integer solution, i.e., the first integer solution is generally optimal or close to optimal.

The solution times obtained for the Lin-TMDPM by using the solution methodology are not as good as the ones obtained for the Lin-LMDPM. However, they still seem promising for real-world applications when compared to the average time of one week needed to develop a deployment plan in the current practice. One reason for the increase in solution times is that the change of time windows from $[a_c, b_c]$ to $[a_c, T]$ increases the size of the feasible set of solutions which in turn increases the times spent in the algorithm.

Table 11. Solution times of Lin-TMDPM-REL and Lin-TMDPM-RES (CPU seconds). A "-" for a corresponding optimality criterion shows that the branch and bound jumps to a solution with a lower optimality criterion.

Pr.Id.	Lin-TMDPM-REL				Lin-TMDPM-RES	**Total Solution Time**
	Root	10%	5%	0%	0%	
1	1.98	-	-	110.24	2.09	**112.33**
2	9.55	-	-	498.98	5.43	**504.41**
3	16.92	-	-	189.45	3.46	**192.91**
4	40.31	-	-	312.17	6.36	**318.53**
5	85.39	-	-	6742.41	4.82	**6747.23**
6	17.00	-	-	393.59	3.19	**396.78**
7	35.62	-	-	662.87	5.62	**668.49**
8	220.28	-	-	4376.65	5.49	**4382.14**
9	142.30	-	-	3468.33	15.87	**3484.20**
10	73.01	-	-	3882.67	27.76	**3910.43**
11	18.60	-	-	380.47	4.39	**384.86**
12	43.43	-	-	632.35	7.15	**639.50**
13	80.89	-	-	1773.67	14.05	**1787.72**
14	71.11	-	-	893.33	29.53	**922.86**
15	65.19	-	-	1018.30	38.64	**1056.94**

CHAPTER 5

A HEURISTIC ALGORITHM

The solution methodology developed to solve the three models, CMDPM, Lin-LMDPM, and Lin-TMDPM, finds optimal solutions to intermediate-size problems in about one hour. The current practice of trial-and-error finds feasible, not optimal, solutions to problems of the same size in about one week (including data collection, communication, ratification, etc.). This shows that using analytical models improves both the solution quality and the solution time and hence embedding the models into a decision support system may be of great value to decision makers.

One of the main considerations in using a large-scale model such as ours in a decision support system is the solution time of the model. The expectation is, of course, to obtain high-quality, i.e., optimal or near-optimal, solutions in a reasonable amount of time. However, this is almost always not possible unless the model of concern has a special structure that facilitates its solution. In such cases, the most commonly used approach is to use a heuristic algorithm to solve the problem. The good of a model at this point is that it helps one to determine empirically how good the heuristic solutions are.

The solution times that we have obtained for the test problems by using the solution methodology lead one to think that the problems of larger sizes will be solved in a reasonable amount of time and that the models and the solution methodology can be embedded into a decision support system. However, a reasonable amount of time cannot be guaranteed for all situations. In the dissertation, the DPP is formulated, in essence, as a hybrid of multicommodity dynamic network flow and vehicle routing problems, which are both NP-hard. Hence, the solution methodology based on direct use of CPLEX is highly likely to be insufficient to solve problems beyond a certain size in a reasonable amount of time. This, combined with the fact that finding a feasible solution in a short time is of utmost value especially in crisis situations, makes it necessary to resort to heuristic solutions.

The purpose of this chapter is to propose a heuristic algorithm to solve the DPP and to compare heuristic objective functions and solution times for the test problems with those of the proposed solution methodology.

5.1. Heuristic Algorithm

The heuristic mainly involves solving a model in an iterative mode such that different sets of variables are fixed, suppressed, and/or set free at each iteration. The output of the heuristic is a solution feasible for the model of concern. In this regard, it is model-based and can be regarded as a means of finding a feasible solution for a model. Now, we define the heuristic for the CMDPM-REL.

The Heuristic Algorithm

Initialization. Suppose that we have a problem with $|G| \geq 2$, i.e., there is more than one deployment component to be moved, and that the problem is modeled using the CMDPM-REL. If there is a single deployment component, solve the CMDPM-REL directly without resorting to the heuristic. Let $G' = \emptyset$ where G' refers to the set of deployment-component indices for which an arrival time at destination has been determined.

Basic Iteration:

As long as $G \neq \emptyset$,

Select a member $\bar{g} \in G$

If $G' = \emptyset$,

Solve CMDPM-REL for only $\bar{g}$ and $c \in C_{\bar{g}}$.

Record the values of the variables IC_{ict} for $i \in N_D, c \in (CFIRST \cap C_{\bar{g}})$ and Y_{lct} for $i \in N_D, l \in AB_i, c \in (CFIRST \cap C_{\bar{g}})$ in the solution $x_{CMDPM-REL}$ as $\overline{IC_{ict}}$ and $\overline{Y_{lct}}$, respectively.

If $G' \neq \emptyset$,

Set IC_{ict} for $i \in N_D, g \in G', c \in (CFIRST \cap C_g)$ and Y_{lct} for $i \in N_D, l \in AB_i, c \in (CFIRST \cap C_g)$ in the CMDPM-REL to the corresponding values $\overline{IC_{ict}}$ and $\overline{Y_{lct}}$, respectively.

Solve the CDPM-REL for $G' \cup \bar{g}$ and $c \in C_{(G' \cup \bar{g})}$ (with fixed $\overline{IC_{ict}}$ and $\overline{Y_{lct}}$ for $c \in C_{(G' \cup \bar{g})}$)

Record the values of the variables IC_{ict} for $i \in N_D, c \in (CFIRST \cap C_{\bar{g}})$ and Y_{lct} for $i \in N_D, l \in AB_i, c \in (CFIRST \cap C_{\bar{g}})$ in the solution $x_{CMDPM-REL}$ as $\overline{IC_{ict}}$ and $\overline{Y_{lct}}$, respectively.

Set $G = G - \bar{g}$.

Set $G' = G' \cup \bar{g}$.

Termination:

If $G = \emptyset$, then the solution $x_{CMDPM-REL}$ obtained in the last iteration is a feasible solution to the CMDPM-REL.

The basic idea in the heuristic is to solve the CMDPM-REL incrementally rather than at one step. In the first iteration, the heuristic solves the model with only one deployment component, say $\bar{g}$, disregarding all other deployment components $g' \in G - \{\bar{g}\}$. That is, all decision variables related to $c \in g'$ where $g' \in G - \{\bar{g}\}$ are suppressed. In the second iteration, a deployment component $\hat{g} \in G - \{\bar{g}\}$ is selected and the model is solved for only deployment components $\bar{g}$ and $\hat{g}$ where the arrival time of deployment component $\bar{g}$ at its destination is fixed at the value obtained in the first iteration. In the third iteration, the model is solved for three deployment components, $\bar{g}$, $\hat{g}$, and a deployment component $\tilde{g} \in G - \{\bar{g}, \hat{g}\}$, but this time the arrival times of both $\bar{g}$ and $\hat{g}$ are fixed at the values obtained in the previous iteration. The procedure goes on in the same manner until all deployment components are processed.

It is clear that the solution obtained by using the heuristic is feasible for the model. However, there is a possibility that the heuristic does not find a solution, i.e., the heuristic concludes that the model is infeasible, when the original model is actually feasible. Such a situation is possible, for example, when the available number of transportation assets is too limited or when the time windows assigned to the units are too narrow. Because the arrival times of deployment components are fixed in the heuristic at each iteration, it is likely that the number of transportation assets used in the heuristic solution will be greater than that used by solving the model in a single pass to simultaneously make all deployment components ready within their time windows. However, because the objectives are the same, there should not be a significant difference in the number of transportation assets used in both solutions. Hence, an infeasibility encountered in the heuristic procedure is a strong indication that the original model itself may be infeasible. Such a quick infeasibility result may be of great value to transportation planners as well as to decision makers because they can reevaluate the fleet size and time restrictions of the units at the beginning of the operation. Even if the actual model is not feasible, infeasible termination in the heuristic strongly indicates that the actual model is tightly constrained and that the feasible set is quite small. In such cases, solving the model exactly using the proposed exact methods will likely generate optimal feasible solution in much less time than would be expected of the average feasible instance for which the feasible set is comparatively quite large.

The solution, hence the objective function, obtained by the heuristic may change depending on the order in which the deployment components are added to the model. In this regard, the heuristic may be repeated with different sequences until a solution with the lowest objective function is obtained. In such a case, the maximum number of repetitions is $|G|!$ if there are no any precedence relations between deployment components.

Instead of enumerating all possible situations to come up with the lowest-objective function, it is possible to develop several approaches to add the deployment components into the model. One approach may be to add deployment components in the order determined by how far their destinations are; starting from the deployment component whose destination is the farthest as well as starting from the one whose destination is the closest. Another approach may be to add the deployment components in the order determined by how long it takes to move a deployment component when it is the only one to be moved, i.e., disregarding all other deployment components. Because there are many factors affecting the solutions, it is not easy to offer an approach that will give good solutions for all possible cases. In the dissertation, we add deployment components by their initially assigned indices but show that changing the order may significantly change the objective function values. What is noteworthy at this point is that the order in which the deployment components are added may determine whether the heuristic finds a solution or not. Hence, changing the order in which the components are added may help obtain a solution when the heuristic cannot find a solution.

Although the proposed heuristic is given for the CMDPM-REL, it can easily be applied to solve other models as well. The heuristic algorithm to solve CMDPM is the same as the one given for CMDPM-REL except that CMDPM-REL in the algorithm is replaced by CMDPM. As to the heuristic algorithm for the Lin-LMDPM and Lin-TMDPM, and their relaxations, the same is valid except that X_{ict} is used, in addition to the variables IC_{ict} and Y_{lct}, to fix the arrival times of deployment components.

The set of variables to fix in the heuristic can be extended. The ones given in the algorithm are selected considering the solution time and the objective function value as well as the possibility of infeasibility. The idea is to set free as many decision variables as possible so that the objective function and the possibility of infeasibility get as low as possible. However, this increases the solution time of the heuristic. In this regard, the tradeoff between these three issues must be taken into account in determining the set of variables to fix. We think that the set of variables given to fix in the heuristic is essentially the minimum.

In addition to fixing arrival times of deployment components at their destinations, it is possible to fix the departure times of deployment components from their origins as well as from transfer points. It is even possible to fix the transportation assets and the route a deployment component follows, i.e., the variables CT_{lcvt} and TF_{lvt}. However, it is highly likely that this will increase the possibility of infeasible termination.

5.2. How to use the heuristic

The heuristic can be used stand-alone or as a part of the solution methodology. When the heuristic is used stand-alone, it is directly used to obtain a feasible solution for CMDPM, Lin-LMDPM, or Lin-TMDPM. The solution obtained by the heuristic is also feasible for the relaxation of the corresponding model. The reservation about this usage of the heuristic is that the expected gains with respect to solution time may not be realized because this usage does not take advantage of improvements in the solution time provided by the proposed solution methodology.

When the heuristic is used as part of the solution methodology, it is used to find a feasible solution for the relaxation of the model of concern. The feasible solution of the relaxation may be used in two ways. In the first one, it is used as an initial starting solution for the relaxation. That is, the "Solve the relaxation" step of the solution methodology is implemented with a feasible solution for the relaxation. It is expected that this will decrease the solution time of the relaxation and hence the total solution time because the relaxation solution time constitutes about 98% of the total solution of the proposed solution methodology. In the second one, the feasible solution of the relaxation is regarded as the final solution of the relaxation and hence the "Obtain the restriction" step of the solution methodology is implemented with the feasible solution obtained by the heuristic. In this usage, the solution methodology may be stopped after obtaining a solution for the restriction as the solution for the restriction will be feasible for the original model. This may be preferred when time is too limited to obtain a deployment plan. Note that the heuristic may also be used in the "Solve the restriction" and "Solve the original model" steps of the solution methodology as well. However, it may not be worth to do that because time spent in these two steps constitutes a small portion of the total solution time of the solution methodology.

In our computational tests regarding the heuristic, we use the heuristic to solve the relaxations of the models to be able to compare results with those presented in the previous chapters.

5.3. Computational Results

One of the main questions in designing and using a heuristic is to determine how far the heuristic solutions are from optimal solutions. In this section, we explore the answer to this question empirically by comparing the objective function values of the test problems obtained by the heuristic for the relaxations with those obtained by solving the corresponding relaxations directly. For comparison, we use the simple statistic $\frac{100\% \times z^*_{\text{Heuristic, REL}}}{z^*_{\text{REL}}}$ where $z^*_{\text{Heuristic, REL}}$ refers to optimal objective function value obtained by the heuristic while z^*_{REL} refers to optimal objective value obtained by solving the relaxation directly. Now, we give the comparison results for the CMDPM-REL, Lin-LMDPM-REL, and Lin-TMDPM-REL, respectively.

Table 12 gives the objective function and the comparison statistic values for the CMDPM-REL. The results presented in the table are obtained by adding deployment components into the algorithm in the order of 3, 4, 2, and 1, respectively. Of these components, deployment components 3 and 4 have the same destination where the deployment component 3 has a higher priority than the deployment component 4. The results in Table 12 show that the heuristic terminates feasible for 14 problems out of 15; the heuristic terminates infeasible for only Problem 5 (which is in fact feasible). The objective function values $z^*_{\text{Heuristic, REL}}$ obtained by the heuristic and objective function values z^*_{REL} obtained by solving the relaxation directly are the same (or the gap is less than 0.01%) for 10 problems (Problems 6 through 15) out of 14. The largest gap between the objective function values is for Problem 4 with a value of 75.65%. The average and standard deviation of the comparison statistic values are 13.96% and 25.42%, respectively.

Table 12. Comparison of the heuristic and optimal objective function values of the CMDPM-REL.

Pr.Id.	$z^*_{\text{Heuristic, REL}}$	z^*_{REL}	$\frac{100\% \times z^*_{\text{Heuristic, REL}}}{z^*_{\text{REL}}}$
1	1817183	1483865	**122.46%**
2	2226944	1595856	**139.55%**
3	4437892	2813069	**157.76%**
4	12073405	6873491	**175.65%**
5	infeasible	13877031	**-**
6	1017431	1017431	**100.00%**
7	1596084	1596062	**100.00%**
8	2955359	2955340	**100.00%**
9	6916499	6916499	**100.00%**
10	13879744	13879744	**100.00%**
11	1017467	1017467	**100.00%**
12	1596305	1596305	**100.00%**
13	2955628	2955594	**100.00%**
14	6875618	6875481	**100.00%**
15	13881081	13880981	**100.00%**

One reason behind the large gaps for Problems 1 through 4 is that the time windows assigned to the units are much tighter for these problems than those of Problems 6 through 15 (they are cut by one half) although the amount of items to be moved are the same. One other reason is related to the type of the objective function. Because the objective is to minimize the cost, the model seeks to use the range of time windows as much as possible to reduce costs. In this regard, the arrival times of the first two components are highly likely to be at the lower and upper bounds of their time windows. This requires that more transportation assets be drawn from the pool so that the remaining deployment components arrive at their destinations within their time windows.

When the deployment components are added by their index numbers, i.e., 1, 2, 3, and 4, respectively, 5 out of 15 problems turn out to be infeasible. The reason for this is as follows. Because the deployment component 4 has a lower priority than the deployment component 3, it is to arrive at its destination later than deployment component 3. However, because of the comments regarding the type of objective function given in the previous paragraph, the model sets the arrival time of the deployment component 3 to the upper bound of the time window which is the latest arrival time of the deployment component 4. Hence, when the deployment component 4 is added, the model becomes infeasible. It is possible to get rid of this type of infeasibility by requiring a deployment component with a higher priority not to arrive at destination at the upper bound of its time window or by changing the time window slightly, e.g., one unit of time period, while adding the component with a lower priority.

Table 13 gives the objective function and the comparison statistic values for the Lin-LMDPM-REL. The results presented in the table are obtained by adding deployment components into the algorithm in the order of 1, 2, 3, and 4, respectively. The results in Table 13 show that feasible solutions are obtained for all problems. The objective function values $z^*_{\text{Heuristic, REL}}$ obtained by the heuristic and objective function values z^*_{REL} obtained by solving the relaxation directly are the same for 7 problems (Problems 1, 3, 4, 5, 8, 9, and 10) out of 15. The largest gap between the objective function values is for Problems 13 and 15 with a value of 38.46%. The average and standard deviation of the comparison statistic values are 11.85% and 15.37%, respectively.

Table 13. Comparison of the heuristic and optimal objective function values of the Lin-LMDPM-REL.

Pr.Id.	$z^*_{\text{Heuristic, REL}}$	z^*_{REL}	$\frac{100\% \times z^*_{\text{Heuristic, REL}}}{z^*_{\text{REL}}}$
1	13	13	**100.00%**
2	14	13	**107.69%**
3	13	13	**100.00%**
4	13	13	**100.00%**
5	16	16	**100.00%**
6	23	22	**104.55%**
7	23	22	**104.55%**
8	24	24	**100.00%**
9	24	24	**100.00%**
10	24	24	**100.00%**
11	32	25	**128.00%**
12	32	25	**128.00%**
13	36	26	**138.46%**
14	32	25	**128.00%**
15	36	26	**138.46%**

As a good example of how the order in which the deployment components are added changes the objective function values, we solve the Problems 11 through 15 by adding the deployment components in the reverse order, i.e., 4, 3, 2, and 1. The results are presented in Table 14. The objective function values turn out to be the same for all five problems. When these results are combined with those in Table 13, the average and standard deviation of the comparison statistic values are 1.12% and 2.42%, respectively.

Table 14. An example of how the objective function values are dependent on the order in which deployment components are added into the model.

Pr.Id.	$z^*_{\text{Heuristic, REL}}$	z^*_{REL}	$\frac{100\% \times z^*_{\text{Heuristic, REL}}}{z^*_{\text{REL}}}$
11	25	25	**100.00%**
12	25	25	**100.00%**
13	26	26	**100.00%**
14	25	25	**100.00%**
15	26	26	**100.00%**

Table 15 gives the objective function and the comparison statistic values for the Lin-TMDPM-REL. The results presented in the table are obtained by adding deployment components in the order of 1, 2, 3, and 4, respectively. The results in Table 15 show that feasible solutions are obtained for all problems. The objective function values $z^*_{\text{Heuristic, REL}}$ obtained by the heuristic and objective function values z^*_{REL} obtained by solving the relaxation directly are the same for 7 problems (Problems 1, 3, 4, 5, 8, 9, and 10) out of 15. The largest gap between the objective function values is for Problems 13 and 15 with a value of 166.67%. The average and standard deviation of the comparison statistic values are 59.11% and 69.72%, respectively.

As another good example of how the order in which the deployment components are added changes the objective function values, we solve the Problems 11 through 15 by adding the deployment components in the reverse order, i.e., 4, 3, 2, and 1. The results are presented in Table 16. The objective function values turn out to be the same for all five problems. When these results are combined with those in Table 15, the average and standard deviation of the comparison statistic values are 8.89% and 18.76%, respectively.

Table 15. Comparison of the heuristic and optimal objective function values of the Lin-TMDPM-REL.

Pr.Id.	$z^*_{\text{Heuristic, REL}}$	z^*_{REL}	$\frac{100\% \times z^*_{\text{Heuristic, REL}}}{z^*_{\text{REL}}}$
1	3	3	**100.00%**
2	4	3	**133.33%**
3	3	3	**100.00%**
4	3	3	**100.00%**
5	6	6	**100.00%**
6	3	2	**150.00%**
7	3	2	**150.00%**
8	4	4	**100.00%**
9	4	4	**100.00%**
10	4	4	**100.00%**
11	12	5	**240.00%**
12	12	5	**240.00%**
13	16	6	**266.67%**
14	12	5	**240.00%**
15	16	6	**266.67%**

Table 16. An example of how the objective function values are dependent on the order in which deployment components are added into the model.

Pr.Id.	$z^*_{\text{Heuristic, REL}}$	z^*_{REL}	$\frac{100\% \times z^*_{\text{Heuristic, REL}}}{z^*_{\text{REL}}}$
11	5	5	**100.00%**
12	5	5	**100.00%**
13	6	6	**100.00%**
14	5	5	**100.00%**
15	6	6	**100.00%**

One other question to answer in designing and using a heuristic is how much the solution time is improved by using the heuristic. To answer this question, we compare the solution times obtained by applying the heuristic to solve the relaxations and those obtained by solving the relaxations directly.

Table 17 through Table 19 give the solution times of the test problems for CMDPM-REL, Lin-LMDPM-REL, and Lin-TMDPM-REL, respectively, obtained in both ways together with the comparison statistic $\frac{T_{\text{REL}}}{T_{\text{Heuristic, REL}}}$, where $T_{\text{Heuristic, REL}}$ refers to solution time obtained by the heuristic while T_{REL} refers to solution time obtained by solving the relaxation directly.

To compare the solution times, we focus only on the problems solved to optimality by the heuristic algorithm. The heuristic finds optimal solutions for the CMDPM-REL for only Problems 6 through 15 and Table 17 shows that the solution times of the heuristic are better with a mean of 2.78 times and with a standard deviation of 1.29 times for those problems. The heuristic finds optimal solutions for Problems 1, 3-5, and 8-15 (including the ones obtained by changing the order of adding deployment components) for the Lin-LMDPM-REL and Lin-TMDPM-REL. Table 18 shows that the solution times of the heuristic are better with a mean of 9.04 times and with a standard deviation of 9.22 times for the Lin-LMDPM-REL for the stated problems. Similarly, Table 19 shows that the solution times of the heuristic are better with a mean of 17.95 times and with a standard deviation of 11.95 times. Combining the results regarding solution times with the ones regarding the optimal objective function values shows that the heuristic can find good feasible solutions in relatively short times and possibly optimal or near-optimal solutions when the problem on hand is analyzed before using the heuristic to determine the order of deployment components to add.

Table 17. Comparison of the solution times of the heuristic and optimal solutions of the CMDPM-REL.

Pr.Id.	T_{REL}	$T_{\text{Heuristic, REL}}$	$\frac{T_{\text{REL}}}{T_{\text{Heuristic, REL}}}$
1	1404.79	46.74	**30.05**
2	1270.85	190.33	**6.67**
3	499.82	186.65	**2.67**
4	572.47	178.54	**3.21**
5	965.26	102.08	**9.46**
6	182.10	98.67	**1.85**
7	519.33	466.98	**1.11**
8	935.56	342.23	**2.73**
9	881.91	291.43	**3.03**
10	1203.05	663.87	**1.81**
11	2983.39	513.55	**5.81**
12	3755.53	991.05	**3.79**
13	3959.07	1415.46	**2.80**
14	2310.08	934.56	**2.47**
15	2375.29	1009.68	**2.35**

Table 18. Comparison of the solution times of the heuristic and optimal solution of the Lin-LMDPM-REL.

Pr.Id.	T_{REL}	$T_{\text{Heuristic, REL}}$	$\frac{T_{\text{REL}}}{T_{\text{Heuristic, REL}}}$
1	36.58	9.87	**3.71**
2	37.16	5.05	**7.36**
3	55.75	5.40	**10.32**
4	319.13	8.89	**35.89**
5	491.11	53.95	**9.10**
6	306.00	45.55	**6.72**
7	178.00	15.33	**11.61**
8	973.20	159.06	**6.12**
9	651.97	93.67	**6.96**
10	903.45	154.05	**5.86**
11	357.78	74.08	**4.83**
12	342.08	57.03	**5.99**
13	1616.11	1318.53	**1.23**
14	636.91	41.62	**15.30**
15	2062.13	656.54	**3.14**

Table 19. Comparison of the solution times of the heuristic and optimal solution of the Lin-TMDPM-REL

Pr.Id.	T_{REL}	$T_{\text{Heuristic, REL}}$	$\frac{T_{\text{REL}}}{T_{\text{Heuristic, REL}}}$
1	110.24	45.95	**2.39**
2	498.98	8.16	**61.15**
3	189.45	7.92	**23.92**
4	312.17	15.05	**20.74**
5	6742.41	208.84	**32.28**
6	393.59	106.72	**3.69**
7	662.87	25.66	**25.83**
8	4376.65	355.41	**12.31**
9	3468.33	113.21	**30.64**
10	3882.67	147.47	**26.33**
11	380.47	99.60	**3.82**
12	632.35	18.83	**33.58**
13	1773.67	326.5	**5.43**
14	893.33	43.60	**20.49**
15	1018.30	295.32	**3.44**

CHAPTER 6

FROM THEORY TO PRACTICE

In this chapter, we explain how the proposed models can be used in the real world. In this respect, we discuss how the output of a model can be used to derive a deployment plan, what other variations of the models might there be, what type of what-if questions can be answered, and how the models can be used in the bottom-up and top-down deployment planning approaches.

6.1. Deriving a deployment plan from the output of the model

The output of the model can be used to determine deployment plans of units as well as the allocation schedules of transportation assets and the transportation infrastructure as follows.

Recall that the first indexed item in each deployment component is designated as the leader item for that component with the understanding that all other items in that component follow the same route and the schedule as does the leader. In this regard, it is sufficient to keep track of only the leader item in each deployment component to obtain the route and movement schedule of that component. A deployment plan for a unit is obtained by combining the movement plans of the deployment components of the unit.

The values of the decision variables CT_{lcvt} are used as the main input to get the movement plan of a deployment component. We first explain how the route and movement schedule for a deployment component g is obtained. Suppose that c_g^{leader} is the leader item of g. For now, assume that there is one unit of c_g^{leader}, i.e., $demand_{c_g^{leader}} = 1$, and that only one unit of transportation asset v' can carry the leader item. Because the decision variable CT_{lcvt} gives the units (amount) of item c that starts moving on arc l at time t via transportation asset v, the route and movement schedule of g is obtained from the arcs l and time periods t for which $CT_{l,c_g^{leader},v',t} = 1$. As an example, suppose that $CT_{4,c_g^{leader},v',7} = 1$, $CT_{9,c_g^{leader},v',12} = 1$, $CT_{6,c_g^{leader},v',15} = 1$, and $CT_{13,c_g^{leader},v',22} = 1$. (Note that $TF_{l,v',t} = 1$ for $l = 4, 9, 6, 13$ and $t = 7, 12, 15, 22$.) Thus, the deployment component g follows the route consisting of the arcs numbered 4, 9, 6, and 13 in the given order. The movement schedule of g on the arcs 4, 9, 6, and 13 is at time periods 7, 12, 15, and 22, respectively. Because g uses the same vehicle on the given route, i.e., single mode, 12-7=5, 15-12=3, and 22-15=7 correspond to travel times of transportation asset v' on arcs 4, 9, and 6, respectively. Supposing that the travel time of v' on arc 13 is 6, g arrives at its destination at time period 28 after 21 time periods it starts moving from its origin at time period 7.

Now, suppose that $demand_{c_g^{leader}}=15$ and that the number of units of transportation asset v' is 10 and hence not sufficient to carry all of c_g^{leader}. In this regard, transportation asset v'' is used to carry 15-10=5 units of c_g^{leader}. In this case, for the same route and movement schedule, the values of the decision variables CT_{lcvt} are $CT_{4,c_g^{leader},v',7}=10$, $CT_{4,c_g^{leader},v'',7}=5$, $CT_{9,c_g^{leader},v',12}=10$, $CT_{9,c_g^{leader},v'',12}=5$, $CT_{6,c_g^{leader},v',15}=10$, $CT_{6,c_g^{leader},v'',15}=5$, $CT_{13,c_g^{leader},v',22}=10$, and $CT_{13,c_g^{leader},v'',22}=5$ because all items are required to move as a whole. (Note that $TF_{l,v',t}=10$ and $TF_{l,v'',t}=5$ for $l=4,9,6,13$ and $t=7,12,15,22$ assuming that both one unit of v' and v'' can carry only one unit of c_g^{leader}.) So, if more than one type of transportation asset is used to carry a leader item on an arc, it is enough just to follow the values of variables CT_{lcvt} for a fixed transportation asset, either v' or v''. Actually, to get the route and movement schedule of a component what is important is at what arcs and time periods the variables CT_{lcvt} are nonzero, rather than their quantities.

If the deployment component g uses a combination of different transportation modes, the route and movement schedule is found for each transportation mode in a similar manner, i.e., follow at what arcs and time periods the decision variables CT_{lcvt} become non-zero for a fixed transportation asset on each mode.

As a convention, loading and unloading times of transportation assets are incorporated into travel times of transportation assets on the arcs. Hence, the time at which a deployment component starts moving from its origin or a transfer node is actually the time at which loading operation for that deployment component starts. The time at which a deployment component arrives at its destination or a transfer point is actually the time at which the unloading operation terminates, i.e., the time at which the unloading operation for the deployment component starts is found by subtracting from the arrival time of the component the unloading time of the transportation asset in which the component has arrived. In the first example above, assuming that loading and unloading times of v' are 1, the loading operation at the home base of g, i_g^{origin}, starts and ends at time periods 7 and 8, respectively. Similarly, the unloading operation at the destination of g, i_g^{dest}, starts and ends at time periods 27 and 28, respectively.

A deployment component may be assumed to be waiting idle at all time periods at which it does not move or undergo any loading or unloading operations. In this respect, the deployment component g waits idle at its origin i_g^{origin} for time periods 1 through 7 and at its destination i_g^{dest} for time periods 28 through T, the planning horizon. In the model, this is tracked by the values of the decision variables IC_{ict}. For our example, the values of the decision variables in the output are $IC_{i_g^{origin},c_g^{leader},t}=1$ for $t=1,...,7$ and $IC_{i_g^{dest},c_g^{leader},t}=1$ for $t=28,...,T$ when $demand_{c_g^{leader}}=1$.

The values of the decision variables CT_{lcvt} and IC_{ict} give the time-based allocation of transportation infrastructure to a deployment component in addition to the route and movement schedule of a deployment component. For example, $CT_{4,c_g^{leader},v',7}=1$ means that arc 4 is allocated to the deployment component at time period 7 for the duration of the travel time of v' on arc 4.

Whether another deployment component is allocated to the same arc at the same time is determined depending on the capacity of the arc. Similarly, $IC_{i_g^{origin}, c_g^{leader}, t} = 1$ for $t = 1,...,7$ means that i_g^{origin} is allocated to the deployment component for $t = 1,...,7$.

Now, we explain how to determine the allocation of transportation assets to deployment components.

The allocation of transportation assets to deployment components can be determined by using the values of the variables CT_{lcvt} and TF_{lvt}. The variables CT_{lcvt} give the number of units of items of index c carried via transportation assets of index v on an arc l at a time period t and the variables TF_{lvt} give the number of units of loaded transportation assets of index v on an arc l at a time t. Thus, the number of units of transportation assets v that are used to carry items of index c on an arc l at a time epoch t can be determined by using the information provided by the values of CT_{lcvt} and TF_{lvt}. In this regard, determining the transportation assets used to carry all items of a deployment component gives the allocation of transportation assets to that deployment component.

Notice that CT_{lcvt} give also a coupling of items of index c with transportation assets of index v on an arc l at a time period t. If transportation assets of index v are coupled with only items of index c, then the value of the decision variable TF_{lvt} on the arc l at time t gives the number of units of transportation assets of index v assigned to carry item c. Returning to our example, for $demand_{c_g^{leader}} = 1$, $CT_{4, c_g^{leader}, v', 7} = 1$ and $TF_{4, v', 7} = 1$

mean that one unit of transportation asset v' is assigned to deployment component g to carry the leader item c_g^{leader} on arc 4 at time period 7 for the duration of the travel time of v' on arc 4. Similarly, for $demand_{c_g^{leader}} = 15$, $CT_{4,c_g^{leader},v',7} = 10$, $CT_{4,c_g^{leader},v'',7} = 5$, $TF_{4,v',7} = 10$ and $TF_{4,v'',7} = 5$ mean that 10 units of transportation assets of index v' and 5 units of transportation assets of index v'' are assigned to carry the leader item on arc 4 at time period 7. Assuming that the deployment component g follows the route and movement schedule given above, v' and v'' are assigned to g from time period 7 to time period 28.

If transportation assets of index v are coupled with items of different indices, then the items can be assigned to transportation assets by taking the capacities of the transportation assets into account. As an example, suppose that the deployment component g has item $c_g^{non-leader}$ in addition to the c_g^{leader} and that one unit of v' can carry 2 units of $c_g^{non-leader}$. (Recall that one unit of v' can carry only one unit of c_g^{leader}.) Assuming that $demand_{c_g^{leader}} = 15$ and $demand_{c_g^{non-leader}} = 10$, 20 units of v' are needed to carry c_g^{leader} and $c_g^{non-leader}$. In compliance with this, the values of the variables on arc 4 at time period 7 are $CT_{4,c_g^{leader},v',7} = 15$, $CT_{4,c_g^{non-leader},v',7} = 10$, and $TF_{4,v',7} = 20$. Of these 20 units of v', 15 are allocated to carry c_g^{leader} and 5 are allocated to carry $c_g^{non-leader}$. If one unit of transportation asset $\tilde{v}$ is sufficient to carry both c_g^{leader} and $c_g^{non-leader}$, then the values of the variables are $CT_{4,c_g^{leader},v',7} = 15$, $CT_{4,c_g^{non-leader},v',7} = 10$, and $TF_{4,\tilde{v},7} = 1$.

A complication may arise when more than one deployment component starts moving on an arc at the same time. In this case, if the transportation asset is big enough to move items of all components, e.g., a ship, and if the routes of all components are the same, then there is not any problem as the transportation asset is assigned to all deployment components at the same time. In the context of the example, we may assume that $TF_{l,\tilde{v},t}=1$ for $l=4,9,6,13$ and $t=7,12,15,22$, respectively. If the routes of deployment components differ, then some caution is needed as this is also related to the rounding of the variables TF_{lvt}. Now, suppose that the deployment component g follows the route $l=4,9,6,13$ and that the deployment component g' follows the route $l=4,10,11,12$ (where arcs 12 and 13 are incoming to different nodes). Thus, although both components start moving on arc 4 at time period 7, their routes differ at the end of arc 4. Assuming that both components are exactly the same, the values of the variables in the output are $TF_{l,\tilde{v},t}=0.5$ for $l=9,6,13$ and $l=10,11,12$. Then, $TF_{l,\tilde{v},t}=0.5$ is rounded up, i.e., two units of $\tilde{v}$ start moving on arc 4 at time 7, as explained in Chapter 3. In this respect, it is suggested to start from the destination of a deployment component in obtaining the movement plan of that component.

If several numbers of units of transportation asset $\tilde{v}$ are used by different components, then the allocation of transportation assets is made by taking into account the size and the number of units of items as well as the capacity of a single unit of transportation asset. The rounding operation is repeated as needed.

The allocation of transportation assets to deployment components also gives the routes and movement schedules of loaded transportation assets. The routes and movement schedules of empty transportation assets are obtained similarly by using the values of the decision variables TE_{lvt}. That the values of the inventory variables IV_{ivt} are greater than zero at a node shows that the transportation assets are waiting idle at that node (e.g., waiting for another deployment component to arrive).

The output of the models gives also a sourcing strategy for transportation assets. Because each transportation asset index refers to transportation assets of a type from a source type at a location, a sourcing strategy for transportation assets can be obtained by using the values of the decision variables TE_{lvt} corresponding to $l \in A_{dum}$.

6.2. How to use the models in creating deployment plans

Deployment plans are created by using bottom-up or top-down approaches. The current practice in the Turkish Armed Forces is the bottom-up approach. In this approach, units, starting from the lowest-level with planning responsibility, develop deployment plans and send their plans to a higher level unit until the highest-level command responsible for the operation develops its deployment plan. For instance, companies prepare their deployment plans and send them to a higher-level unit, battalion, in accordance with the military force structure. Battalions receiving plans from their companies bring their companies' deployment plans together, prepare their deployment plans and send them to a higher-level command, brigade. The problem with this approach is that deployment plans of units may conflict in the sense that they may demand the usage of the same resources, transportation infrastructure and transportation assets, at the same time. For example, deployment plans of companies of a battalion or different battalions may conflict because each company develops its deployment plan independent of other companies. For this reason, what is expected at each level while plans go up in the hierarchy is that conflicting deployment plans are resolved (de-conflicted) and lower-level units are asked to change their plans appropriately. For instance, a battalion resolves conflicting plans of their companies and a brigade conflicting plans of its battalions. However, it is very difficult or even impossible depending on the number of units participating in an operation to resolve conflicting plans manually. Even though de-conflicting is possible, the approach does not create cost-effective plans.

The latter top-down approach aims to prevent conflicts beforehand. In this approach, the process starts from the highest level and ends when the lowest-level commands prepare their deployment plans. To prevent conflicts in advance, starting from the highest level, units provide guidance to their sub-units and sub-units develop their plans based on the guidance provided to them. For example, a brigade (battalion) may determine time intervals at which each of its battalions (companies) may use a transportation mode(s) and a transfer point(s). Then, each battalion (company) prepares its deployment plan based on the given information. After the lowest-level commands prepare their plans, they send them through the hierarchy up to the highest-level command. At each level, plans of sub-units are united and revised. Such an approach, in addition to preventing conflicts from the beginning, will help develop cost-effective plans because the process considers the entire system.

Despite the disadvantages of the bottom-up planning approach, it is preferred by high-level commands to impose responsibility on the low-level commands. It is also preferred by low-level commands because they would like to have some initiative in the operation they are going to carry out. One other reason for the preference of the bottom-up approach is that there does not exist tools to help develop deployment plans using top-down approach.

The models developed in this dissertation can be used in both planning approaches. The idea in the heuristic algorithm can directly be applied to develop a deployment plan by using the bottom-up planning approach. For example, each company of a battalion develops its deployment plan independent of other companies of the battalion by using a model, say, CMDPM. The battalion receiving the plans of its companies may use the models in the first place to see whether the plans of the companies are implementable, i.e., to see whether there are any conflicts. This can easily be done by fixing some information, e.g., the allocation of transportation assets to units. If there are any conflicts or if there are not any conflicts but there is a need for revision of the plans, the battalion may use the heuristic algorithm to develop a deployment plan for the battalion. The order in which the companies are added can easily be determined by the battalion commander depending on the criticality of the mission a company is expected to execute. The battalion may either create a plan from scratch or fix anything desired. In any case, the battalion informs its companies of any changes made in the plans of the companies. After developing its deployment plan, the brigade sends its plan to the brigade. The brigade receiving plans from its battalions goes through the same process. The process goes on similarly until the deployment plan for the highest level command is developed.

When the top-down planning approach is used, one should not expect to get deployment plans of all deployable units by running a model only once at the highest level except when the problem of concern is of moderate size. In this approach, there is a need for a set of models that can create deployment plans with different levels of detail depending on the planning level. For example, a model capable of creating coarse deployment plans that show when and using what transportation modes the units are moved may be sufficient for the highest level command. On the other hand, a model capable of developing deployment plans that show what transportation assets carry what items of each unit is of value at lower levels command. However, this is not sufficient for the lowest level command as the loading plan of each transportation asset must be known at this level. Hence, a model that keeps track of individual transportation assets and items are needed at the lowest level command.

In the top-down approach, a model appropriate for the highest level command is run and necessary information that will guide the lower-level units to develop their deployment plans is revealed. For example, the day at which a unit is to be at destination or the transportation mode a unit is to use is revealed to the units. Units receiving the guidance from the highest level develop deployment plans appropriate for their planning levels using appropriate models and reveal necessary information to their lower-level units. This process goes on until the lowest level commands develop their detailed deployment plans. The plans are then sent to the highest level through the military hierarchy where at each level the plans are united and revised.

The modeling structure presented in the dissertation allows the models to be run with data of different levels of detail. For example, for a high-level planning, items for each unit may be aggregated as pax and cargo or as pax and bulk, oversize, and outsize cargo as in Baker et al. (1999, 2002). In this case, a line on the item list represents the number of pax or the amount of bulk, oversize, and outsize cargo. Data structure regarding transportation assets is modified accordingly, e.g., the capacities are expressed to represent carrying capacities for bulk, oversize, and outsize cargo. Similarly, the node and arc representations in the underlying network may be also aggregated. On the other hand, the model can be modified to track individual transportation assets and items for a low-level planning. In this case, each item index and transportation asset index refers to an individual item and a transportation asset. The decision variables regarding transportation assets and items are defined as binary variables, e.g., CT_{lcvt} is defined to be 1, if item c starts moving on arc l via transportation asset v at time t; 0, otherwise. Constraints (3) , (14), (16), and (20)-(22) are modified accordingly. The right-hand sides of constraints (3) , (14), and (16) are set to one. Constraints (20) are removed and constraints (21) and (22) are modified, respectively, as follows:

$$\sum_{v} CT_{lc'vt} = \sum_{v} CT_{lcvt} \qquad i \in (N - n_d), l \in AF_i, g \in G, c \in CFIRST, c \in C_g, c' \in C_g, c' \neq c \tag{39}$$

$$\sum_{v} CT_{lc'v,t-trv_{lv}} \leq \sum_{l' \in AB_i, v, e_c \leq t' < t} CT_{l'cv,t'-trv_{l'v}} \qquad i \in N_D, l \in AB_i, c \in CFIRST, c' \in CFIRST, c' \in CP_{ic}, t \tag{40}$$

6.3. Some other model variations

The CMDPM, LMDPM, and TMDPM are actually three variations of the deployment planning problem. In this section, we mention some other variations.

When the resources are not sufficient to deploy all units within their time windows, the TMDPM can be used to create deployment plans in which the arrival times of units not deployed within their time windows are drawn to their latest arrival times. In the TMDPM, all units are of the same importance. For this reason, the model does not differentiate between units and determines freely which units to deploy within their time windows and which units not to deploy. Thus, the solution of the TMDPM may have some high-priority units not deployed within their time windows. (Note that this is different from precedence relations between deployment components/units.) To prevent this situation from occurring, one approach might be to assign priorities to the units and require that the sum of the priorities of the units deployed within their time windows is maximized. Note that this approach allows the non-delivery of some units. To formulate this problem, we define a new decision variable β_c that shows the amount of item c not delivered within its time window. Note that it is sufficient to define this variable only for the leader items that $\beta_c = 0$ or $\beta_c = demand_c$ due to the unsplittable flow requirement. Assuming that the priority of an item, $prior_c$, is obtained properly from the priorities of the units, the *Priority Maximization Deployment Planning Model* (PMDPM) is given below.

Priority Maximization Deployment Planning Model (PMDPM)

Objective Function

$$\underset{TF,TE,CT,IV,IC,Y,\beta}{\text{Maximize}} \sum_{i\in N_D,\, l\in AB_i,\, c\in(CFIRST\cap CD_i),v,t} prior_c \times \left(\frac{CT_{lcvt}}{demand_c} \right) \quad (41)$$

Constraints

In addition to constraints (2) -(12) and (14)-(24),

$$IC_{ict} - IC_{ic,t-1} - \sum_{l\in AB_i, v} CT_{lcv,t-trv_{lv}^{loaded}} - \beta_c = 0 \qquad i\in N_D, c\in CD_i, c\in CFIRST, t\in TD_c \quad (42)$$

The objective function (41) maximizes the sum of the priorities of the units deployed within their time windows while the constraints (42) allow the non-delivery of items at destination nodes.

When the resources are not sufficient, one other approach might be to assign penalties to the units for not delivering them. In this case, the objective is to minimize the sum of the penalties of units not deployed within their time windows. This problem is closely related to the PMDPM and can easily be formulated by changing only the objective function of the PMDPM. Assuming that the penalty for not delivering an item c, $penalty_c$, is obtained properly, the *Penalty Minimization Deployment Planning Model* (PnMDPM) is given below.

Penalty Minimization Deployment Planning Model (PnMDPM)

Objective Function

$$\underset{TF,TE,CT,IV,IC,Y,\beta}{\text{Maximize}} \sum_{i\in N_D, l\in AB_i, c\in(CFIRST\cap CD_i),v,t} penalty_c \times \left(\frac{\beta_c}{demand_c}\right) \tag{43}$$

Constraints

Constraints (2) -(12), (14)-(24), and (42).

Other model variations may be obtained by adding several constraints to the current models. For example, a budget constraint such as $\sum_{l\in A_{dum}, v, t} TE_{lvt} \leq Budget$ may be added to the CMDPM. Constraints that set minimum and maximum limits on the number of transportation assets for each type, location, and/or source used to deploy units may be added, e.g., $\sum_{l\in A_{dum}, v, t} TE_{lvt} \geq Min_v, \forall v$ or $\sum_{l\in A_{dum}, v, t} TE_{lvt} \leq Max_v, \forall v$.

6.4. Sample questions that can be answered by using the models

The first usage of the models is to evaluate and assess investment decisions regarding transportation assets and transportation infrastructure. In developing a good investment plan, the trade-off between the deployment cost and deployment time must be observed, not only the cost. For this reason, the CMDPM must be used in conjunction with the LMDPM and the TMDPM. A figure that shows lateness and tardiness values for different cost values may be an invaluable support to the decision makers. Such figures will help the decision makers to develop investment plans robust enough to support many different scenarios.

The second usage of the models is to develop a deployment plan. Depending on the situation, one of the three models may be used. For example, when updating operations plans for potential contingencies on a routine basis or when deploying units for exercises, international operations, e.g., peace support operations, or other purposes during peacetime, the CMDPM may be used. On the other hand, during a crisis, the LMDPM and TMDPM may be of more value as they create plans directed towards the execution of an operation. In both cases, the analysts or the decision makers need to answer several questions before making the final decision. We now give examples of questions that can be answered or that should be asked.

The first question is of course whether it is possible to carry out a deployment successfully with the given number, type, and locations of transportation assets and the given capacities of transportation infrastructure. If not possible, i.e., if the problem is infeasible, one question is what the earliest times are at which units can be deployed. This question can be answered by solving the TMDPM. One other question to ask is what additional numbers of transportation assets of each type are required to carry out deployment within the required time windows. This can be obtained by solving the CMDPM without constraints (2) -(3) and with constraints that set lower limits on the number and type of transportation assets for each type to their available numbers. Another question in the case of infeasibility is to determine what units cannot be delivered. This can be found by using the output of the TMDPM. However, assigning priorities or penalties and then determining the units not delivered within their time windows may be of more value to decision makers, i.e., solving PMDPM or PnMDPM.

If the problem is feasible, one may ask what the earliest time is to deploy units and this can be answered by solving the LMDPM. In case of feasibility, the decision makers may want to know how long the movements of units may be delayed, i.e., how robust the plan is. This can be obtained by subtracting the arrival times of units from their latest permissible arrival times, i.e., slack times of units are found.

Regardless of whether or not a problem is feasible, the following appropriate what-if questions may be asked by the analysts or decision makers: What happens if the available numbers of some types of transportation assets are increased/decreased? What happens when the time windows of some units are enlarged/tightened? What happens if the processing capacity of an airfield/port is increased/decreased? What happens if a transportation asset not in the inventory is procured? What happens when the sizes of the deployment components are changed? What happens when the travel times on certain arcs are increased? What happens when the waiting time at a node suddenly increases? What happens if a bridge or an airport is destroyed by an attack? What happens when certain number of transportation assets are destroyed? What happens when a unit on the move is destroyed? What happens when some units are required to use certain routes or transportation assets?

The answers to the above questions change depending on whether the problem is feasible or depending on which model(s) is used. For example, increasing the available number of some transportation assets may turn an infeasible problem into a feasible one. On the other hand, in case of feasibility, this may draw the arrival times of the units to their earliest allowable times if the LMDPM is used. If the CMDPM is used, the additional transportation assets may not change anything if they are more expensive than the other ones. On the other hand, if they are cheaper, then the allocation of transportation assets to the units and hence arrival times at destinations may change. When the processing capacity of an airfield/airport is increased, an infeasible problem may become feasible. However, it may not have any effect on the solution if the airfield is not a busy one. If the waiting time at a node increases suddenly, a feasible problem may be infeasible. However, if the waiting time remains within the range of the parameters in which the problem is feasible, then only the arrival times of units may be increased.

Possible effects of all of the above questions may be answered similarly. However, they are beyond the scope of this study. The point is that one should not try to solve the models from scratch to answer the effects of all questions. The focus should be on determining the effects without having to solve the models from scratch. For example, when a bridge is destroyed, there is no need to change the deployment plans of all units but only the plans of those units that use the bridge. For this purpose, the deployment plans of all units not required to use the bridge are fixed including allocation of transportation assets and transportation infrastructure and the model is solved with fixed plans. If a unit required to use the bridge is on the move at the time the bridge is destroyed, the movement of the unit may be delayed if it is possible to repair the bridge in the slack time of the unit. If it is not possible to repair the bridge in the slack time, the model is solved by regarding the last location of the unit as the origin of the unit and the allocated transportation assets and deployment plans of all other units as fixed. Similar arguments may be made to answer the effects of other what-if questions.

CHAPTER 7

CONCLUSION

In this dissertation, we study the deployment planning problem (DPP) that may roughly be defined as the problem of the planning of the physical movement of military units, stationed at geographically dispersed locations, from their home bases to their designated destinations while obeying constraints on scheduling and routing issues as well as on the availability and use of various types of transportation assets that operate on a multimodal transportation network.

The DPP is a large-scale, real-world problem. However, there does not exist a complete and academic definition of the problem. Hence, we first give a detailed and formal description of the DPP in Chapter 1. In Chapter 2, we give an analysis of the problem with respect to the scientific literature. Our analysis shows that several features of the problem are studied individually in different research areas related to dynamic network flow, network design, vehicle routing, dynamic resource allocation, and mobility analysis problems. However, there seems to be no study that deals with the problem as a whole at a level of detail and complexity we have undertaken. The existing models are far from being sufficient to model and solve the DPP.

In Chapter 3, we first give an abstraction of the problem. Specifically, we define the underlying network, transportation assets, items (commodities) to be moved, and sets and data related to these three. The abstraction is such that it gives a basis for a database. Next, we define and formulate the *Cost Minimization Deployment Planning Model (CMDPM)* where the purpose is to plan the movements of units with a given fleet of transportation assets such that the sum of fixed and variable transportation costs is minimized. The CMDPM is expected to be of use for investment decisions in transportation resources during peacetime and for deployment planning in cases where the operation is not imminent and there is enough time to do a deliberate planning that takes costs into account. As the solution times of the CMDPM are too poor to be used in a real application, we develop a solution methodology based on an effective use of a relaxation and restriction of the model that significantly speeds up a CPLEX-based branch and bound. The solution times for intermediate sized problems are around one hour whereas it takes about a week in the Turkish Armed Forces to produce a suboptimal feasible solution based on trial and error methods.

In Chapter 4, we present two min-max models, *Lateness Minimization Deployment Planning Model (LMDPM)* and *Tardiness Minimization Deployment Planning Model (TMDPM)*. *Lateness* in the LMDPM is defined as the difference between the arrival time of a unit and its earliest allowable arrival time at its destination while *tardiness* in the TMDPM is defined as the difference between the arrival time of a unit and its latest allowable arrival time at its destination. In this regard, the objectives in the LMDPM and TMDPM are to minimize maximum lateness and tardiness, respectively. These models are expected to be of use in cases where quick deployment is of utmost concern. The LMDPM is an appropriate model for cases when the given fleet of transportation assets is sufficient to deploy units within their allowable time windows while the TMDPM is appropriate for cases when the given fleet is not sufficient. We solve the LMDPM and the TMDPM by using the same solution methodology developed in Chapter 3. Solution times for these models are also around one hour for intermediate sized problems.

In Chapter 5, we present a heuristic algorithm where the basic idea is to solve a model incrementally, adding one unit at a time, rather than solving the model for all units simultaneously. At iteration k, the model is solved with k deployment components given the arrival times of $k-1$ deployment components obtained in the previous $k-1$ iterations. The process goes on in the same manner until all deployment components are added to the model.

Our experience shows that the order in which the deployment components are added greatly affects the objective function value of the heuristic algorithm. However, computational results show that if the time windows and the number of transportation assets are not too restricted, the heuristic finds optimal or near-optimal solutions. Of the 15 test problems, 10 problems are solved to the optimality for the relaxation of the CMDPM and 12 problems are solved to optimality for the relaxations of the LMDPM and TMDPM. The solution times of the heuristic are also much better than those obtained by solving the models directly. The results regarding both the objective function values and the solution times show that the heuristic is highly likely to give good feasible solutions in short times.

In Chapter 6, we explain how the output of a model can be used to derive a deployment plan, what other variations of the models might be, what type of what-if questions can be answered, and how the models can be used in the bottom-up and top-down deployment planning approaches.

Our contribution in this study is many-fold. We formally describe a large-scale, real-world, and complicated problem for which an academic and detailed description does not exist. We organize and simplify the problem such that it is understandable. We do an academic analysis of the problem with respect to the literature in several related research areas. Our review shows that there does not exist a study that deals with the problem as a whole and hence the problem is new to the literature. Our review also shows that some properties of the problem may lead to new research topics in the well-known problems such as network flow, dynamic network flow, and vehicle routing and scheduling. We abstract the problem and develop large-scale MIP models that can be used for different purposes. We propose an effective solution methodology and heuristic to solve the models. Finally, we relate the models' solutions to the real world in such a way that a layman can understand and explain what kinds of what-if questions can be answered by using the models. Thus, we provide the basis for a decision support system together with its database structure that can be used by decisionmakers at different levels.

Some future research directions may be as follows. The models that we have developed are flow-based. Another approach in modelling the problem may be to use assignment variables instead of flow variables. Then, the results of both models may be compared. In the models, the network, items and transportation assets are represented in the detail that the results of the models can be used for operational planning. However, the models can easily be modified for tactical and strategic planning such that the network, transportation assets, and items are aggregated to a certain level. One study may use an aggregated formulation and compare the results to see how resolution is lost, if any. In the dissertation, we use a solution methodology based on an effective use of a relaxation and restriction of the model to solve the models. Another approach may be to use decomposition techniques such as Benders or Lagrangean relaxation. In the dissertation, we develop a heuristic whose results are dependent on the order in which deployment components are processed in the heuristic. However, we cannot make suggestions on an appropriate sequence. Hence, a study may focus on experimenting with the heuristic by using problems of different characteristics to come up with suggestions on the sequence in which deployment components are processed. Finally, another study may focus on how to modify a current solution given that some conditions in the problem change.

REFERENCES

1. Ahuja, R.K., Magnanti, T.L., Orlin, J.B., and Reddy, M.R. Applications of networks optimization. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors, *Network Models,* volume 7 of *Handbooks in Operations Research and Management Science*, North-Holland, Amsterdam, , pp. 1-83. 1995.

2. Arntzen, B.C., Brown, G.G., Harrison, T.P., and Trafton, L.L. Global supply chain management at Digital Equipment Corporation. *Interfaces* **25**, pp. 69-93, January-February 1995.

3. Aronson, J.E. A survey of dynamic network flows. *Annals of Operations Research* **20**, pp. 1-66, 1989.

4. Assad, A.A. Models for rail transportation. *Transportation Research A: Policy and Practice* **14**A, pp. 205-220, 1980.

5. Baker, L. Overview of computer based models applicable to freight car utilization. Report prepared for U.S. Department of Transportation, NTIS, Springfield VA, U.S.A, 1977.

6. Baker, S.F. A cascade approach for staircase linear programs with an application to Air Force mobility. Ph.D. Dissertation, Naval Postgraduate School, Monterey, California, 1997.

7. Baker, S.F., Rosenthal R.E. A cascade approach for staircase linear programs. Technical Report NPS-OR-98-004, Naval Postgraduate School, Monterey, California, 1998.

8. Baker, S.F., Morton D.P., Rosenthal R.E., and Williams L.M. Optimizing strategic airlift. Technical Report NPS-OR-99-004, Naval Postgraduate School, Monterey, California, 1999.

9. Baker, S.F., Morton D.P., Rosenthal R.E., and Williams L.M. Optimizing military airlift. *Operations Research* **50(4),** pp. 582-602, 2002.

10. Balakrishnan, A., Magnanti, T.L., and Wong, R.T. A dual-ascent procedure for large-scale uncapacitated network design. *Operations Research* **37(5)**, pp. 716-740, 1989.

11. Ball, M. and Roberts, A. A graph partitioning approach to airline crew scheduling. *Transportation Science* **19**, pp. 107-126, 1985.

12. Barnhart, C. and Talluri, K. Airlines Operations Research. In A.E. McGarity and C. ReVelle, editors, *Civil and Environmental Engineering Systems: An Advanced Applications Text*, John Wiley, New York, 1997.

13. Barnhart, C, Jin, H., and Vance, P.H. Railroad blocking: A network design application. *Operations Research* **48(4)**, pp. 603-614, 2000.

14. Beamon, B.M. Supply chain design and analysis: Models and methods. *International Journal of Production Economics* **55**, pp. 281-294, 1998.

15. Beaujon, G.J. and Turnquist, M.A. A model for fleet sizing and vehicle allocation. *Transportation Science* **25(l)**, pp. 19-45, 1991.

16. Bell, W., Dalberto, L., Fisher, M., Greenfield, A., Jaikumar, R., Kedia, P., Mack, R., and Prutzman, P. Improving the distribution of industrial gases with an online computerized routing and scheduling optimizer. *Interfaces* **13**, pp. 4-23, 1983.

17. Berlin, G.N. The use of directed routes for assessing escape potential. National Fire Protection Association, Boston, MA., 1979.

18. Bodin, L.D., Golden, B.L., Assad, A., and Ball, M. Routing and scheduling of vehicles and crews: the state of the art. *Computational Operations Research* **10**, pp. 63-211, 1983.

19. Bodin, L.D. Twenty years of routing and scheduling. *Operations Research* **38**, pp. 571-579, 1990.

20. Braklow, J.W., Graham, W.W., Hassler, S.M., Peck, K.E., and Powell, W.B. Interactive optimization improves service and performance for Yellow Freight System. *Interfaces* **22(1)**, pp. 147-172, 1992.

21. Brown, G. and Graves, G. Real-Time dispatch of petroleum tank trucks. *Management Science* **27**, pp. 19-32, 1981.

22. Burkard, R.E., Dlaska, K., and Klinz, B. The Quickest flow problem. *Methods and Models of Operations Research* **37**, pp. 31-58, 1993.

23. Campbell, J.F., Ernst, A.T., and Krishnamoorthy, M. Hub location problems. In Drezner, Z. and Hamacher, H., editors, *Facility Location:Applications and Theory*, Springer-Verlag, Berlin, 2002.

24. Carvalho, T.A. A New Approach to Solving Dynamic Resource Allocation Problems. PhD Thesis, Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ, U.S.A., 1996.

25. Carvalho, T.A. and Powell, W.B. A multiplier adjustment method for dynamic resource allocation problems. *Transportation Science* **34(2)**, pp. 150-164, 2000.

26. Cascetta, E. *Transportation Systems Engineering: Theory and Methods.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001.

27. Chalmet, L.G., Francis, R.L., and Saunders, P.B. Network models for building evacuation. *Management Science* **28**, pp. 86-105, 1982.

28. Chen, Y.L. and Chin, Y.H. The quickest path problem. *Computers and Operations Research* **17**, pp. 153-161, 1990.

29. Chih, K. C.-K. A real time dynamic optimal freight car management simulation model of the multiple railroad, multicommodity, temporal spatial network flow problem. PhD Thesis, Princeton University, Princeton, NJ, U.S.A., 1986.

30. Cohen, M.A., Kamesam, P.V., Kleindorfer, P., Lee, H.L., and Tekerian, A. Optimizer: IBM's multi-echelon inventory system for managing service logistics. *Interfaces* **20**, pp. 65-82, January-February 1990.

31. Congressional Budget Office of the Congress of the United States. Moving US Forces: options for strategic mobility. February 1997. ftp://ftp.cbo.gov/0xx/doc11/StratMob.pdf

32. Cordeau, J.-F., Toth, P., and Vigo, D. A survey of optimization models for train routing and scheduling. *Transportation Science* **32(4)**, pp. 380-404, 1998.

33. Crainic, T.G. A comparison of two methods for tactical planning in rail freight transportation. In J.P. Brans, editor, *Operational Research'84,* North-Holland, Amsterdam, pp. 707-720, 1984.

34. Crainic, T.G., Ferland, J.-A., and Rousseau, J.-M. A tactical planning model for rail freight transportation. *Transportation Science* **18(2)**, pp. 165-184, 1984.

35. Crainic, T.G. and Rousseau, J.-M. Multicommodity, multimode freight transportation: A general modeling and algorithmic framework for the service network design problem. *Transportation Research B: Methodology* **20B**, pp. 225-242, 1986.

36. Crainic, T.G. and Rousseau, J.-M. The column generation principle and the airline crew scheduling. *INFOR* **25(2)**, pp. 136-151, 1987.

37. Crainic, T.G. Rail tactical planning: Issues, models and tools. In Bianco, L. and Bella, A. L., editors, *Freight Transport Planning and Logistics,* Springer-Verlag, Berlin, pp. 463-509, 1988.

38. Crainic, T.G., Florian, M., and Leal, J.-E. A model for the strategic planning of National Freight Transportation by rail. *Transportation Science* **24(l)**, pp. l-24, 1990.

39. Crainic, T.G., Florian, M., Guelat, J., and Spicss, H. Strategic planning of freight transportation: STAN, An interactive-graphic system. *Transportation Research Record* 1283, pp. 97-124, 1990.

40. Crainic, T.G. and Roy, J. Design of regular intercity driver routes for the LTL motor carrier industry. *Transportation Science* **26(4),** pp. 280-295, 1992.

41. Crainic, T.G., Florian, M., and Larin, D. STAN: New developments. In S. Khade and R. Brown, editors, *Proceedings of the 23rd Annual Meeting of the Western Decision Sciences Institute,* School of Business Administration, California State University, Stanislaus CA, pp. 493-498, 1994.

42. Crainic, T.G. and Laporte, G. Planning models for freight transportation. *European Journal of Operational Research* **97(3)**, pp. 409-438, 1997.

43. Crainic, T.G., Dufour, G., Florian, M., and Larin, D. Path recovery/reconstruction and applications in nonlinear multimodal multicommodity networks. In Gendreau, M. and Marcotte, P., editors, *Transportation and Network Analysis - Current Trends*, Kluwer Academic Publishers, Norwell MA, (2002).

44. Crainic, T. A survey of optimization models for freight transportation. In: Hall RW, editor, *Handbook of transportation secience*, 2 nd Edition, Dordrecht: Kluwer Academic Pub, pp. 451-516, March 2003.

45. Croom, S., Romano P., and Giannakis, M. Supply chain management: an analytical framework for critical literature review. *European Journal of Purchasing & Supply Management* **6**, pp. 67-83, 2000.

46. Daskin, M.S. Network and discrete location: models, algorithms, and applications. John Wiley & Sons, New York, NY, 1995.

47. Dejax, P.J. and Crainic, T.G. A review of empty flows and fleet management models in freight transportation. *Transportation Science,* **21(4)**, pp. 227-247, 1987.

48. Delorme, L., Roy, J., and Rousseau, J.-M. Motor carrier operation planning models: a state of the art. In Bianco, L. and Bella, A. L., editors, *Freight Transport Planning and Logistics,* Springer-Vcrlag, Berlin, pp. 510-545, 1988.

49. Desaulniers, G., Desrosiers, J., Ioachim, I., Solomon, M.M., Soumis, F., and Villeneuve, D. A united framework for deterministic time constrained vehicle routing and crew scheduling problems. In T.G. Crainic and G. Laporte, editors, *Fleet Management and Logistics,* Kluwer Academic Publishers, Norwell, MA, pp. 57-93, 1998b.

50. Desrochers, M., Lenstra, J.K., Savelsbergh, M.W.P., and Soumis, F. Vehicle routing with time windows. In Golden, B.L., Assad, A.A., editors, *Vehicle Routing: Methods and Studies*, Elsevier Science, 1988.

51. Desrochers, M., Lenstra, J.K., and Savelsbergh, M.W.P. A classification scheme for vehicle routing and scheduling problems. *European Journal of Operations Research* **46**, pp. 322-332, 1990.

52. Desrochers, M., Jones, C.V., Lenstra, J.K., Savelsbergh, M.W.P., and Stougie, L. Towards a model and algorithm management system for vehicle routing and scheduling problems. Working Paper, GERAD, Montreal, Canada, 1998.

53. Desrosiers, J., Dumas, Y., Solomon, M.M., and Soumis, F. Time constrained routing and scheduling. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors, *Network Routing,* volume 8 of *Handbooks in Operations Research and Management Science,* North-Holland, Amsterdam, pp. 35-139, 1995.

54. Drezner, Z., editor. Facility location. A survey of applications and methods. Springer-Verlag, New York, NY, 1995.

55. Dror, M. and Trudeau, P. Split Delivery Routing. *Naval Research Logistics* **37**, pp. 383-402, 1990.

56. Farvolden, J.M. and Powell, W.B. Subgradient methods for the service network design problem. *Transportation Science* **28(3)**, pp. 256-272. 1994.

57. Federgruen, A. and Simchi-Levi, D. Analysis of vehicle routing and inventory-routing problems. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., Editors, *Network Routing,* Volume 8 of *Handbooks in Operations Research and Management Science,* North-Holland, Amsterdam, pp. 297-373, 1995.

58. Fisher, M.L. Vehicle routing. In Ball, M.O., Editor, *Handbooks in OR and MS*, Vol. 8, Elsevier Science, 1995.

59. Fisher, M.L., Tang, B., and Zheng, Z., "A network Flow Based Heuristic for Bulk Pick-up and Delivery Routing," *Transportation Science* **29,** pp. 45-55, 1995.

60. Fleischer, L. and Skutella, M. The Quickest multicommodity flow problem. In Cook, W.J., Schulz, A.S., Editors, *Integer Programming and Combinatorial Optimization*, Volume 2337 of *Lecture Notes in Computer Science*, Springer, Berlin, pp. 36-53, 2002.

61. Fleischer, L., Skutella, M. Minimum cost flows over time without intermediate storage. In Proceedings *of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, Baltimore, MD, pp. 66-75, 2003.

62. Florian, M. and Hearn, D. Networks equilibrium models and algorithms. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., Editors, *Network Routing*, volume 8 of *Handbooks in Operations Research and Management Science*, North Holland. Amsterdam, pp.485-550, 1995.

63. Ford, L.R. and Fulkerson, D.R. Constructing maximal dynamic flows from static flows. *Operations Research* **6**, pp.419-433, 1958.

64. Ford, L.R. and Fulkerson, D.R. *Flows in Networks*, Princeton University Press, Princeton, NJ, 1962.

65. Fuller, D. Reduction of a large-scale global mobility optimization model through aggregation. Master's Thesis, Naval Postgraduate School, Monterey, California, 1996.

66. Gale, D. Transient flows in networks. *Michigan Mathematical Journal* **6**, pp. 59-63, 1959.

67. Gendreau, M., Guertin, F., Potvin, J.Y., and Taillard, E. Parallel tabu search for real-time vehicle routing and dispatching. *Transportation Science* **33**, pp. 381-390, 1999.

68. Goggins D. Stochastic modeling for airlift mobility. Master's Thesis, Naval Postgraduate School, Monterey, California, 1995.

69. Golden, B.L. and Assad, A., editors. Special issue on time windows. *American Journal of Mathematical Management Analysis* **6**, pp. 251-399, 1986.

70. Golden B.L. and Assad, A. *Vehicle routing: methods and studies*. North Holland, Amsterdam, 1988.

71. Gorman, M.F. An application of genetic and tabu searches to the freight railroad operating plan problem. *Annals of Operations Research* **78**, pp. 51-69, 1998.

72. Gorman, M.F. Santa Fe Railway uses an operating-plan model to improve its service design. *Interfaces* **28(4),** pp. 1—12**,** 1998a.

73. Grünert, T. and Sebastian, H.-J. Planning models for long-haul operations of postal and express shipment companies. *European Journal of Operational Research* **122,** pp. 289-309, 2000.

74. Guelat, J., Florian, M., and Crainic, T.G. A multimode multiproduct network assignment model for strategic planning of freight flows. *Transportation Science* **24(l)**, pp. 25-39**,** 1990.

75. Haghani, A.E. Formulation and solution of combined train routing and makeup, and empty car distribution model. *Transportation Research B: Methodological* **23B(6)**, pp. 433-431, 1989.

76. Hall, A., Hippler, S., and Skutella, M. Multicommodity flows over time: efficient algorithms and complexity, In Lenstra, J.K., Woeginger, G.J., Editors, *Automata, Languages, and Programming, Lecture Notes in Computer Science*, Springer, Berlin, 2003.

77. Halpern, J. A generalized dynamic flows problem. *Networks* **9**, pp. 133-167, 1979.

78. Hane, C.A., Barnhart. C., Johnson, E.L., Marsten, R.E., Nemhauser, G.L., and Sigismondi, G. A fleet assignment problem: solving a large-scale integer program. *Mathematical Programming* **70**, pp. 211-232, 1995.

79. Heal, G.N. and Garnett, I.K. *Allied Deployment and Movement System (ADAMS) Version 3.0 Tutorial*, Technical Note 669, NATO Consultation, Command and Control Agency, The Hague, September 2001.

80. Herren, H. Computer controlled empty wagon distribution on the SSB. *Rail International* **8(1)**, pp. 25—32, 1977.

81. Herren, H. The distribution of empty wagons by means of computer: an analytical model for the Swiss Federal Railways (SSB). *Rail* International **4(l)**, pp. 1005-1010, 1973.

82. Holmberg, K., Joborn, M., and Lundgren, J.T. Improved empty freight car distribution. *Transportation Science* **32(2)**, pp. 163-173, 1998.

83. Hung, C.-Y. and Chen, G.-H. On the quickest path problem. In *Proc. ICCI Lecture Notes in Computer Science 497* , Springer-Verlag, pp.44-46, 1991.

84. Hoppe, B. *Efficient Dynamic Network Flow Algorithms*. PhD Thesis, Cornell University, 1995.

85. Hoppe, B. and Tardos, E. The Quickest Transshipment Problem. *Mathematics of Operations Research* **25**, pp. 36-62, 2000.

86. Jarvis, J.R. and Ratliff, D.H. "Some Equivalent Objectives for Dynamic Network Flow Problems," *Management Science* **28**, pp. 106-109, 1982.

87. Joborn, M. Empty Freight Car Distribution at Swedish Railways - Analysis and Optimization Modeling. PhD thesis, Division of Optimization, Department of Mathematics, University of Linkoping, Linkoping, Sweden, 1995.

88. Joborn, M., Crainic, T.G., Gendreau, M., Holmberg, K., and Lundgren, J.T. Economies of Scale in Empty Freight Car Distribution in Scheduled Railways. Publication CRT-2001-12, Centre de recherche sur les transports, University de Montreal, Montreal, QC, Canada, 2001.

89. Jordan, W.C. and Turnquist, M.A. A stochastic dynamic network model for railroad car distribution. *Transportation Science* **17, pp.** 123-145, 1983.

90. Keaton, M.H. Designing optimal railroad operating plans: Lagrangian relaxation and heuristic approaches. *Transportation Research B: Methodological* **23B(6)**, pp. 415-431, 1989.

91. Keaton, M.H. Service-cost trade-offs for carload freight traffic in the U.S. Rail Industry. *Transportation Research A: Policy and Practice* **25A(6)**, pp. 363-374, 1991.

92. Keaton, M.H. Designing optimal railroad operating plans: a dual adjustment method for implementing Lagrangian relaxation. *Transportation Science* **26**, pp. 263-279, 1992.

93. Killingsworth, P., Henry, K., Melody, L., and Stucker, J. Tankers: air mobility roles for the 1990s. Report DRR-913-AF, Rand Corporation, Santa Monica, California, 1994.

94. Kim, D., Barnhart, C., Ware, K., and Reinhardt, G. Multimodal express package delivery: a service network design application. *Transportation Science* **33(4)**, pp. 391-407, 1999.

95. Klinz, B. and Woeginger, G.J. Minimum cost dynamic flows: the series-parallel case. In Balas, E., Clausen, J., Editors, *Integer Programming and Combinatorial Optimization*, volume 920 of *Lecture Notes in Computer Science*, Springer, Berlin, pp. 329-343, 1995.

96. Kress, M. Optimizing supply chains in military operations. 4 January 2005. http://www.ima.umn.edu/talks/workshops/9-23-27.2002/kress/kress.ppt

97. Kuby, M.J. and Gray, R.G. The hub network design problem with stopovers and feeders: the case of Federal Express. *Transportation Research A: Policy and Practice* **27A(1),** pp. 1-12, 1993.

98. Labbe, M., Peeters, D., and Thisse, J.-F. Location on networks. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., Editors, *Network Routing,* volume 8 of *Handbooks in Operations Research and Management Science,* North-Holland, Amsterdam, pp. 551-624., 1995.

99. Labbe, M. and Louveaux, F.V. Location problems. In DeirAmico, M., Mafnoli, F., and Martello, S., Editors, *Annotated Bibliographies in Combinatorial Optimization,* John Wiley & Sons, New York, NY, pp. 261-281, 1997.

100. Lamar, B.W., Sheffi, Y., and Powell, W.B. A capacity improvement lower bound for fixed charge network design problems. *Operations Research* **38(4)**, pp. 704-710, 1990.

101. Laporte, G. Location-Routing Problems. In Golden, B.L. and Assad, A.A., Editors, *Vehicle Routing: Methods and Studies,* North-Holland, Amsterdam, pp. 163-197, 1988.

102. Laporte, G. The vehicle routing problem: an overview of exact and approximate algorithms. *European Journal of Operations Research* **59**, pp. 994-1004, 1992.

103. Leddon, C.D. and Wrathall, E. Scheduling empty freight car fleets on the Louisville and Nashville Railroad. In *Second International Symposium on the Use of Cybernetics on the Railways*, pp. 1-6, 1967.

104. Lee, H.L. and Billington, C. The evolution of supply chain management models and practice at Hewlett-Packard. *Interfaces* **25**, pp. 42-63, September-October 1995.

105. Magnanti, T.L. and Wong, R.T. Network design and transportation planning: models and algorithms. *Transportation Science* **18(l)**, pp. l-55. 1986.

106. Malandraki, C. and Daskin, M.S. Time dependent vehicle routing problems: formulations, properties, and heuristics algorithms. *Transportation Science* **26**, pp. 185-199, 1992.

107. Markowicz, B.P. and Turnquist, M.A. Applying the LP solution to the daily distribution of freight cars. Technical Report, Cornell University, 1990.

108. Marsten, R.E. and Shepardson, F. Exact solution of crew scheduling problems using the set partitioning model: Recent successful applications. *Networks* **11**, pp. 167-177, 1981.

109. Martin, C.H., Dent, D.C., and Eckhart, J.C. Integrated production, distribution, and inventory planning at Libbey-Owens-Ford. *Interfaces* **23**, pp. 68-78, May-June 1993.

110. McGaughey, K., Gohring, W., and McBrayer, R.N. Planning locomotive and caboose distribution. *Rail International* **3**, pp. 151-158, 1973.

111. McKinzie, K. and Barnes, J.W. A review of strategic mobility models supporting the defense transportation system. June 2003.

http://www.me.utexas.edu/~barnes/research/files/update_06_2003/RSMMSDTS.pdf

112. Min, H. and Zhou, G. Supply chain modeling: past, present and future. *Computers and Industrial Engineering* **43**, pp. 231-249, 2003.

113. Minieka, E. Dynamic network flows with arc changes. *Networks* **4**, pp. 255-265, 1974.

114. Minieaka, E. Maximal, lexicographic, and dynamic network flows. *Operations Research* **21**, pp. 517-527, 1979.

115. Minoux, M. Network synthesis and optimum network design problems: models, solution methods and applications. *Networks* **19**, pp. 313-360. 1986.

116. Mirchandani, P.S. and Francis, R.L., editors. *Discrete Location Theory.* John Wiley & Sons, New York, NY, 1990.

117. Misra, S. Linear programming of empty wagon disposition. *Rail International* **3**, pp. 151-158, 1972.

118. Morlok, E.K. and Peterson, R.B. A final report on a development of a geographic transportation network generation and evaluation model. In *Proceedings of the Eleventh Annual Meeting*, Transportation Research Forum, pp. 99-103, 1970.

119. Morton, D.P., Rosenthal, R.E., and Weng, L.T. Optimization modeling for airlift mobility. *Military Operations Research* **1,** pp. 49-68, 1996.

120. Newton, H.N., Barnhart, C., and Vance, P.H. Constructing railroad blocking plans to minimize handling costs. *Transportation Science* **32(4)**, pp. 330-345, 1998.

121. Nemhauser, G.L. and Wolsey, L.A. *Integer and Combinatorial Optimization.* Wiley, New York, NY, 1988.

122. Niemi, A. Stochastic Modeling for the NPS/RAND Mobility Optimization Model. August 2000. http://ie.engr.wisc.edu/robinson/Niemi.htm.

123. Nobert, Y. and Roy, J. Freight handling personnel scheduling at air cargo terminals. *Transportation Science* **32(3)**, pp. 295-301, 1998.

124. O'Kelly, M.E. A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research* **32**, pp. 393-404, 1987.

125. Özdamar, L. and Ekinci, E. A logistics planning decision support system for deployment of military forces in military crises and natural disasters. *Savunma Bilimleri Dergisi* (Defense Sciences Journal) **1 (1)**, pp. 1-28, May 2002.

126. Powell, W.B. and Sheffi, Y. The load-planning problem of motor carriers: problem description and a proposed solution approach. *Transportation Research A: Policy and Practice* **17(6)**, pp. 471-480, 1983.

127. Powell, W.B. and Sheffi, Y. Interactive optimization for motor carrier load planning. *Journal of Business Logistics* **7(2)**, pp. 64-90, 1986.

128. Powell, W.B. A stochastic model of the dynamic vehicle allocation problem. *Transportation Science* **20**, pp. 117-129, 1986.

129. Powell, W.B. A local improvement heuristic for the design of less-than-truckload motor carrier networks. *Transportation Science* **20(4)**, pp. 246-357, 1986a.

130. Powell, W.B. and Sheffi, Y. Design and implementation of an interactive optimization system for the network design in the motor carrier industry. *Operations Research* **37(l)**, pp. 12-29, 1989.

131. Powell, W.B., Jaillet, P., and Adoni, A. Stochastic and dynamic networks and routing. In Ball, M.O., Magnanti, T.L., Monma, C.L., Nemhauser, G.L., Editors, *Handbooks in Operations Research and Management Science: Networks*, Elsevier Science Publishers B.V., 1995.

132. Powell, W.B., Carvalho, T.A., Godfrey, G.A., and Simao, H.P. Dynamic fleet management as a logistics queueing network. *Annals of Operations Research* **61**, pp. 165-188, 1995a.

133. Powell, W.B. A stochastic formulation of the dynamic assignment problem, with an application to truckload motor carriers. *Transportation Science* **30 (3)**, pp. 195-219, 1996.

134. Carvalho, T. and Powell, W.B. Dynamic control of multicommodity logistics queueing networks. *European Journal of Operations Research* **98 (3)**, 1997.

135. Powell, W.B. and Carvalho, T.A. Real-time optimization of containers and flatcars for intermodal operations. *Transportation Science* **32 (2)**, pp. 110-126, May 1998.

136. Powell, W.B. and Carvalho, T. Dynamic control of logistics queuing networks for large scale fleet management. *Transportation Science* **32 (2)**, pp. 90-109, 1998a.

137. Powell, W.B. Dynamic models of transportation operations, prepared for Handbooks in Operations Research Volume on Supply Chain Management, October 2002.

138. Ribeiro, C.C. and Soumis, F. A column generation approach to the multi-depot vehicle scheduling problem. *Operations Research* **42,** pp. 41-52, 1994.

139. Rodriguez, P., Nussbaum, M., Baeza, R., Leon, G., Sepulveda, M., and Cobian, M. Using global search heuristics for the capacity vehicle routing problem. *Computers and Operations Research* **25**, pp. 407-417, 1998.

140. Rosen, J.B., Sun, S.-Z., and Xue, G.-L. Algorithms for the quickest path problem and the enumeration of quickest paths. *Computers and Operations Research* **18**, pp. 579-584, 1991.

141. Roy, J. and Delorme, L. NETPLAN: A network optimization model for tactical planning in the less-than-truckload motor-carrier industry. *INFOR* **27(l)**, pp. 22-35, 1989.

142. Salkin, H.M. and Mathur, K. *Foundations of Integer Programming.* North-Holland, Amsterdam, 1989.

143. Schank, J., Mattock M., Sumner, G., Greenberg, I., Rothenberg, J., and Stucker, J.P. A review of strategic mobility models and analysis. Report R-3926-JS, RAND Corporation, Santa Monica, California, 1991.

144. Shan, Y. A dynamic multicommodity network flow model for real-time optimal rail freight car management. PhD Thesis, Princeton University, Princeton, NJ, U.S.A., 1985.

145. Shapiro, J.F. Approaches for integrating inventory with other supply chain decisions, 2003. Accessed at http://shapiro.ASCET.com.

146. Slats, P.A., Bhola B., Evers J.J.M., and Dijkhuizen, G. Logistics chain modeling. *European Journal of Operational Research* **87**, pp. 1-20, 1995.

147. Solomon, M. Algorithms for the vehicle routing and scheduling problem with time window constraints. *Operations Research* **35**, pp. 254-265, 1987.

148. Stevens, G.C. Integrating the supply chain. *International Journal of Physical Distribution and Materials Management* **19(8)**, pp. 3-8, 1998.

149. Tan, K.C. A framework of supply chain management literature. *European Journal of Purchasing & Supply Management* **7,** pp. 39-48, 2001.

150. Thomas, D.J. and Griffin, P.M. Coordinated supply chain management. *European Journal of Operational Research* **94**, pp. 1-15, 1996.

151. Toy, A. O. Route prioritization for an airlift mobility optimization model. Master's Thesis, Naval Postgraduate School, Monterey, California, 1996.

152. Turker, Y. Route and column generation methods for airlift mobility optimization. Master's Thesis, Naval Postgraduate School, Monterey, California, 1995.

153. Turnquist, M.A. and Markowicz, B.P. An interactive microcomputer-based planning model for railroad car distribution. Technical Report, Cornell University, 1989.

154. Turnquist, M.A. Economies of scale and network optimization for empty car distribution. Technical Report, Cornell University, 1994.

155. Weng L.T. Strategic airlift optimization model. Master's Thesis, Naval Postgraduate School, Monterey, California, 1994.

156. White, W.W. and Bomberault, A.M. A network algorithm for empty freight car allocation. *IBM Systems Journal* **8 (2)**, pp. 147-171, 1969.

157. White, W.W. A program for empty freight car allocation. Technical Report 360D.29.002, IBM Contributed Program Library, IBM Corporation, Program Information Department, Hawthorne, N.Y., 1968.

158. White, W. Dynamic transshipment networks: An algorithm and its application to the distribution of empty containers. *Networks* **2(3)**, pp. 211-236, 1972.

159. Wilkinson, W.L. An algorithm for universal maximum dynamic flows in a network. *Operations Research* **19**, pp. 1602-1612, 1971.

160. Wing, V.F., Rice, R.E., Sherwood, R., and Rosenthal, R.E. Determining the optimal mobility fix. Joint Staff (J8), Force Design Division, The Pentagon, Washington, DC, 1991.

161. Yost, K.A. The THRUPUT strategic airlift flow optimization model. Air Force Studies and Analyses Agency, The Pentagon, Washington, DC, 1994.