Distribution Planning for Rail and Truck Freight Transportation Systems

Yazhe Feng

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Industrial and Systems Engineering

Kimberly P. Ellis, Chair
Ebru K. Bish
Roberta S. Russell
Yasemin M. Uzgoren

June 21, 2012
Blacksburg, Virginia

Keywords: Transportation, Railroad Trip Planning, Industrial Gas Distribution, Fleet Planning, Heuristic

Distribution Planning for Rail and Truck Freight Transportation Systems

Yazhe Feng

(ABSTRACT)

Rail and truck freight transportation systems provide vital logistics services today. Rail systems are generally used to transport heavy and bulky commodities over long distances, while trucks tend to provide fast and flexible service for small and high-value products. In this dissertation, we study two different distribution planning problems that arise in rail and truck transportation systems.

In the railroad industry, shipments are often grouped together to form a block to reduce the impact of reclassification at train yards. We consider the time and capacity constrained routing (TCCR) problem, which assigns shipments to blocks and train-runs to minimize overall transportation costs, while considering the train capacities and shipment due dates. Two mathematical formulations are developed, including an arc-based formulation and a path-based formulation. To solve the problem efficiently, two solution approaches are proposed. The sequential algorithm assigns shipments in order of priority while considering the remaining train capacities and due dates. The bump-shipment algorithm initially schedules shipments simultaneously and then reschedules the shipments that exceed the train capacity. The algorithms are evaluated using a data set from a major U.S. railroad with approximately 500,000 shipments. Industry-sized problems are solved within a few minutes of computational time by both the sequential and bump-shipment algorithms, and transportation costs are reduced by 6% compared to the currently used trip plans.

For truck transportation systems, trailer fleet planning (TFP) is an important issue to improve services and reduce costs. In this problem, we consider the quantities and types of trailers to purchase, rent, or relocate among depots to meet time varying demands. Mixed-integer programming models are developed for both homogeneous and heterogeneous TFP problems. The objective is to minimize the total fleet investment costs and the distribution costs across multiple depots and multiple time periods.

For homogeneous TFP problem, a two-phase solution approach is proposed. Phase I concentrates on distribution costs and determines the suggested fleet size. A sweep-based routing heuristic is applied to generate candidate routes of good quality. Then a reduced mathematical model selects routes for meeting customer demands and determines the preferred fleet size. Phase II provides trailer purchase, relocation, and rental decisions based on the results of Phase I and relevant cost information. This decomposition approach removes the interactions between depots and periods, which greatly reduces the complexity of the integrated optimization model.

For the heterogeneous TFP problem, trailers with different capacities, costs, and features are considered. The two-phase approach, developed for the homogeneous TFP, is modified. A rolling horizon scheme is applied in Phase I to consider the trailer allocations in previous periods when determining the fleet composition for the current period. Additionally, the sweep-based routing heuristic is also extended to capture the characteristics of continuous delivery practice where trailers are allowed to refill products at satellite facilities. This heuristic generates routes for each trailer type so that the customer-trailer restrictions are accommodated. The numerical studies, conducted using a data set with three depots and more than 400 customers, demonstrate the effectiveness of the two-phase approaches. Compared to the integrated optimization models, the two-phase approaches obtain quality solutions within a reasonable computational time and demonstrate robust performance as the problem sizes increase. Based on these results, a leading industrial gas provider is currently integrating the proposed solution approaches as part of their worldwide distribution planning software.

Dedication

To my parents

Lihe Feng & Caixia Fan

for their love and support

Acknowledgments

This dissertation would not have been finished without the guidance of my committee members, help from friends, and support from my family. I would like to express my sincerely gratitude to these individuals for their support and assistance.

Dr. Kimberly Ellis has been a strong and supportive advisor to me throughout my graduate studies. She has always given me great freedom to think and work independently, and she has also provided her insightful comments when I encountered difficulties. I appreciate the great opportunities that she provided to work with industrial companies, which made my research more interesting and meaningful. To me, Dr. Ellis is not only an academic mentor, but she is always a strong support in my studies, life, and career. When I felt distressed and depressed, I was influenced and encouraged by her energetic voice after our "marathon" meetings. She has made my life in the U.S. easier and more delightful.

I would also like to express my sincere gratitude to my committee members for their contributions and support. Dr. Yasemin Merzifonluoglu (I would like to use her original name here even it has been replaced) started my academic research on the right foot and provided me confidence that I could complete research work successfully. Dr. Ebru Bish has provided me warmth and understanding during the most distressing time in the Ph.D. journey. I still remember she told me that "Ph.D. work is hard but it is a great opportunity to grow and learn." These words have given me immeasurable courage when I encountered challenges. I also appreciate Dr. Robin Russell for her input, valuable discussions, and accessibility. She generously gave her expertise to improve my work and helped me gain a different perspective on the research problems.

Undoubtedly, my parents have always been the strongest support in all my endeavors. While I have been thousands miles away from them, their love has helped me gain so much drive and an ability to tackle challenges head on.

I would like to thank all my friends I've made through my life. Thank you, Shiguang Xie, for always cheering me up and standing by me through the good and bad times. Thank you, Charlie Crawford, Andrew Henry, and Steven Roesch. Without you, the lab would have been much more lonely and boring. In addition, many thanks to Shengzhi Shao, Ming Cheng, Sourish Sakar, Lingrui Liao, Yanfeng Li, Juqi Liu, Xiaomeng Chang, and all good friends. Regardless of where you are, in Blacksburg or in other places, in the U.S. or in China, I appreciate all the time we have spent together and I wish all of you the brightest future.

Lastly, I also gratefully acknowledge the institutional support that I have received for the research involved in this dissertation. In particular, I thank Innovative Scheduling and Air Liquide for providing valuable suggestions to improve my work and supporting me with generous assistantships. The collaboration with them made my research more meaningful and successful. I also appreciate the funding and fellowships provided by the National Science Foundation, the Institute of Industrial Engineers E.J. Sierleja Memorial Fellowship, and the Grado Department of Industrial and Systems Engineering Dover Fellowship.

Contents

List of Figures

List of Tables

Chapter 1

Overview of Transportation in Business Logistics

The Council of Logistics Management (1991) defined *business logistics* as the process of planning, implementing, and controlling the efficient and effective flow and storage of raw materials, in-process inventory, finished goods, services, and related information from the point of origin to the point of consumption for the purpose of conforming to customer requirements. Effective logistics processes integrate transportation, inventory levels, warehouse space, materials handling systems, packaging, and other related activities to meet cost expectations and service requirements.

Transportation systems play an important role in logistics system. The management of transportation is concerned with planning and controlling of the movement systems used by a company to achieve their logistics objectives. An effective transportation system sends goods to the right place at right time in order to satisfy customer demands and provides a connection between producers and consumers.

Transportation costs comprise a substantial portion of the costs of business logistics systems. The 21st Annual State of Logistics Report [6], released by the Council of Supply Chain Management Professionals (CSCMP), states that the cost of the United States business lo-

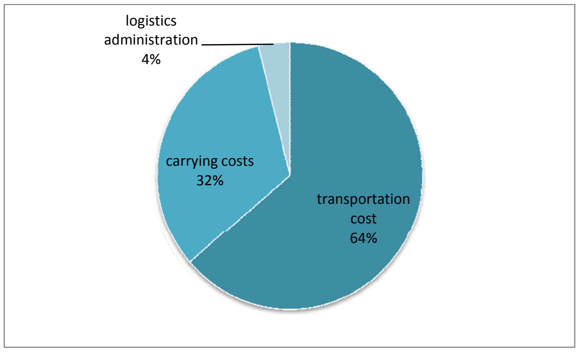

Figure 1.1: Components of Logistics Costs in 2009

gistics system from 2005 to 2008 remains between 9 percent and 10 percent of the United States gross domestic product (GDP). Generally, freight movement accounts for around one third to two thirds of total logistics costs [15]. As shown in Figure 1.1, transportation costs in 2009 accounts for 64 percent of the total logistics costs, while inventory-carrying costs accounts for 32 percent and logistics administration accounts for 4 percent.

A variety of options are available for individuals, companies, or countries to move products from one point to another. One or more of five transportation modes - truck, rail, air, water, or pipeline - may be selected. Table 1.1 summarizes the main characteristics of the five transportation modes. Trucks and air tend to be fast but costly, and are generally used to handle small and high-value products. Rail and water carriers are more suitable for movements of heavy, bulky, low-value-per-unit commodities in situations where speed is not of primary importance. Pipelines offer shippers a high level of service dependability at a relatively low cost. However, only a limited number of products can be transported by pipelines, including natural gas, crude oil, petroleum products, water, chemicals, and slurry products. In addition to single mode service, modal combinations are also available in freight transportation, including rail-truck, truck-water, truck-air, and rail-water. Such *intermodal* combinations offer specialized or lower cost services that are not generally available when using a single transport mode.

Table 1.2 summarizes the shipment characteristics transported by different modes in

Table 1.1: Comparison of Transportation Modes

Modes	Advantages	Disadvantages
Rail	- low cost (on a weight basis) - energy efficiency	- higher loss and damage ratio - limited to fixed track facilities - longer transited time and frequency of service
Truck	- flexibility, door-to-door delivery - versatile in transporting many sizes and weights - reliable service with little damage	- air pollution - congestion on road - high cost
Air	- shortest time-in-transit - high-level service - international shipping	- terminal, delivery delay, congestion - high cost
Water	- low cost - international shipping	- slow - limited movement by waterways
Pipeline	- low cost - dependability - rare damage	- limited number of products - one direction

Table 1.2: Shipment Characteristics by Mode of Transportation: 2002 and 2007

Transportation	Value (mil. dol.)		Tons (1,000)		Ton-miles (mil.)	
Modes	2002	2007	2002	2007	2002	2007
Truck	6,235,001	8,335,789	7,842,836	8,778,713	1,255,908	1,342,104
Rail	310,884	436,420	1,873,884	1,861,307	1,261,612	1,344,040
Water	89,344	114,905	681,227	403,639	282,659	157,314
Air	264,959	252,276	3,760	3,611	5,835	4,510
Pipeline	149,195	399,646	684,953	650,859	(S)	(S)

S Data do not meet publication standards due to high sampling variability or other reasons.
Source: U.S. Department of Transportation, Bureau of Transportation Statistics, and U.S. Census Bureau, 2007 Commodity Flow Survey. Transportation Commodity Flow Survey, preliminary; http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml ?pid=CFS_2007_00P1&prodType=table

2002 and 2007. As shown, trucks play a dominant role in freight transportation in terms of value, tons, and ton-miles[1]. Railroads are similar to trucks when considering ton-miles, since railroads often transport heavy shipments over long distances. In this dissertation, we focus on these two important transportation modes: rail and truck.

1.1. Railroad Industry

Railroads played a significant role in the economic and social development of the United States for over a century (1850-1950). Since World War II, rail has transported about two-thirds of the ton-mile traffic in the United States. During the last half of the twentieth century, however, the railroad industry declined in relative importance. In 1997, railroads shipped about 40.6 percent of all ton-miles moved by all transport modes in the United States, which is approximately 39 percent less than 1929 on a relative basis [2]. In 2007, freight rail systems carried 29 percent of ton-miles, 14 percent of the nation's freight by

[1]Ton-mile is a unit of freight transportation equivalent to a ton of freight moved one mile.

tonnage, and 5 percent of value [4]. However, it is important to note that the actual rail ton-miles have been steadily increasing and railroads are still the largest carrier in terms of intercity ton-miles.

With large carrying capacity, the railroads are able to handle large-volume movements of raw materials or low-value manufactured commodities over long distances. Truck carriers, on the other hand, are constrained by volume and weight to the truckload (TL) and less-than-truckload (LTL) markets. However, railroads are constrained by fixed right-of-ways and thus have different degrees of service flexibility. For example, door-to-door service can only be provided when both the shipper and receiver have rail sidings.

Rail transportation providers have high fixed cost and relatively low variable cost. The major cost elements in the railroad industry are the operation, maintenance, and ownership of right-of-way. Initially, a large capital investment is required and annual maintenance costs are incurred. Capital expenditures in 1996 alone amounted to $6.1 billion [1]. In addition, the investment for equipment in rail transportation, principally for locomotives and various types of rolling stock, also adds to the level of fixed cost. As a result, railroad companies tend to limit excess rail trackage and strive to use the limited resources efficiently. To remain profitable, railroad companies concentrate on improved service and efficient capacity utilization.

As introduced in Chapter 2, railroads typically develop plans for a tactical planning horizon, perhaps three to six months, to satisfy various operating constraints and business rules and use the resources efficiently. For instance, railroad companies create *blocking plans* to group shipments together and prevent the delay time and operational cost of loading, unloading and reclassifying. In addition, train schedules are also an important part of railroad tactical plans. Our research focuses on the time and capacity constraint routing (TCCR) problem, which uses existing tactical plans and develops efficient trip plans for each day. The objective is to minimize total transportation costs while ensuring train capacities are not exceeded and due dates of shipments are met.

1.2. Trucking Industry

During the late 1960s, trucks replaced railroads as the dominant form of freight transport in the United States. Generally, trucks compete with air for small shipments and rail for large shipments. The inherent advantages of trucking is the door-to-door service between almost any origin-destination combination such that no intermediate loading or unloading is required, which saves time and labor and also reduces the possibility of damage. Moreover, the frequency and availability of service also compete with other transportation modes.

The amount of freight transported by trucks has steadily increased over the recent years. Truck represents a vital part of most firms' logistics network, because the characteristics of the trucking industry are more compatible with the service requirements of the customers. While providing fast and efficient service at rates between those offered by rail and air, trucking industry has continued to prosper relative to other transport modes.

In contrast to railroads, the trucking industry has lower fixed cost since they do not own the roadways over which they operate. The tractor-trailer represents a relatively small economic unit, and terminal operations do not require expensive equipment. Therefore, a trucking company is able to increase or decrease the number of vehicles used in short periods of time and in small increments of capacity. On the other hand, the variable costs for trucks tend to be relatively high due to fuel costs. Also, highway construction and maintenance costs are charged to the users in the form of fuel taxes, tolls, and weight-mile taxes. Approximately, 70 to 90 percent of a motor carriers' cost is variable, and 10 to 30 percent is fixed [29].

For truck transportation providers, one of the challenges is to determine a fleet of suitable size for a particular area. Fleet planning is generally considered a strategic level decision due to the expense incurred and the expected life of a fleet. Strategic fleet decisions involve considerable capital investment with significant financial implications. An oversized fleet increases fixed costs associated with purchase, maintenance, fleet management, and related infrastructure (garages, parking lots, technical back-up facilities). On the other hand, an undersized fleet increases the possibility of unsatisfied demands and inefficient deliveries. The trailer fleet planning (TFP) problem that arises in an industrial gas distribution system is discussed in Chapter 3 and Chapter 4. In this problem, both the long-term fleet acquisition decisions and the medium-term vehicle relocation and rental decisions are considered so that total distribution and fleet management costs are minimized.

1.3. Contributions

The dissertation research is motivated by important distribution planning issues in rail and truck industry. The main contribution of this research is developing decision support tools to improve the rail and truck distribution systems and reduce overall system costs. Specific to the problems we studied, the contributions are summarized as follows:

Railroad Planning:

1. The time and capacity constrained routing (TCCR) problem is explored to support railroad trip planning.
2. A Time-Space-Train-Block (TS-TB) network is developed, based on which the TCCR problem is formulated as a multicommodity flow problem with additional train capacity and due time constraints.
3. Two integer programming formulations are presented. The first formulation is a conventional arc-based network flow model with respect to the TS-TB network. Due to the large size of the TS-TB network, the second path-based formulation is more applicable since it contains fewer variables and constraints.

4. Two heuristics are proposed to solve the TCCR problem efficiently. The first algorithm schedules shipments sequentially so that each shipment is guaranteed to satisfy both the due date and train capacity, while the second algorithm tends to prevent the repetitive operations for each shipment by considering all shipments at one time and then bumping and rescheduling the ones which violate the train capacity.

5. Computational investigations of these algorithms are conducted using data provided by a major U.S. railroad. Both the algorithms are implemented in C++ and solve the industry size of data within a few minutes of computational time. Additionally, compared to the trip planning approach which is currently used in railroad, the proposed algorithms provide better performance in terms of all the key cost factors. We also conduct sensitivity analysis on the impact of distance weight factor and train capacity.

Trailer Fleet Planning:

1. The trailer fleet planning (TFP) problem is explored to improve industrial gas distribution system. Both homogeneous and heterogeneous trailer fleets are discussed.

2. Mixed integer programming models are developed to formulate the problem across multiple depots and multiple periods. Both the delivery routes decisions and fleet management decisions are considered.

3. Decomposition approaches are proposed to solve the problem in two phases, where Phase I focuses on the distribution cost and Phase II determines the trailer long-term purchase, medium-term relocation, and medium-term rental decisions. For heterogeneous TFP, Phase I applies a rolling scheme to consider the allocation in previous period and the frequency that one trailer type has been used.

4. Sweep-based heuristics are developed to generate candidate routes which allows refill product during a route.

5. Solution approaches are implemented in OPL CPLEX and C# and numerical studies are conducted by using actual dataset. The effectiveness and efficiency of the decompo-

sition approaches are evaluated by comparing with the integrated model. In addition, the impact of trip size and the effect of multiple trailer types are analyzed.

The reminder of the dissertation is organized as follows. In Chapter 2, we study the time and capacity constrained routing problem in railroad planning. Optimization models and effective solution approaches are developed. Computational studies illustrate the effectiveness of the approaches on an industry data set. Chapter 3 discusses the vital trailer fleet planning problem that arises in industrial gas distribution. To support effective decisions, both the strategic and operational issues are considered in addressing the trailer fleet planning problem. A mathematical model and a two-phase approach for homogeneous fleet are developed. Chapter 4 further explores the trailer fleet planning problem for heterogeneous fleet. A mathematical model and a modified two-phase approach are developed to determine the appropriate number and types of trailers with varying customer demand. Finally, Chapter 5 summarizes the research and outlines potential future work.

Chapter 2

Time and Capacity Constrained Routing Problem in Railroad Planning

2.1. Introduction

Railroads are a vital component of freight transportation systems and play an important role in the economy. In the U.S., railroads account for 43 percent of intercity freight, more than any other mode of transportation. According to a model of U.S. Department of Commerce, America's freight railroads generate nearly $265 billion in total economic activity each year including direct, indirect, and induced effects [7]. Rail systems provide shippers with cost-effective transportation, especially for heavy and bulky commodities. It is also fuel efficient and generates less air pollution than truck.

In recent years, freight railroads are experiencing rapid growth in demand for their services. Between 2000 and 2006, the ton-miles of increased by more than 21 percent [3] and long-term growth is expected to continue. In the U.S., the American Association of State

Highway and Transportation Officials (AASHTO) predicts that the demand for freight rail services will increase 69 percent based on tons and 84 percent based on ton-miles by 2035 [4]. Railroads face a capacity shortage, which requires improvement for both infrastructure and operations. The development of railroad infrastructure, however, requires substantial capital investment to purchase land, lay tracks, build bridges, provid terminals, etc. Therefore, enhancing operational efficiencies is becoming vital to effectively manage the limited resources.

For major U.S. railroads, a typical *shipment* comprises a set of individual cars that share a common origin and destination. This chapter primarily focuses on the time and capacity constrained routing (TCCR) problem which aims to create efficient trip plans for multiple shipments from their origins to their destinations by considering the due date as well as the train capacities.

In this chapter, we develop integer programming formulations of the TCCR problem and explore algorithms to solve it. The paper makes the following contributions.

1. A Time-Space-Train-Block (TS-TB) network is developed, based on which the TCCR problem is formulated as a multicommodity flow problem with additional side constraints.

2. Two integer programming formulations of the TCCR problem are presented. The first formulation is a conventional arc-based network flow model with respect to the TS-TB network. Due to the large size of the TS-TB network, a second path-based formulation is developed with fewer variables and constraints.

3. Two heuristics are proposed to solve the TCCR problem efficiently. The first algorithm schedules shipments sequentially so that each shipment is guaranteed to satisfy both the due date and train capacity, while the second algorithm tends to prevent the repetitive operations for each shipment by considering all shipments at one time and then bumping and rescheduling the ones which violate the train capacity.

4. Computational investigations of these algorithms are conducted using data provided by a major U.S. railroad. Both the sequential and bump-shipment algorithms are implemented in C++ and solve the industry size of data within a few minutes of computational time. Additionally, compared to the trip planning approach which is currently used in railroad, both the algorithms provide better performance in terms of all the key cost factors. We also conduct sensitivity analysis on the impact of distance weight factor and train capacity.

The remainder of the chapter is organized as follows. We first provide an overview of four related railroad problems in Section 2.2 and then describe the inputs, outputs, objective, and constraints of the TCCR problem in details in Section 2.3. In Section 2.4, two integer programming formulations are presented. Section 2.5 proposes two heuristic algorithms to solve the TCCR problem, and numerical results are summarized in Section 2.6. Section 2.7 provides the conclusion.

2.2. Background

To transport millions of shipments annually, railroad companies typically develop plans for a medium-term (tactical) planning horizon, perhaps three to six months, to satisfy various operating constraints and business rules as well as use resources efficiently [46]. Below we introduce three tactical plans that are important in railroads: blocking plan, train schedule, and block-to-train assignment. The trip planning problem is described based on the tactical environment.

2.2.1 The Blocking Problem

A shipment often passes through many classification yards where the incoming shipment is reclassified (sorted and grouped) for placement on outgoing trains. These reclassification operations delay the movement of shipments and incur additional operational costs. To reduce the impact of reclassification, railroads group several shipments together to form a *block*. Each block is associated with an origin-destination pair that may or may not be the origin or destination of the individual shipments contained in group. Once placed into a block, the shipments are not reclassified until they reach the destination of the block.

> ***The blocking problem*** *determines how to aggregate a large number of shipments into blocks as they travel from origins to destinations.*

The objective of the blocking problem is to minimize total transportation cost, which includes the total miles that shipments travel and the costs of classifications. One of the first models for blocking problem, attributed to Bodin et al. [20], is formulated as a nonlinear mixed integer programming problem to simultaneously determine the optimal blocking strategies for all the classification yards in a railroad system. More practically, Ahuja et al. [8] solve the real-life railroad blocking problem using an advanced technique known as very large-scale neighborhood (VLSN) search. The proposed algorithm solves the problem to near optimality in one to two hours of computational time on a standard workstation computer.

2.2.2 The Train Scheduling Problem

Another important operating plan for a railroad company is the train schedule.

> ***The train scheduling problem*** *determines the origin, destination and route of each train, the arrival and departure times for each station at which the train stops, and the weekly operating schedule for each train.*

The train scheduling problem (also referred as *the train timetable problem*) seeks to minimize the total system-wide transportation cost while satisfying various practical constraints. Since the train scheduling problem is a large-scale and multi-objective optimization problem, numerous mathematical approaches and heuristic algorithms have been developed (Brännlund et al. [23], Ghoseiri et al. [40], Ghoseiri and Morshedsolouk [39] and Caprara et al. [25]). More recently, Mu and Dessouky [55] develop optimization-based approaches for scheduling freight trains. Two mathematical formulations of the scheduling problem are first introduced for moderate size rail networks. One assumes the path of each train is given and the other one relaxes this assumption. For large networks, two decomposition-based methods are developed and are outperform other existing algorithms. Dessouky et al. [31] consider trains operating in densely populated metropolitan areas where double or triple track lines may be used in some portions of the rail network. A branch-and-bound procedure is developed to determine the optimal dispatching times for trains traveling.

2.2.3 The Block-to-Train Assignment Problem

The blocking plan is generally created independently of train schedule, so the following problem determines the train route for each block.

> ***The block-to-train assignment (BTA) problem*** *determines which trains carry which blocks.*

The *BTA problem* is proposed by Jha et al. [45] with the goal of transporting the blocks at a minimum cost and with minimum delay. The problem considers both the requirement of the minimum cost train path for each shipment from its origin to its destination and the capacity of the trains. Hence, one block may be assigned to multiple train paths, resulting in different *train-blocks*. A train-block indicates how a block have been assigned to a train path. Jha et al. [45] propose an effective IP formulation and an exact algorithm to solve it by CPLEX. Additionally, they present heuristic algorithms including a Lagrangian relaxation-based method as well as a greedy construction to solve the BTA problem more efficiently. Nozick and Morlok [56] study a intermodal model of minimizing the cost of delivering the required shipments given a fixed train schedule, and satisfying due dates. They develop a heuristic by solving the linear programming relaxation iteratively and rounding some of the resulting fractional values until a feasible integral solution is found. Kwon et al. [48] identify several ways to improve the given blocking plan. They formulate the car-to-block and block-to-train assignment as a linear multi-commodity flow, which is solved using column generation technique.

2.2.4 The Trip Planning Problem

The blocking problem, the train scheduling problem, and the block-to-train assignment problem address important tactical decisions based on historical shipment data. The trip planning problem is motivated by the following considerations: (1) since the shipments vary weekly and daily, it is impractical for railroad planners to create blocking plans, train schedules and BTAs frequently; (2) typically the train-blocks are generated in the BTA problem by assuming that all trains run every day, and thus only consider *general train* information. However, trains may have different frequencies. For instance, one train may run every day of the week, while another may run only from Monday to Friday. Therefore, decisions are required for the train on a particular working day, which is referred as a *train-run*. Table 2.1 summarizes the definitions of the important terminology used in this paper.

> *Given the blocking plans, train schedules, and block-to-train assignments,* ***the trip planning problem*** *determines the assignment of shipments to train-blocks and train-blocks to particular train-runs so that various operational constraints are satisfied and the total cost of transportation is minimized.*

Two main variants of the trip planning problem include the static trip planning problem and dynamic trip planning problem. For the *static trip planning problem*, each shipment is fixed to a sequence of blocks. That is, shipment-block assignment is given. If each of the given blocks only contains one train-block, then we only need to determine the train-run path. If there are multiple train-blocks for a given block, we should determine both the train-blocking path and the train-run path. For the *dynamic trip planning problem*, the pre-

Table 2.1: Terminology and definition

Terminology	Definition in this context
Block	A group of shipments moving together in the network
Train-block	A group of shipments moving along a specific train path
Train-run	An individual train operating on a specific day
Train-segment	A segment of train route between two consecutive stations.
Trip	An instance of train-segment for a specific day

determined shipment-block assignment is relaxed. Shipments can be assigned to any block sequence and train route as long as the operational constraints are satisfied. Note that the dynamic trip planning problem is a generalized form of the static trip planning problem. Therefore, in this paper, we primarily discuss the dynamic trip planning problem.

We are aware of only a few papers that consider either the static or dynamic trip planning problem indirectly in their problem formulation. Barnhart et al. [18] consider an origin-destination integer multi-commodity flow problem to decide an optimal path from origin to destination for each commodity. As part of a branch-and-bound approach, a new branching rule is devised to generate columns efficiently at each node. Keaton [46] considers the train routing problem to determine optimal train connections, frequencies, and routing plans for freight cars. The objective is to minimize train cost, car time cost, and yard classification cost, subject to limits on train size, number of blocks formed by yard, and maximum origin-to-destination trip times. A heuristic method is proposed based on Lagrangian relaxation for this problem. Ahuja et al. [10] study dynamic shortest path problems that determine the shortest path in a dynamic network. The arc travel times change dynamically by time. A pseudo-polynomial time algorithm is developed to minimize the total travel time. Avella et al. [13] formulate a Intelligent Tourist Problem (ITP) which consists of scheduling the visit of a tourist on a given network in order to maximize the satisfaction degree while respecting time windows restrictions. The problem is formulated as a set packing problem with side constraints. Due to the huge number of variables, the LP-relaxation is solved using a column-and-row generation approach.

2.3. TCCR Problem Description

In this paper, we study a dynamic trip planning problem, referred as the *time and capacity constrained routing (TCCR) problem*. The problem is to determine the shortest cost train-blocking paths and train-run paths for a set of shipments by satisfying their due dates and honoring train capacities. Formally, we introduce the input data, decision variables, constraints and objectives in the following.

Input:

- Blocks: the origin-destination of each block;
- Train schedule: train route, weekly operation schedule[1] , its arrival and departure times at various stations, and train capacity[2];
- Block-to-train assignment: all permissible train paths for each block, the reclassification time required to form a block, and the time to transfer a block from one train to another;
- Shipments: shipment origin-destination, number of cars needed, release time and due date.

Decision variables:

- Train-blocking path of each shipment;
- Train-run path of each shipment, including the trains, the origin and destination stations, and the departure and arrival time at each station.

[1]Weekly operation schedule includes the frequency of operations for a train and on which days does the train operate. For instance, 'NYYYYYN' indicates that the train operates from Monday to Friday in a week.

[2]Each train has a specified capacity in terms of the maximum number of cars that can be carried

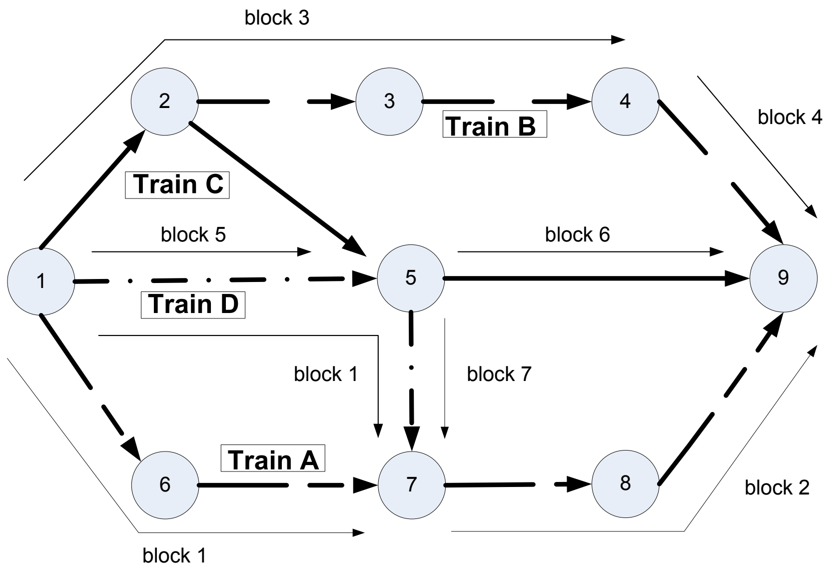

Figure 2.1: Trip planning problem illustration

<u>**Constraints:**</u>

- Use specified blocks and block-to-train assignments;
- Meet due date of each shipment;
- Train capacities should not be exceeded.

<u>**Objective Function:**</u>

- Minimize the following weighted cost terms:
  - Train-mile costs: the total train distance traveled multiplied by a cost per distance;
  - Station costs: (i) Reclassification costs, which occur when moving a shipment from one block to another block; (ii) Block swap costs, which occur when swapping a block from one train to another train.

Figure 2.1 illustrates a simple example of TCCR problem. Four trains (Trains A, B, C and D, represented by different line styles) travel in the network that includes nine stations. Each train has its own route, schedule, and weekly frequency. Additionally, the problem includes seven blocks with their block-to-train assignments. For instance, block 3 takes

Train C from station 1 to 2 and changes to Train B to travel from station 2 to 4. Thus a block swap cost is incurred at station 2 and a reclassification cost is incurred at station 4 when a shipment is moved from block 3 to block 4. When a shipment goes from station 1 to 9, it can either take block 1 and block 2, block 3 and block 4, or block 5 and block 6. In this example, block 1 is assigned to two different train paths: Train D over the path 1-5-7 or Train A over the path 1-6-7. Thus, we further distinguish block 1 as two separate train-blocks, which will be shown in the following section. The TCCR problem aims to find the train-blocking path for each shipment and the train-run path as well.

2.4. Mathematical Formulations

In this section, we first describe the network representations over which the TCCR problem can be formulated as a network flow problem. Based on the appropriate network, two different IP formulations are developed: a conventional arc-based formulation and a path-based formulation with fewer constraints and variables.

2.4.1 Network Representations

The TCCR problem essentially is a network flow problem. The following subsection describes two important network representations: a *train-block network* which contains the general train and block information and a *time-space-train-block* network which adds the detailed block-to-train assignment and train time schedule information.

Train-Block Network

The train-block network illustrates the blocks and their train assignment. Each arc in the network, which is referred to as a *train-block arc*, represents a block carried by certain train path. Regardless of the detailed train route, the train-block network simply replicates each

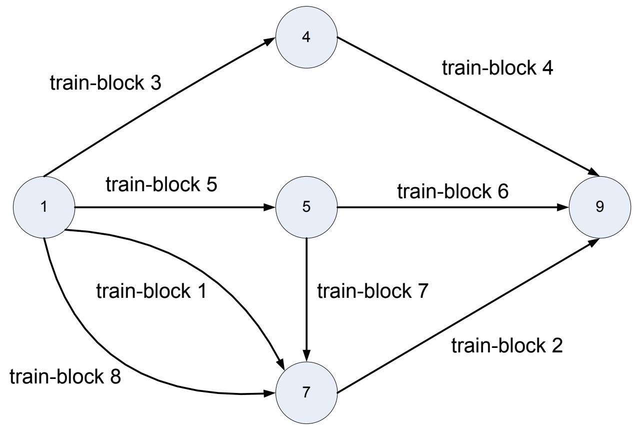

Figure 2.2: Train-block network illustration

block for each train path assigned to that block. For example in figure 2.1, block 1 from station 1 to 7 can be carried on two different train paths, therefore they are distinguished as train-block 1 and train-block 8 in the train-block diagram shown in Figure 2.2. The train-block network is relatively simple and contains an aggregated cost structure. That is, each train-block arc is associated with a fixed transport cost, which includes the train distance costs and block swap costs on the associated train path, and the reclassification cost at the origin station.

Time-Space-Train-Block Network

The train-block network includes the main cost components, but it does not specify either the detailed train routes or the train schedules. To accommodate due dates, the time-space-train-block (TS-TB) network incorporates the information about time, place, train and block. We assume that there are multiple train runs in a week and trains have different weekly frequencies. On their working days, all train runs follow the same departure and arrival schedule at each station.

Our TS-TB network is similar to the space-time network developed by Ahuja et al. [9] to formulate the flow of locomotives in a train network, with additional information on blocks. The following important components are used to construct the TS-TB network:

> *Train-block-trip (TB-trip)*: The itinerary of a train run between two consecutive valid-stops is referred to as a *trip*. There may be multiple train-blocks carried on one trip, and we specify the trip associated with each train-block. In the TS-TB network, we use *TB-trip arc* (l', l'') to represent a trip carrying a certain train-block, where l' denotes a *train-departure node* and l'' denotes a *train-arrival node*;

> *Ground nodes*: For each train-arrival node, corresponding *ground-arrival nodes* are created with the same station, time and train attributes. Likewise, for each train-departure node, corresponding *ground-departure nodes* are created

> *Ground arcs*: Each train-arrival node is connected to the associated ground-arrival node with a directed arc, referred to as a *arrival-connection arc*. Likewise, each ground-departure node is connected to the associated train-departure node with a directed arc, referred to as a *departure-connection arc*

> *Ground loop*: to capture the movement among blocks and trains, the ground nodes at each station are sorted in chronological order by their time attributes. Each ground node is connected to the next ground node in this order through a directed *ground arc*. Furthermore, to allow a block to take a train in the next week, the latest ground node is connected to the earliest ground node at the same station. Therefore, ground arcs form a loop in each station.

The TS-TB network is constructed based on the train-block network by presenting a single train-block arc in detail with the intermediate stations, train runs, and ground connections. Figure 2.3 provides the TS-TB network corresponding to the example in Figure 2.1. Suppose that in each week we have two runs for Train A, two runs for Train B, three runs for Train C, and two runs for Train D. The labeled nodes are train-departure (-arrival) nodes

Table 2.2: Attributes of the arcs of TS-TB network

Arc Type	**Attributes**
TB-trip arcs	train distance, traversal time, capacity
Ground arcs	waiting time
Arrival-connection arcs	none
Departure-connection arcs	If the arc connects two different trains for the same block, a block swap cost incurs. If the arc connects two different blocks, there should be an reclassification cost and cutoff time to setup a new block.

corresponding to different stations and the small nodes are the ground-departure (-arrivel) nodes. The thicker lines are TB-trip arcs representing trips with different train-blocks and the thinner lines are ground connection arcs. At each station, there is a ground loop.

Each arc in the TS-TB network has attributes which are summarized in Table 2.2. The TB-trip arcs, the main component in the network, have the following attributes: train distances, traversal times, and capacities. Ground arcs have associated waiting times. Departure-connection arcs contains the station costs. Since all TB-trips are defined by both train-blocks and train-runs, we can easily tell which station cost applies in transit. That is, block swap costs are incurred when the connection occurs between two different trains with the same train-block, while reclassification costs are incurred when the connection occurs between two different train-blocks. Both the block swap cost and reclassification cost are node-based (independent of blocks and trains).

2.4.2 Arc-based IP Formulation

When using the TS-TB network for TCCR problem, an *origin node* and a *destination node* are added for each shipment. The origin node is connected to the earliest ground-departure

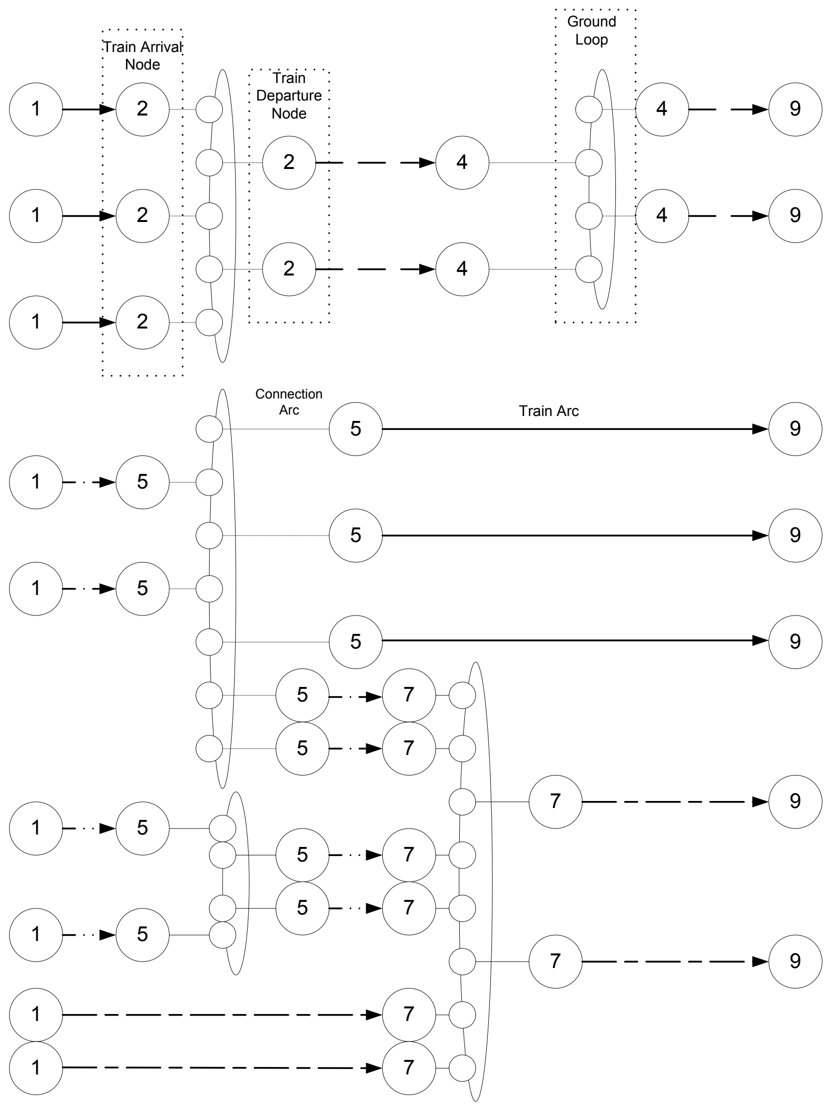

Figure 2.3: TS-TB network illustration

node after the release time at the shipment's origin station. The destination node is connected to all the ground-arrival nodes at the shipment's destination station. We define the TCCR problem as finding a path for each shipment from its origin node to the destination node in the TS-TB network, such that the overall cost is minimized and both the due time and train capacity are satisfied.

Note each arc in the TS-TB network includes information about both trips and train-blocks, hence we define T to be the set of trips and the associated ground connections and define B as the set of train-blocks. As a result, each arc in the TS-TB network is represented by the combination of (t, b), which indicates the particular trip t carrying train-block b.

With the TS-TB network as a basis, the TCCR problem is formulated as an arc-based integer programming problem (TCCR-ab) that determines whether an arc is selected in the optimal path for a shipment. The objective is to find a connected path in the TS-TB network for each shipment from its origin-node to the destination-node while considering all constraints. The following notation is used in our formulation.

Parameters of network:

- G TS-TB network with origin-node set O and destination-node set D;
- N Set of nodes in G, where nodes are denoted by i;
- T Set of trips and the corresponding ground connections, where trips and the ground connections are denoted by t;
- B Set of train-blocks, where train-blocks are denoted by b;
- A Set of arcs in G, where arcs are denoted by (t, b);
- A_i^+ Set of outgoing arcs at node $i \in N$;
- A_i^- Set of incoming arcs at node $i \in N$;
- $\tau_{(t,b)}$ Traversal/waiting time on arc (t, b) for TB-trip arcs and ground arcs; otherwise, $\tau_{(t,b)}$ is zero;

$c_{(t,b)}$ Cost on arc (t, b), which can be a reclassification cost on departure-connection arcs, block swap cost on departure-connection arcs, train distance cost on TB-trip arcs, or zero on other arcs;

q_t Capacity on trip or ground connection t (TC_t is infinity for ground connections).

Parameters of shipments:

W Set of shipments, where shipments are denoted by w;

o_w Origin node of shipment w;

d_w Destination node of shipment w;

v_w Number of cars needed for shipment w;

r_w Release time of shipment w;

s_w Due time of shipment w.

Decision variables:

$$x^w_{(t,b)} = \begin{cases} 1 & \text{shipment } w \text{ travels on arc } (t,b) \in A \\ 0 & \text{else} \end{cases}$$

In the TCCR-ab formulation, objective function (2.1) minimizes the cost of selected arcs. Constraint (2.2) ensures that each shipment follows a connected path from its origin node to its destination node. Constraint (2.3) ensures that the number of cars flow on each trip is no more than the allowed capacity. The capacity limits only apply for the trips while the capacities of other ground connections are set to infinity. Note that a trip may carry multiple train-blocks, thus we add the shipments on various train-blocks to determine the total used capacity. Constraint (2.4) ensures that the total travel time for each shipment is less than the duration between due date and available time.

Formulation (TCCR-ab):

$$\text{Minimize} \quad \sum_{w \in W} \sum_{(t,b) \in A} c_{(t,b)} x^w_{(t,b)} \tag{2.1}$$

Subject to

$$\sum_{(t,b) \in A_i^+} x^w_{(t,b)} - \sum_{(t,b) \in A_i^-} x^w_{(t,b)} = \begin{cases} 1 & i = o_w \\ 0 & i \neq o_w \text{ or } d_w \\ -1 & i = d_w \end{cases} \quad \forall w \in W \tag{2.2}$$

$$\sum_{w \in W} \sum_{b \in B} v_w x^w_{(t,b)} \leq q_t \qquad \forall t \in T \tag{2.3}$$

$$\sum_{(t,b) \in A} \tau_{(t,b)} x^w_{(t,b)} \leq s_w - r_w \qquad \forall w \in W \tag{2.4}$$

To estimate the size of the problem in formulation TCCR-ab for a U.S. railroad, suppose a network has 3,000 train-blocks, each train-block is transported by 2 trains on average, and the average frequency for each train is 4 in a week. In this case, there will be 24,000 ($3,000 \times 2 \times 4$) TB-trip arcs available for each shipment. A rail company transports many shipments, say 500,000, resulting in $24,000 \times 500,000$ decision variables and more than a million constrains. This magnitude is formidable to solve with commercially available optimization software. So another IP formulation is described based on paths rather than arcs.

2.4.3 Path-based IP Formulation

This path-based formulation is similar to the formulation developed by Jha [45]. For each shipment, an enumeration of *feasible paths* with respect to the TS-TB network is required as input and the model selects the optimal path.

This formulation often reduces the number of decision variables, since only a set of "possible" paths is considered. This approach is consistent with current railroad practice. Railroads do not allow some shipments to be sent along arbitrary trains or blocks. For instance, high-wide shipments cannot go on bridges, and hazardous materials are not routed through densely populated cities. Instead, some shipments are restricted to a set of paths which are manageable. Another advantage of this formulation is that we can eliminate the paths that do not meet operational constraints by excluding them in the feasible path set. Particularly, when a candidate path is given, the traversal time and arrival time can be determined accordingly. As a result, all the paths in the feasible set can be limited to those that satisfy due dates and only the train capacity constraints need to be ensured. Below we introduce additional notation to formalize the path-based formulation (TCCR-pb).

Parameters of network:

P^w Set of feasible paths in G for shipment $w \in W$;

$P^w_{(t,b)}$ Set of paths in P_w which contain arc (t, b);

c_p Total cost of traveling on a path $p \in P$; Here, the cost should the sum of all the reclassification costs, block swap costs, and train distance costs.

Decision variables:

$$x_p^w = \begin{cases} 1 & \text{shipment } w \text{ travels through path } p \\ 0 & \text{else} \end{cases}$$

In the TCCR-pb formulation, objective function (2.5) minimizes the cost of selected paths. Constraint (2.6) ensures that exactly one path is assigned to each shipment, while Constraint (2.7) ensures that the train capacity is not violated. For each trip, we sum over all shipments, paths, and train-blocks to evaluate the capacity used. The due date constraints are not needed since only paths that meet the due date requirements are included in P^w.

Formulation (TCCR-pb):

$$\text{Minimize} \quad \sum_{w \in W} \sum_{p \in P^w} c_p x_p^w \tag{2.5}$$

Subject to

$$\sum_{p \in P^w} x_p^w = 1 \qquad \forall w \in W \tag{2.6}$$

$$\sum_{w \in W} \sum_{b \in B} \sum_{p \in P^w_{(t,b)}} v_w x_p^w \leq q_t \qquad \forall t \in T \tag{2.7}$$

Consider the example with 24,000 TB-trip arcs and 500,000 shipments. The path-based formulation reduces the problem size in two ways: (1) the number of decision variables is significantly reduced, since it only depends on the size of feasible path set. Suppose there are 100 candidate paths for each shipment, then we only have $100 \times 500,000$ decision variables (rather than $24,000 \times 500,000$ in TCCR-ab); and (2) Since only the paths satisfying the due date are allowed in the feasible set, the due date constraints are no longer needed, which reduces the size by 500,000 constraints.

2.5. Solution Approaches

In order for the path-based formulation to be of a reasonable size, it is critical to restrict the size of the feasible path set. Jha et al. [45] propose several algorithms to enumerate the paths under various restrictions and solve the model optimally. Their computational results show that "the computational effort to reach the optimality can be considered reasonable only if the problem was solved once with a given data set." In railroad practice, however, trip planning decisions are frequently analyzed using different data sets. Thus, this section presents two heuristic algorithms which produce a high-quality solution in a relatively short time.

2.5.1 Sequential Algorithm

The first algorithm treats shipments sequentially. That is, the shipments are sent one by one along their optimal train-run path while satisfying due dates and not violating the train capacities (in terms of the number of cars). Generally, three steps are followed for *each shipment*.

Step 1: Find *the (possible) shortest train-blocking path*. When time and capacity constraints are not taken into account, we only focus on the simpler train-block network (example in Figure 2.2). Each arc (i, j) in the network is associated with a fixed train-blocking cost c_{ij} which includes the reclassification costs, block swap costs and train distance costs on a specific train route.

Step 2: Find *the (possible) fastest train-run path*. Although the train-blocking path produced in Step 1 provides the minimum cost solution, there is no guarantee for feasibility as we also have additional capacity and due date constraints. Therefore, based on the obtained train-blocking path, we check the BTA and train schedule so as to send the shipment with the earliest possible train-run (as long as the remaining capacity is allowed) at each station. Then check the arrival time of this path. If the due time is not met, go back to step 1; otherwise go to step 3.

Step 3: After sending each shipment, the trip capacities are updated accordingly.

It is noteworthy that even the fastest train-run path produced in step 2 still may not satisfy the due date. In that case, we have to go back to step 1 and find the *second* shortest train-blocking path (or 3^{rd}, 4^{th}, $\cdots$). This problem needs a list of shortest paths, which is different than the traditional shortest path problem. In the following, we describe *Yen's Algorithm*, which is able to find the K shortest paths connecting a given origin-destination pair in the digraph with minimum total cost. Furthermore, the algorithm to find the fastest train-run path is investigated.

Yen's Algorithm: Finding the K Shortest Paths

In graph theory, the shortest path problem is a well-solved problem of finding a path between two nodes such that the sum of the weights of the constituent arcs is minimized. The shortest path problem can be solved using Dijkstra's algorithm in polynomial time [32]. The K shortest paths problem is a classical and long-studied generalization of the shortest path problem. Yen [68] first proposes the algorithm to find the K shortest loopless paths. Later, Martins and Pascoal [54] shows it can be implemented in $O\left(Kn\left(m + n\log n\right)\right)$, where m is the number of arcs and n is the number of nodes in the network. Two important definitions are used in Yen's algorithm [68]:

P^k: the kth shortest path, $k = 1, 2, \cdots, K$; Particularly, $P^k = (1)-(2^k)-\cdots-(Q_k^k)-(N)$, where (1) is the origin, (N) is the destination, Q_k is the length of the kth shortest path and $(2^k), \cdots, (Q_k^k)$ are respectively the $2nd, \cdots, Q_k$ th node in kth shortest path;

P_i^k: the shortest *deviation path* from P^{k-1} at ith node, $i = 1, 2, \ldots, Q_k$; Particularly, $P_i^k = (1) - (2^k) - \cdots - (i^k) - ((i+1)^k) - \cdots - (Q_k^k) - (N)$, where the subpath $(1)-(2^k)-\cdots-(i^k)$ coincides with P^{k-1} and then deviates to a node that is different from the $(i+1)$th nodes of all the $P^j, j = 1, 2, \cdots, k-1$ that have the same path from 1st to the ith nodes.

Essentially, Yen's algorithm is developed from the fact that P^k is a deviation from P^j, $j = 1, 2, \cdots, k-1$. Therefore, given all $k-1$ shortest paths, we just need to find all of their shortest deviations and pick the one with shortest length as the kth shortest path. Algorithm 1 finds the kth shortest path P^k then $P^1, \cdots, P^{k-1}$ are given. We use List A to store K shortest paths and List B to store all candidates for the kth shortest path.

Initially, List A contains $k-1$ shortest paths while List B contains all the deviations of P^j, $j = 1, 2, \cdots, k-2$.

Algorithm 1 Find the kth Shortest Path

```
for i = 1, 2, ···, Q^{k-1} do
  for j = 1, 2, ···, k-1 do
    if the first i nodes of P^{k-1} coincide with the first i nodes of P^j then
      Set c_{iq} = M where q is the (i+1)th node in P^j
    end if
  end for(prevent recalculating P^j, j = 1, 2, ···, k-1)
  Find the shortest path from (i) to (N)
  Construct P_i^k by joining (1) to (i) in P^{k-1} and shortest path from (i) to (N) just found
  Add P_i^k into List B
end for
Select the path that has the minimum length in List B, denoted as P^k, and move it from
List B to A
(Remark: List A and B are not cleared until finding all K shortest paths.)
```

Finding the Fastest Train-run path

After the train-blocking path for each shipment has been determined, the fastest train-run path is determined. The solution for this problem is more straightforward: we check the BTA and train schedule, and determine the earliest day when the train-run has enough capacity remaining. Figure 2.4 illustrates the procedure to find the fastest train-run path. Suppose the *train-block sequence* is provided as train-block 3 to train-block 4 with the BTA and train schedules given in Figure 2.1. By checking BTA, we can build a *BTA sequence* which demonstrates the sequence of train-segments carrying the shipment. By considering the train schedule, we can determine a *train sequence* which includes the frequencies and time schedule of each train-run.

For the train sequence, we check if the earliest trip has enough remaining capacity. If so, the shipment is added to this trip. Otherwise, the shipment waits for the next available train-run. After the fastest train-run path is completed, the total traversal time is checked to see if the due time for the shipment is met. If so, then this train-run path is output.

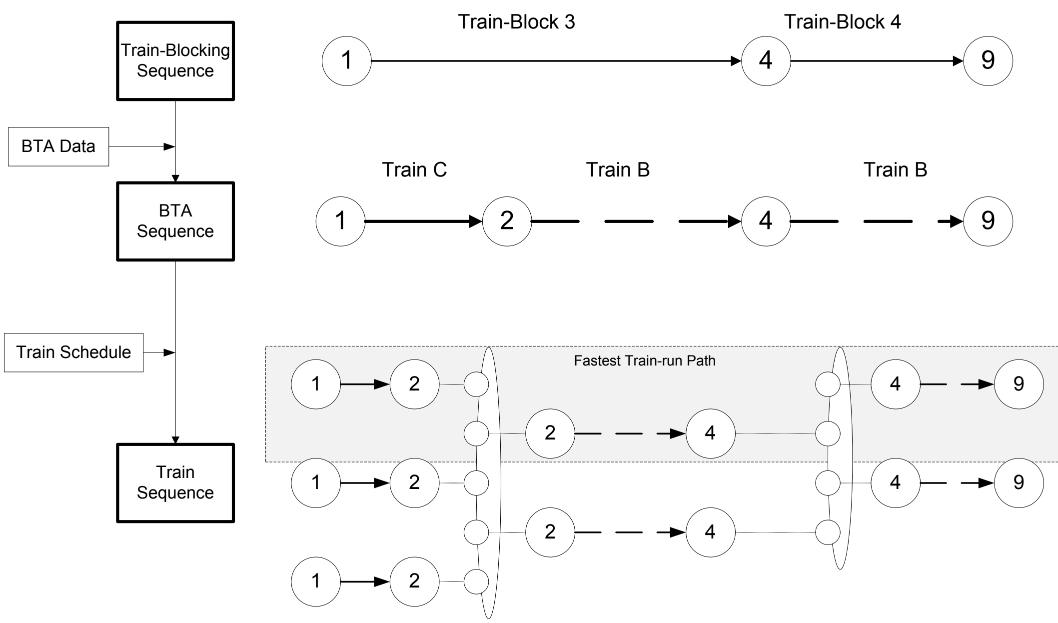

Figure 2.4: Procedures to find the fastest train-run path

Otherwise, the given train-blocking path is not feasible and we need an alternative train-blocking path. The formal approach is summarized in Algorithm 2, where DepartureNode tracks the nodes visited on the train-run path, TotalTime tracks the total used travel time, and the TrainRunPathList returns the obtained train-run path.

Summary of TCCR Sequential Algorithm

When applying the sequential algorithm, the first issue is to decide the order of sending shipments. Below are some criteria by which the shipments can be prioritized:

- Shipments with earlier due dates;
- Shipments which are released earlier;
- Shipments which require more cars.

In this research, we follow the last criterion.

Algorithm 2 Find the fastest train-run path

```
Read given train-blocking path, build train-blocking sequence
Read BTA data, build BTA sequence
Read train schedule, build train sequence
Add origin-node into TrainRunPathList
DepartureNode ← origin-node of the shipment
TotalTime ← 0
while DepartureNode != destination-node of the shipment do
  if The trip capacity from DepartureNode is enough then
    DepartureNode ← Arrival Node of the trip
    TotalTime ← TotalTime + CutoffTime + TransitTime
  else
    DepartureNode ← DepartureNode at next train run
    TotalTime ← TotalTime + WaitingTime
  end if
  Add DepartureNode into TrainRunPathList
end while
if TotalTime ≤ DueTime - AvailableTime of the shipment then
  return TrainRunPathList
else
  return NULL
end if
```

Now we present our sequential algorithm for TCCR problem formally in Algorithm 3. Let N denote the number of shipments (sorted) and k denote the index of the shortest path. Note that k increases until the shipment finds a feasible train-run path. This is not practical for implementation since computational time will increase if the due date is very tight or the remaining capacity is very limited. Therefore, generally we set an upper bound, K, for the number of k, say $k \leq K = 3$. If feasible train-run path is not found after exploring K train-blocking paths, then this shipment is marked as unscheduled and is scheduled manually later.

Algorithm 3 TCCR Sequential Algorithm

```
for n = 1, 2, ··· , N do
  Set k ← 1
  while k ≤ K do
    Determine the kth shortest train-blocking path for shipment n in train-block network.
    Based on that obtained train-blocking path, determine the fastest train-run path
    if The fastest train-run path exists (feasible) then
      Record the train-run path for shipment n and update the capacities on related trips
    else
      Set k ← k + 1
    end if
  end while
end for
```

An important advantage of the sequential algorithm is that it is primarily based on the train-block network that is much smaller than the TS-TB network. Thus it is relatively easy to solve the shortest train-blocking paths. The main disadvantage of the sequential algorithm is the potential repetitions when the algorithm solves for the shipments one by one. A railroad company often has millions of shipments to schedule. Any unnecessary repetitive calculations increase computational time and effort. To overcome this deficiency, a modified approach is developed.

2.5.2 Bump-Shipment Algorithm

Since the sequential algorithm sends shipments one by one, the remaining capacity on each trip needs to be checked again and again. In addition, the trip capacities need to be updated after sending each shipment. To avoid these repetitive processes, an alternative heuristic is proposed by relaxing the capacity constraint and allowing all shipments to be sent "simultaneously" on the fastest train-run path. Then the trips where the capacity is exceeded are evaluated to determine potential shipments to remove (bump). Informally, we follow three main steps to present this algorithm.

Step 1: Find the (possible) shortest train-blocking path in the train-block network, and find the fastest train-run path for *all* shipments based on the obtained train-blocking paths. At this time, each shipment is just required to meet the due date.

Step 2: Identify the trip where the capacity constraint is violated and remove shipments until the trip capacity is satisfied. Denote the removed shipments as *bumped*.

Step 3: Resend the *bumped* shipments. Unlike Step 1, we do not allow the bumped shipments to take the trips where the capacities have been (or will be) violated in order to ensure the algorithm can terminate within finite iterations. Repeat Step 2 and Step 3 until all trips satisfy the capacity constraints.

The first step can be solved by a *simplified* sequential algorithm, where (1) we do not sort the shipments since every shipment has an equal opportunity to select the optimal routes; (2) each shipment is sent along the fastest train-run path without checking the remaining capacity on each trip; (3) there is no need to update the trip capacities. In step 2, the shipments are bumped according to priority from low to high. Note that there may be several trips of the same train-segment which exceed the capacity limit. Instead of analyzing each trip individually, we consider the bumped shipments from all of these trips together. That is, all shipments removed from the same train-segment are stored in one set and rescheduled later in step 3 by priority from high to low. The shipments are resent only through the trips which have enough space. This step is similar to the sequential algorithm. The only difference is that we have obtained an optimal train-blocking path in step 1, so we can simply update the fastest train-run path based on that train-blocking path while considering the capacity constraint. If the updated train-run path cannot meet the due time, then we go to find the next optimal train-blocking path. Still an upper bound for the number of candidate train-blocking paths is maintained, i.e. $k \leq K$. If any feasible train-run plan for a shipment is not found after exploring K train-blocking paths, then this shipment is recorded for further consideration. Figure 2.5 illustrates the bump-shipment algorithm in details.

Compared to the sequential algorithm, the bump-shipment approach addresses the shipments simultaneously in Step 1 thus preventing potential repetitions. However, Step 3 has the same complexity as the sequential algorithm. Hence in the case of limited train capacity, the bump-shipment algorithm may be not practical as more trip capacities may be violated and more shipments may need to be rescheduled.

2.6. Numerical Studies

Before presenting the numerical results of applying TCCR algorithms, recall that the static trip planning problem uses predetermined shipment-block assignments as input, while the TCCR problem (i.e. dynamic trip planning problem) has no fixed relationship between shipment and block selection. Thus, the static trip planning problem is a reduced problem of TCCR problem, the details of the static trip planning problem are summarized in the Appendix, which can be solved by the sequential algorithm with a reduced size network. This algorithm is referred to as a static trip planning approach.

In this section, we present the computational results of the proposed algorithms, which are implemented in C++ with the data provided by a major U.S. railroad for an actual problem. This railroad network has more than 3000 train stations, 7000 blocks, and 2000 trains with

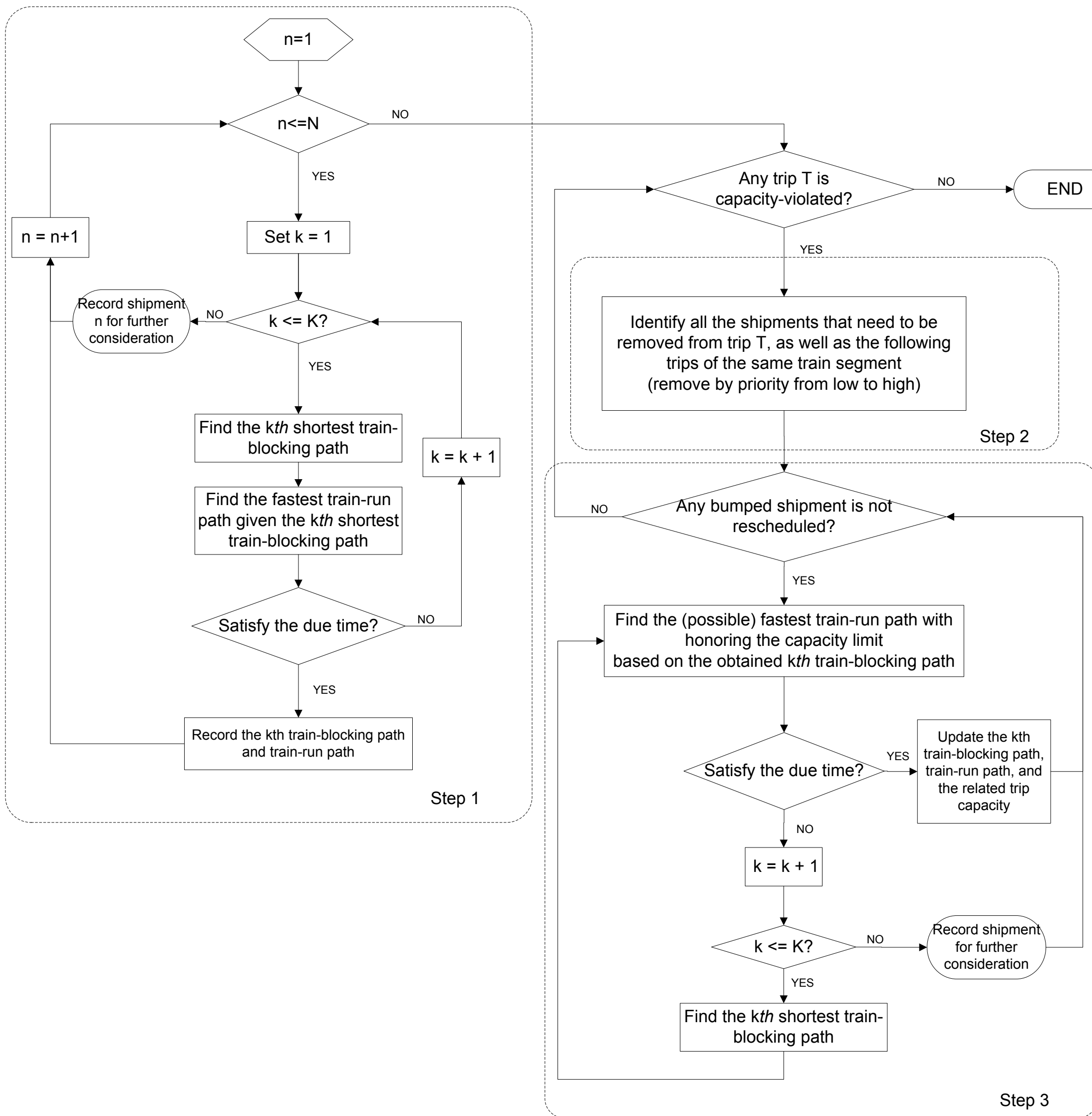

Figure 2.5: TCCR bump-shipment algorithm

Table 2.4: Running times for different trip planning problems

Trip Planning Problem	Static TCCR Problem	Dynamic TCCR Problem	
		Sequential	Bump-shipment
Running times	45 s	approx. 6 min	approx. 5 min

specified frequencies. The goal is to transport approximately 500,000 shipments through the railroad network. The shipments are routed using a static trip planning approach and both dynamic trip planning approaches (sequential and bump-shipment algorithms). All computational experiments are tested on a 2.93 GHz Pentium PC. Table 2.4 summarizes the computational times for different trip planning approaches. As shown, the static trip planning problem is solved very quickly by the sequential algorithm. The dynamic TCCR problem takes longer but is still solved within 6 minutes. As expected, the bump-shipment algorithm reduces computational time by preventing the repetitive operations. Since the trip planning problem is typically solved each week, computational times of 5 and 6 minutes are reasonable. Implementation issues are summarized in Appendix B.

2.6.1 Cost Saving

We compare the results obtained from the static trip planning approach and the dynamic trip planning approaches in Table 2.5.The static trip planning problem adopts the blocking paths similar to what the railroad company currently follows. Since it is possible to have shipments that cannot find a feasible train-run path, we first summarize the percentage of the unscheduled shipments. Furthermore, the average values of the key factors in railroad industry, weighted by the number of cars each shipment taken, are summarized, including travel and waiting times, travel miles, associated operational costs, and the capacity utilization of merchandize trains.

As shown in Table 2.5, the dynamic trip planning approaches have fewer unscheduled shipments. With a static trip plan, the blocking path is fixed and shipments may not be transported when a vital trip has reached the maximum capacity limit. With the static trip plans, unassigned shipments account for 4.24% of the data set. In contrast, the dynamic trip planning approaches provide additional capability to determine the train-blocking path. Moveover, more options are found to send shipments by considering K shortest train-blocking paths. Since we limit the number of the shortest train-blocking paths to no more than three, there are still shipments that cannot find feasible train-run paths. But the percentages of unassigned shipments are reduced to 2.23% for the sequential algorithm and 2.16% for the bump-shipment algorithm. Both the sequential and bump-shipment algorithms generate high quality solutions that are better than the current railroad practice. Regarding to the total transportation cost, the dynamic trip plans save approximately $7 (6%) for each railcar while also performing well for other cost factors. When we compare the total transportation costs from static trip planning approach and the sequential algorithm of dynamic trip planning, the total saving is very substantial ($1,885,870).

To further illustrate the benefits of dynamic trip plans, we focus on the results with dynamic plan generated by our dynamic trip planning approaches that are different than the results generated with the static planning approach. With the sequential algorithm, approximately 30% shipments are assigned to blocking paths different from their static plans (excluding shipments that cannot find feasible train-run paths through the static planning approach). Table 2.6 evaluates the average performances of these 30% shipments and shows the improvements of these distinct dynamic trip plans. As shown, the total transportation cost is saved by 12.88% and other cost factors are also substantially reduced. Similarly, Table 2.7 presents the improvement of results obtained by the bump-shipment algorithm which are different than the results from the static trip planning approach.

Table 2.5: Average improvements of dynamic trip plans

Statistics	Static Trip Plan	Dynamic Trip Plan		% Improvement	
		Sequential	Bump	Sequential	Bump
% of Unscheduled shipments	4.24%	2.23%	2.16%	2.01%	2.08%
Travel time[a](min)	5509.58	5339.12	5336.26	3.09%	3.15%
Waiting time[b](min)	3058.88	2916.09	2905.51	4.67%	5.01%
Travel distance (mile)	560.95	534.65	536.12	4.69%	4.43%
Block swap cost ($)	24.94	23.93	23.93	4.07%	4.07%
Reclassification cost ($)	38.95	35.15	35.34	9.76%	9.27%
MER capacity utilization[c](%)	52.83%	46.27%	45.10%	6.56%	7.73%
Transportation cost ($)	119.98	112.54	112.87	6.20%	5.92%

[a] Travel time is the total time between the release time at origin station and the arrival time at destination station.

[b] Waiting time includes all the time when the shipment is staying at stations.

[c] Capacity utilization of merchandize trains, where

$$\text{Capacity utilization} = \frac{\sum(\text{number of cars used} \times \text{trip length})}{\sum(\text{total capacity of the trip} \times \text{trip length})}$$

Table 2.6: Improvements of dynamic trip plans which are different from the static trip plans (by sequential algorithm)

Statistics	Static Trip Plan	Dynamic Trip Plan	% Improvement
Travel time (min)	7641.03	7036.06	7.91%
Waiting time (min)	4197.85	3714.79	11.50%
Travel distance (mile)	791.42	700.72	11.46%
Block swap cost ($)	37.12	33.50	9.77%
Reclassification cost ($)	57.57	47.88	16.83%
Transportation cost ($)	173.83	151.45	12.88%

Table 2.7: Improvements of dynamic trip plans which are different from the static trip plans (by bump-shipment algorithm)

Statistics	Static Trip Plan	Dynamic Trip Plan	% Improvement
Travel time (min)	7689.29	7037.86	8.47%
Waiting time (min)	4197.85	3714.79	11.50%
Travel distance (mile)	790.4	701.53	11.24%
Block swap cost ($)	37.12	33.64	9.38%
Reclassification cost ($)	57.38	47.69	16.88%
Transportation cost ($)	173.54	151.48	12.71%

2.6.2 Sensitivity Analysis

To gain more insights into the dynamic planning problem, we also investigate the impacts of the distribution cost per distance and capacity limits. Due to similar results from the sequential algorithm and bump-shipment algorithm, we only analyze the results for the sequential algorithm in the following discussion.

Impact of distribution cost per distance

Recall that our objective comprises the station cost (block swap cost and reclassification cost) and train-mile cost (i.e. traveled distance multiplied by a cost per distance). The results shown above use the cost per distance as 0.1 so that the train-mile costs are close to those in the current railroad's solution. However, the cost per distance can vary. Table 2.8 summarizes the impact of costs per distance ranging from 0 to 1 on the travel distance, station cost, and the transportation cost. As the cost per distance increases, the travel distance decreases since a larger weight is put on train-miles while the station cost increases. Figure 2.6 illustrates the tradeoff between travel distance and station cost:

Table 2.8: Impact of distance factor

Cost per distance ($/mile)	**Travel Distance (mile)**	**Station Cost ($)**	**Transportation Cost ($)**
0.00	550.49	49.50	49.50
0.01	516.47	49.51	54.67
0.05	508.25	49.77	75.18
0.10	502.58	50.17	100.43
0.20	496.98	50.92	150.32
0.40	494.09	52.03	249.67
0.60	491.96	52.89	348.06
0.80	490.16	53.69	445.82
1.00	489.99	53.74	543.74

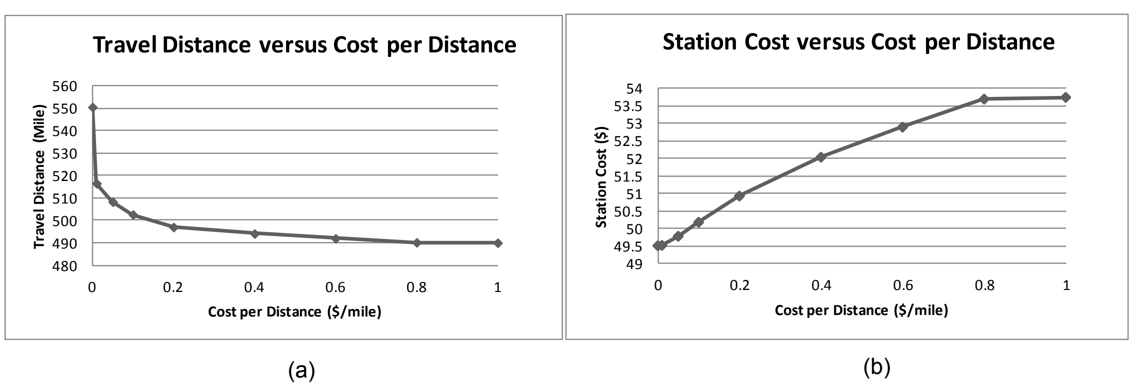

Figure 2.6: Impact of cost per distance

Table 2.9: Impact of capacity

Capacity Adjustment (car/train)	**Traveling Time (min)**	**Waiting Time (min)**	**Total Cost ($)**
-20	7797.02	4076.32	102.93
-10	7474.55	3865.62	101.43
-	7178.52	3789.10	100.87
+10	7036.34	3651.79	100.46
+20	6987.17	3603.84	100.43

Impact of capacity

To better understand the effect of capacity tightness on TCCR problem, we generate instances by modifying the capacity given by the rail company for the actual data set. We analyze five instances by adjusting the given capacities (either increase or decrease) in steps of 10 or 20 cars for all train-runs. Note that lower capacity limits would result in shipments that are unable to find a feasible train-blocking path to meet their due dates. For purpose of comparison, we accept the resultant paths for all shipments but add a penalty cost to the paths which do not meet the due date. Table 2.9 shows the impact of capacity on the traveling time, waiting time, and total cost. The new term *total cost* includes the transportation costs and penalty costs. As shown in Figure 2.7, increased capacities provide two benefits: firstly, more shipments fit in the fastest train-run paths resulting in decreased traveling times; secondly, more shipments are able to travel through their shortest train-blocking paths instead of requiring the next shortest paths. However, the rate of improvement decreases as the train capacities increase.

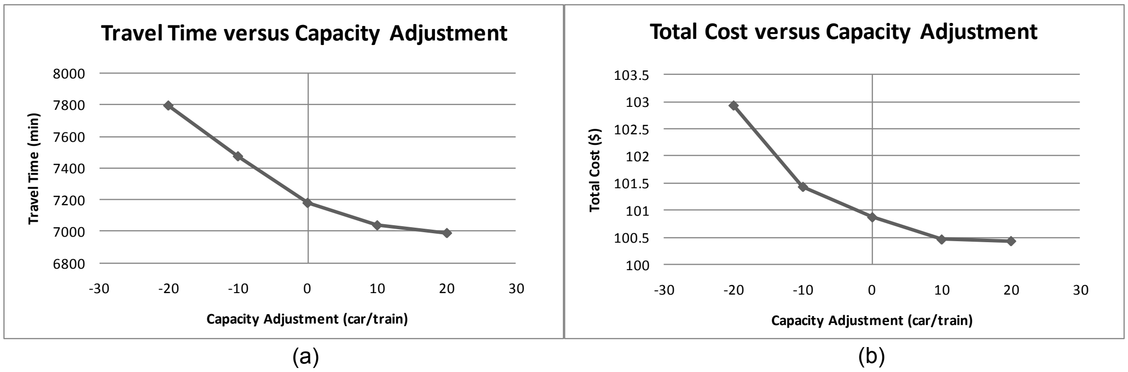

Figure 2.7: Impact of capacity limits

2.7. Conclusions

In this chapter, the time and capacity constrained routing problem is explored to support dynamic trip planning for railroads. Based on the tactical results of the blocking problem, the train scheduling problem, and the block-to-train assignment problem, the trip planning problem schedules multiple shipments through a railroad network. Specifically, the TCCR problem determines the optimal train-blocking paths and the optimal train-run paths for multiple shipments so as to minimize the total transportation cost while ensuring train capacities are not exceeded and the due dates of the shipments are met.

The problem is defined on a time-space-train-block network which is an extension of the train-block network. Based on these network representations, two integer programming models are formulated. Compared to the arc-based formulation, the path-based formulation tends to contain fewer variables and constraints and thus is more manageable.

Due to the complexity of the TCCR problem, two heuristic algorithms are proposed to solve the problem efficiently. The sequential algorithm sends shipments one by one and the trip for each shipment is optimized to meet its due date while considering train capacities. An alternative algorithm, the bump-shipment algorithm, schedules all shipments simultaneously without considering capacity limits and then reschedules shipments that exceed the capacity constraints. Both algorithms are implemented in C++ and solve the TCCR problem in a reasonable amount of time for representative data provided by a major U.S. railroad. Moreover, the dynamic trip plans provided by our algorithms have better performance for every cost factor than the static plan which is currently used in the railroad industry. To gain more insights, we also analyze the impact of distance factor and capacity limits. It is observed that larger costs per distance reduce travel miles but increase station costs, while larger capacity limits lower both the travel time and total cost.

In the trip planning problem, the blocking plan, the train schedule, and the BTA are all considered as given and fixed. This assumption helps to restrict the trip planning problem in a defined system. However, in practice the railroad companies may also need to adjust their current blocking plan, train schedules, or BTA if appropriate to save costs. Therefore, one area for future research is considering the situation when these pre-determined plans can be revised to improve overall effectiveness.

Chapter 3

Homogeneous Trailer Fleet Planning Problem

3.1. Introduction

Truck freight transportation plays a dominant role in many countries. In the U.S. logistics market, for instance, truck carrier cost was 542 billion in 2009, which accounted for 78% of total transportation cost [6]. For companies which rely on transportation in their business, fleet management is important to their operational and financial performance. From the Aberdeen survey of 317 logistics professionals [5], leading companies have found that effective fleet management solutions lead to benefits such as a 12.2% increase in service profitability, 13.0% improvement in vehicle utilization, 14.8% reduction in average travel time per job, 9.9% decrease in overtime pay and 27.9% increase in operator compliance.

For transportation providers, one of the challenges in fleet management is to establish a fleet of suitable size for a particular area. Fleet planning is generally considered a strategic level decision due to the expense incurred and the expected life of a fleet. Strategic fleet decisions involve considerable capital investment with significant financial implications. An oversized fleet increases fixed costs associated with purchase, maintenance, fleet management, and related infrastructure (garages, parking lots, technical back-up facilities). On the other hand, an undersized fleet decreases the ability to satisfy demands with an efficient delivery process. The goal of fleet planning is to minimize the risks associated with vehicle investment, improve efficiency and productivity, and reduce overall transportation and staff costs.

Our research is motivated by Air Liquide, an international industrial gas company that produces and distributes gases such as hydrogen, oxygen, argon, helium, and nitrogen for multiple business lines. For each geographic region, the gases are stored in liquified state at centralized sources and transported by trailers in bulk quantities to customer sites. When replenishing the gas inventory, trailers may visit a single customer or multiple customers on a route. Since trailers are engineered to safely carry a particular type of gas, they are dedicated to a specific product.

At the strategic level, an important decision is to define a suitable trailer fleet size in each region. Where no previous fleet exists, the company determines how many trailers to purchase. For an area where a fleet has been allocated, the company may adjust the existing fleet size by purchasing trailers or moving trailers from one region to another. In addition, the company acquires additional capacity by leasing trailers from the third party over shorter period to increase the flexibility and reduce the impact of demand fluctuations.

The fleet planning decisions are closely interrelated with trailer routing and scheduling practices. The routing practices influence the efficiency of the trailer resources. Thus, in most cases an integration of fleet planning decisions and routing considerations is important. In this research, we provide the strategic level fleet planning decisions and also consider the operational level details when adjusting current trailer fleet and designing delivery routes. Below a list of characteristics of the fleet planning problem considered in this research.

1. The optimal trailer fleet at each depot over time periods should be recommended.
2. The planning horizon is generally 1-2 years, and the length of time period can be a

month or a quarter.

3. The proposed fleet can be obtained by either investing in new trailers, transferring trailers from different locations, or renting trailers. New purchase decisions are made for the whole planning horizon, while relocation and rental decisions are made for each time period.
4. Customers are assigned to different depots, and can only be served by the trailers that are allocated to the assigned depot.
5. The seasonality of customer demands should be considered, i.e., the customer demands are varying by periods.
6. Estimates of the distribution cost should reflect routing practices.
7. Trailers return to the assigned depot after a maximum allowed working time.
8. Trailers transferred from another location may need to wait for approval to work in new depot. The waiting times varies by depot and they can be one or multiple periods.

In this chapter, we focus on the homogeneous trailer fleet planning (TFP) problem where all trailers are considered identical. In Section 2, we review the relevant fleet management and routing problems. Section 3 formalizes the problem as a mixed integer programming model, and a two-phase solution approach is developed in Section 4. In Section 5, we demonstrate the effectiveness of the solution approach using industry data. Finally, the conclusions and future research areas are given in Section 6.

3.2. Literature Review

In operation research literature, the vehicle routing problem (VRP) determines optimal routes for a fleet of identical vehicles, based at a central depot, to supply customers with known demands such that each customer is visited exactly once and the variable traveling costs (or distance) is minimized. Compared to classical VRP setting, our problem is distinct by the following characteristics:

1) Each trailer has capacity limit for freight, similar to the *capacitated vehicle routing problem* (CVRP).

2) Customers may be visited by more than one route, since customers have demand that is greater than even the largest trailer capacity. The related literature is referred as the *split delivery vehicle routing problem* (SDVRP).

3) The planning horizon is relatively long. The VRP generally addresses daily routing where each vehicle only operates one route. Therefore, the number of vehicles required is the same as the number of routes originated from central depot. When considering a long time period, a large trailer traverses multiple routes. The related literature includes the *inventory routing problem* (IRP), the *periodic vehicle routing problem* (PVRP), and the *vehicle routing problem with multiple trips* (VRPM).

In this section, we first review these relevant VRP variants, which concentrate on short-term decisions (e.g. days) at the tactical and operational level. Our problem affects medium- and long-term decision making (e.g. months or years), therefore, we also survey a strategic level problem which is called *strategic inventory routing problem* (SIRP) that addresses the fleet size decisions.

3.2.1 Tactical and Operational VRPs

Golden and Raghavan [44] systematically examine the significant technical advances that have evolved over the past few years for modeling and solving the VRPs and variants. Below we focus on the variants related to our problem.

Capacitated Vehicle Routing Problem (CVRP)

CVRP is a VRP with a fixed fleet of capacitated vehicles. That is, the CVRP incorporates the additional constraints that the total demand delivered on a route does not exceed the vehicle capacity. Generally, CVRP deals with homogeneous fleet. There are three basic formulations for the CVRP. A two-index vehicle flow formulation is originally proposed by Laporte et al. [50] where arcs are selected for a vehicle and subtour elimination constraints are developed. Another multicommodity flow formulation is proposed in an oil delivery problem by Garvin et al. [35]. This formulation combines assignment constraints for modeling vehicle routes and flow constraints for modeling commodity movements. The third general model is called set partitioning formulation [11], which associates binary variable with each feasible route in a set. This formulation has a possible exponential number of variables, however, the route feasibility (e.g. time and capacity) is implicity considered in the definition of the feasible route set. In addition, the practical importance of the CVRP provides the motivation for the effort involved in the development of heuristic and metaheuristic algorithms (see the surveys of Golden et al. [42], Laporte and Semet [51], Toth and Vigo [65]). In particular, Laporte et al. [49] report a computational comparison of some of the most important families of heuristics.

Split Delivery Vehicle Routing Problem (SDVRP)

The SDVRP is a relaxation of the VRP wherein it is allowed for a customer to be served by more than one route, if it reduces overall costs. This relaxation is very important when the customer demands are larger than the capacity of a vehicle. An additional variable is needed to decide the amount delivered to each customer on a route. Dror et al. [33] formulate the SDVRP as an integer program with several new classes of valid constraints. A constraint relaxation algorithm using branch and bound is developed for solving the SDVRP exactly. The computational results show that various constraints could successfully reduce the gap between the lower and upper bounds of the optimal solution value of SDVRP. Archetti et al. [12] propose an optimization-based approach to solve the SDVRP, which integrates meta-heuristic search with integer programming. The first step uses the information provided by a tabu search heuristic to identify the areas of the solution space that most likely contain good solutions, while the second step explores these parts of the solution space using a suitable integer programming model. The computational experiments show encouraging results.

Inventory Routing Problem (IRP)

The IRP arises where a vendor has the ability to decide the timing and planning of deliveries, as well as the routing, with the restriction that customers are not allowed to run out of product. There are no customer orders which must be met for everyday. In the IRP, however, the delivery company deals with a longer horizon and all decisions must be made keeping in mind their impact on what has to be done in the future. Campbell and Savelsbergh [24] present a two-phase approach to solve the IRP, where the first phase applies integer programming to create a delivery planning, and the second phase employs constructive heuristics to schedule the delivery routes. The computational experiments demonstrate the effectiveness of this approach for large-scale real-life instances. Bard et al. [17] use a similar decomposition scheme, which assigns each customer to a given day by solving a balanced assignment problem and then find good feasible solution for each day in the planning horizon. A unique aspect of the daily subproblems is the presence of satellite facilities for drivers to load product during a route. In other words, satellite replenishment allows the drivers to continue making deliveries without necessarily returning to the central depot. Three heuristics to solve VRP with satellite facilities (VRPSF) include a randomized version of the Clarke-Wright algorithm [28], a modified greedy randomized adaptive search procedure (GRASP) [47], and a revised sweep algorithm [41].

Periodic Vehicle Routing Problem (PVRP)

Typically, the planning period for the classical VRP is a single day, while PVRP extends the planning period to multiple days. With the PVRP, each customer is served with a given frequency during the multiple-day period. A solution to the PVRP consists of multiple sets of routes that jointly satisfy the demand constraints and the frequency constraints. A very popular approach decomposes the complex problem into two phases: customers are assigned to days in the first phase, and then a VRP is solved for each day for all customers scheduled on that day. For instance, Beltrami and Bodin [19] present two procedures for the municipal waste collection of the City of New York. In the first phase, routes are initially created according to the Clarke and Wright [28] heuristic, and then assigned to days of the week. In the second, customers are randomly assigned to days, and then routes are created using the Clarke and Wright method. Their analysis is limited since customers require service either three or six times per week. Chao et al. [26] follow the two-stages scheme in their approach. A integer program is solved in the first phase to assign delivery days to customers such that the total amount of demand delivered on each day is balanced. The second phase applies an interchange heuristic, where only the movement of one customer at a time to a new set of routes is considered. Gaudioso and Paletta [36] address a variation of PVRP, where the objective is not to minimize the total routing costs over the period but also to minimize the fleet size. They devise a one-customer-at-a-time allocation procedure, which determines the fleet size to accommodate the first i customers, once the fleet size to meet the first $i-1$ customers has been determined.

Vehicle Routing Problem with Multiple Trips (VRPM)

The classical VRP assumes that one vehicle is associated with exactly one route. However, the VRPM is characterized by vehicles and drivers working multiple routes within a given time period. Clearly, the one-to-one relationship between route and vehicle in classical VRP is relaxed, so VRPM generally solves a bin-packing problem to assign routes to vehicle. Taillard et al. [64] propose a two phase approach, where a set of VRP solutions are constructed from a set of routes generated using the Tabu Search (TS) heuristic of Rochat and Taillard [61] before bin-packing is used to allocate routes to vehicles. To compare with the benchmark of Taillard et al. [64], Brandao and Mercer [21] modified their heuristic based on the nearest neighbor rule and the insertion criterion to assign customers to routes within vehicles. Petch and Salhi [59] propose and test a multi-phase constructive heuristic for this problem. The main phases consist of two methods to generate a variety of VRP solutions, a bin-packing problem to allocate routes to vehicles, and improvement modules. The paper shows that this approach can be also adopted to solve the vehicle Beet mix with multiple trips.

3.2.2 Strategic VRPs

The classical VRP solves daily routing. While some of the variants address longer-term decisions, the time horizon is still limited to several days. Below we introduce a strategic inventory routing problem (SIRP), which focuses on the required fleet size instead of actual vehicle routes.

Strategic Inventory Routing Problem (SIRP)

Larson [52] introduced the SIRP, which is motivated by the fact that there is usually a significant amount of time between the purchase or lease agreement of the vehicles and their actual use for logistic operations. In the SIRP, the objective is to minimize the required fleet size to supply inventory. A strategic inventory/routing saving algorithm (SIRSA) is used to solve this problem, which divides customers into a set of clusters and assumes that all customers in the cluster are visited on a single route. The SIRSA fleet size estimate is based on a set of replenishment routes for a deterministic demand problem (DSIRP). Ideally, these routes are optimal or near-optimal for all possible realizations of the tactical IRP the vehicle fleet may eventually meet. Since the number of such realizations is enormous, it is not practical to use a fully detailed model of the tactical problems in developing these routes. Therefore, a set of routes is generated for the DSIRP based on a Simplified Tactical Model which is simple enough to allow estimation of the fleet size but also retains many fundamental characteristics of the tactical IRP's. As a strategic model, the author realizes that considerable attention must be devoted to the integration of the results of such a strategic model with scheduling and related concerns of tactical models.

Later, Larson and Webb [53] generalize the SIRSA through the introduction of two additional decision variables to address the period and phase of individual customer replenishment. Similarly, routing solutions based on customer-specific period and phase of replenishment are developed for a Simplified Tactical Model. Estimates of the fleet size are developed on the basis of these routing solutions. Computational results show that the period/phase approach generally yields significant reductions in average vehicle requirement.

3.2.3 Comparison to the existing literature

A substantial portion of the VRP literature focuses on tactical and operational routing details. The time period for a tactical VRP is usually limited to several days. Our problem combines the characteristics of CVRP, SDVRP, IRP, and VRPM, but is focused on a longer time period. At a strategic level, the focus is determining the fleet size that should be acquired - either for a long term contract or short-term adjustment instead of the details of daily routes.

In addition, our problem is distinguished by the combination of long-term and medium-term decisions. Traditional fleet acquisition at the strategic level addresses long-term decisions (i.e. purchase). Due to the varying demand, we also need to consider the medium-term adjustment options (i.e. relocation and rental). Our solution approach is inspired by some decomposition principles and practical routing algorithms.

3.3. Formulation

The homogeneous TFP problem is formulated over multiple periods and multiple depots to capture time vary demand and potential trailer relocation between depots. Information on depots, periods, customers, and trailers is considered. The objective is to minimize the total distribution cost as well as the costs associated with fleet adjustment. The customer demands must be met with the suggested trailer fleet while complying with the mandatory maximum working hours in each period. The balance of trailer relocation also need to be guaranteed.

One challenge in the formulation is to estimate the distribution cost. Designing routes for each day is impractical for this strategic level problem. We assume a set of candidate routes is given for each depot and period. A path-based model is used to select delivery routes and determine the delivery amount to each customer on a route so that customer demands are satisfied. The fleet decisions (i.e. long-term purchase, medium-term rental, and relocation) are made based on the selected routes. The following notation is used in this formulation:

Sets and Indices:

P set of depots, indexed by p

T set of periods, indexed by t

I set of customers, indexed by i

I^p subset of customers assigned to depot p

R set of routes, indexed by r

R^{pt} subset of routes originating at depot p during period t

Depot Parameters:

η_p available working time of a trailer at depot p in a time period (excluding the maintenance time)

Δ_p number of period(s) required for approval for incoming trailers at depot p

$d_{pp'}$ distance between depot p and depot p'

B budget for new trailer purchase

Customer Parameters:

δ_{it} demand requirement for customer i in period t

σ_i working capacity, which is the maximum mass that customer i can hold

Trailer Parameters:

q capacity of a trailer

α cost to purchase a new trailer

β_p cost to rent a trailer for one period at depot p

γ_p relocation cost per distance to transport a trailer from depot p to other depots

a_p number of trailers which are currently at depot p

Route Parameters:

c_r the cost of route r, including the fixed cost to load product and serve customers as well as the variable cost to transport product

τ_r the duration of route r, including the working time at sources and customer sites as well as the driving time on road

y_{ir} indicates if customer i is visited on route r

Decision Variables:

Ψ_r the number of times route r is selected

D_{ir} the amount of gas delivered to customer i on route r

X_{pt} the number of trailers allocated (owned by the company) to depot p during period t

N_p the number of trailers purchased for depot p at the beginning of planning horizon

$Y_{pp't}$ the number of trailers relocated from depot p to p' at the beginning of period t

Z_{pt} the number of trailers rented for depot p during period t

Homogeneous TFP Model:

$$\text{Minimize} \quad \sum_{t\in T}\sum_{p\in P}\sum_{r\in R^{pt}} c_r\Psi_r + \sum_{p\in P}\alpha' N_p + \sum_{t\in T}\sum_{p\in P}\beta_p Z_{pt} + \sum_{t\in T}\sum_{p\in P}\sum_{p'\in P}\gamma_p d_{pp'}Y_{pp't} \tag{3.1}$$

Subject to

$$\sum_{r\in R^{pt}} D_{ir}y_{ir} = \delta_{it}, \qquad \forall i\in I^p, \forall p\in P, \forall t\in T \tag{3.2}$$

$$D_{ir} \le \min\{q,\sigma_i\}y_{ir}\Psi_r, \qquad \forall i\in I^p, \forall r\in R^{pt}, \forall p\in P, \forall t\in T \tag{3.3}$$

$$\sum_{i\in I^p} D_{ir}y_{ir} \le q\Psi_r, \qquad \forall r\in R^{pt}, \forall p\in P, \forall t\in T \tag{3.4}$$

$$\sum_{r\in R^{pt}} y_{ir}\Psi_r \ge \max\left\{\left\lceil\frac{\delta_{it}}{q}\right\rceil, \left\lceil\frac{\delta_{it}}{\sigma_i}\right\rceil\right\}, \qquad \forall i\in I^p, \forall p\in P, \forall t\in T \tag{3.5}$$

$$\sum_{r\in R^{pt}} \tau_r\Psi_r \le \eta_p(X_{pt}+Z_{pt}), \qquad \forall p\in P, \forall t\in T \tag{3.6}$$

$$X_{pt} = a_p + N_p + \sum_{p'\in P} Y_{p'pt} - \sum_{p'\in P} Y_{pp't} \qquad \forall p\in P : \Delta_p = 0, t=1 \tag{3.7}$$

$$X_{pt} = X_{p,t-1} + \sum_{p'\in P} Y_{p'pt} - \sum_{p'\in P} Y_{pp't}, \qquad \forall p\in P : \Delta_p = 0, t=2,\cdots,N \tag{3.8}$$

$$X_{pt} = a_p + N_p - \sum_{p'\in P} Y_{pp't} \qquad \forall p\in P : \Delta_p \ge 1, t=1 \tag{3.9}$$

$$X_{pt} = X_{p,t-1} - \sum_{p'\in P} Y_{pp't}, \qquad \forall p\in P : \Delta_p \ge 2, t=2,\cdots,\Delta_p \tag{3.10}$$

$$X_{pt} = X_{p,t-1} + \sum_{p'\in P} Y_{p'p,t-\Delta_p} - \sum_{p'\in P} Y_{pp't}, \qquad \forall p\in P : \Delta_p \ge 1, t=\Delta_p+1,\cdots,N \tag{3.11}$$

$$\sum_{p\in P}\alpha N_p \le B \tag{3.12}$$

$$\Psi_{rt}, X_{pt}, Y_{pp't}, Z_{pt}, N_p \text{ integer } \ge 0, \qquad D_{irt} \ge 0 \tag{3.13}$$

The objective function (3.1) minimizes the distribution cost and the investment costs to acquire the trailer fleet. The first term indicates the total distribution cost over all depots and all periods. The second term captures the total "effective" purchase cost. Generally, a trailer is utilized for more than 10 years, so trailers may have a *salvage value* at the end of the planning horizon. Thus, we calculate the *effective purchase cost,* α', by deducting the salvage value at the end of planning horizon from the new purchase cost, α. The third term indicates the total rental cost and the forth term is the relocation cost across all the depots considered.

Constraint (3.2) ensures that the amounts delivered to customer i across all routes r meet the demand of customer i in period t. Since each customer has been assigned to a given depot p, we only need to consider the subset of routes R^{pt} to serve this customer. Constraint (3.4) ensures that the total amount delivered to all the customers on route r in period t does not exceed the trailer capacity q multiplied by the number of times route r is executed, or equals to zero if the route is not selected. For each individual customer i on route r, constraint (3.3) ensures that the total delivery amount is less than or equal to either the trailer capacity or the largest capacity which the customer can hold multiplied by the number of times route r is selected. Constraint (3.5) strengthens the model. The minimum number of visits required to customer i in period t is determined by the ceiling of the demand divided by either the trailer capacity or the maximum working capacity. Constraint (3.6) ensures that the total time required to execute all the selected routes in R^{pt} cannot exceed the available working time per trailer multiplied by the number of trailer allocated (either trailers owned by the company or rented) to depot p during period t.

Constraints (3.7) to (3.11) ensure the trailer movements among depots are correct. Recall that the purchase decision is made at the beginning of planning horizon, while the relocation and rental decisions are made at the beginning of each time period. Rented trailers can be used immediately, however relocated trailers may need to wait for an approval to provide service in a new depot. Specifically, constraints (3.7) and (3.8) address the simple case when there is no waiting period. Basically, the first period considers the number of trailers initially allocated, the number of new purchased trailers, and, if possible, the number of trailers that are transferred in and transferred out of the depot. The following periods consider the number of trailers previously allocated and the number of trailers that are transferred in and transferred out. When the waiting period is one, the in coming trailer capacities are lost in period 1 as captured by constraint (3.9). When waiting periods are greater than two, we consider the number of trailers in the previous period and the number of trailers transferred out for $t = 2, \cdots, \Delta_p(\geq 2)$, as captured by constraint (3.10). After the waiting period(s), constraint (3.11) considers the trailers which are transferred from other depots (the decisions for trailers transfer out have to be made Δ_p periods before). Lastly, Constraint (3.12) addresses the budget constraint for purchased trailers.

With this model, it is possible that a trailer may serve more than one route in a time period. In constraint(3.6), note that estimate of the fleet size, $X_{pt} + Z_{pt}$, is based on the total time of selected routes and the available working time for each trailer, which represents a lower bound to the true fleet size. For a more accurate estimate, we could solve a bin-backing problem to ensure all the selected routes are assigned to a specific trailer so that the total route time does not exceed the available working time for *each* trailer. However, we believe the obtained fleet size will in fact be reasonable for a strategic level estimate. Similar estimates were also applied in Larson [52] and Larson and Webb [53] when they studied a strategic inventory routing problem.

This formulation is complicated to solve due to the substantial number of decision variables and the interrelationships between depots and periods. The path-based formulation requires a large set of candidate routes. These routes are different by depot and period due to time varying customer demands. Furthermore, the potential relocation decisions of trailers link the routing and trailer fleet decisions together. In addition, frequently the rental and reallocation cost is much less than the total distribution cost. When the MIP model is terminated within a certain optimality gap, an undesirable solution may result. For instance, it may be suggested to move one trailer from depot A to B and one trailer from depot B to C instead of the more efficient movement from A to C, since the difference of these two relocation costs is negligible in the objective value. Therefore, we need an effective and efficient approach to solve this problem.

3.4. Two-Phase Approach

To keep the computational times within acceptable limits, we develop a heuristic approach that considers the demand fluctuations and the long-term nature of the strategic decisions. The solution approach decomposes the problem into two phases as shown in Figure 3.1. In the first phase, candidate routes are generated and a reduced model is solved to select routes so that the distribution cost is optimized and the suggested fleet sizes for all depots and periods are provided. The second phase addresses the fleet purchase, relocation, and rental decisions based on the initial allocation.

3.4.1 Phase I

The first phase concentrates on the distribution cost, which is a significant component of the objective (3.1). In the Phase I model, the number of trailers *needed* for each depot and each period are determined. Since the route selection and delivery decisions are independent of depots and periods, the problem can be separated by depots and periods, which greatly reduce the complexity of the original problem. Therefore, in this approach Phase I model is solved for each period. Below we remove the information on period t in the notation.

For a given period, a set of potential routes is generated based on customer demands. Then a reduced path-based model is used to select routes and determine the delivery amount to each customer in a given period. Based on the selected routes, the number of trailers needed for each depot, V_p, is estimated. In Phase I we do not distinguish among trailers owned by the company and rented from the third party. In addition, a new parameter is

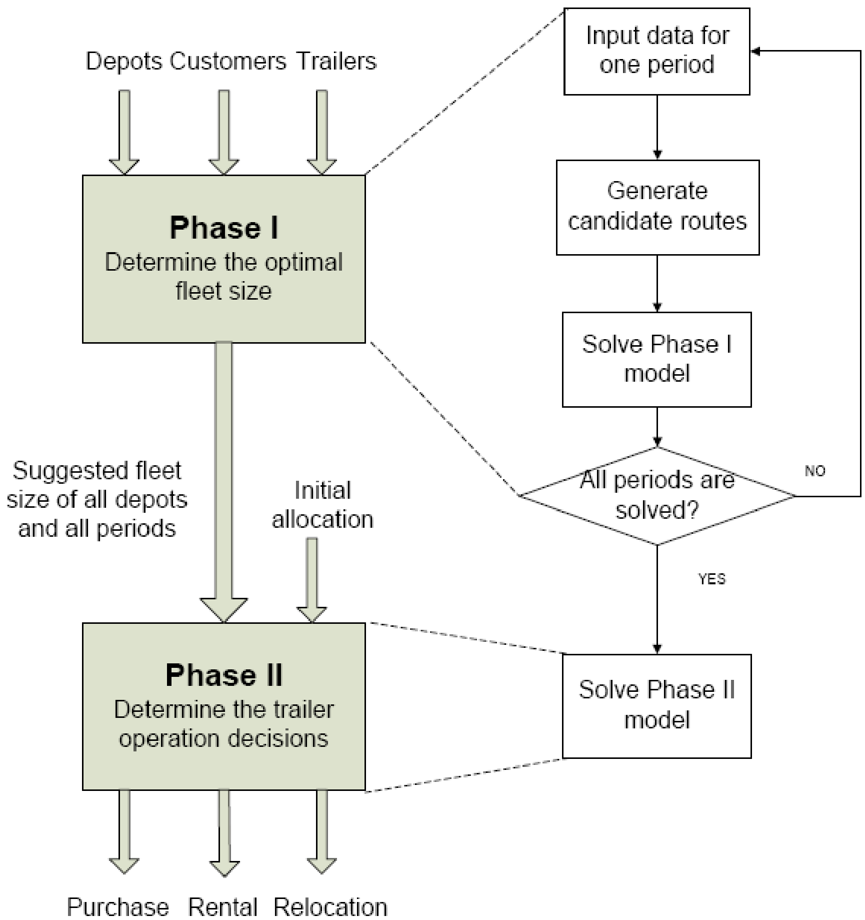

Figure 3.1: Two-Phase Approach for Homogeneous Trailer Fleet Planning Problem

introduced to represent the cost to use one trailer in one period. This *fixed cost*, h, can be considered as a weighted rental cost or a weighted purchase cost. The parameters, decision variables, and Phase I model are presented in the following:

Updated Parameters:

R^p	subset of routes originated at depot p in this period
δ_i	demand requirement for customer i in this period
h	fixed cost of a trailer

Decision Variables:

V_p	the number of trailers required at depot p
Ψ_r	the number of times route r is selected
D_{ir}	the amount of gas delivered to customer i on route r

Phase I model:

$$\text{Minimize} \quad \sum_{p \in P} \sum_{r \in R^p} c_r \Psi_r + \sum_{p \in P} h V_p \tag{3.14}$$

Subject to

$$\sum_{r \in R^p} D_{ir} y_{ir} = \delta_i, \qquad \forall i \in I^p, \forall p \in P \tag{3.15}$$

$$D_{ir} \leq \min\{q, \sigma_i\} y_{ir} \Psi_r, \qquad \forall i \in I^p, \forall r \in R^p, \forall p \in P \tag{3.16}$$

$$\sum_{i \in I^p} D_{ir} y_{ir} \leq q \Psi_r, \qquad \forall r \in R^p, \forall p \in P \tag{3.17}$$

$$\sum_{r \in R^p} y_{ir} \Psi_r \geq \max\left\{ \left\lceil \frac{\delta_i}{q} \right\rceil, \left\lceil \frac{\delta_i}{\sigma_i} \right\rceil \right\}, \qquad \forall i \in I^p, \forall p \in P \tag{3.18}$$

$$\sum_{r \in R^p} \tau_r \Psi_r \leq \eta_p V_p, \qquad \forall p \in P \tag{3.19}$$

$$\Psi_r \text{ integer } \geq 0, D_{ir} \geq 0 \tag{3.20}$$

Similar to the integrated model, the trailer fleet size, V_p, is determined based on the selected route times and the available working time of each trailer. By minimizing the objective, V_p is restricted to

$$V_p = \left\lceil \frac{\sum_{r \in R^p} \tau_r \Psi_r}{\eta_p} \right\rceil \tag{3.21}$$

3.4.2 Phase II

After solving Phase I model for each period, V_{pt} is used to represent the number of trailers needed for all depots and periods. Since the proposed trailer fleet can be either owned by the company or rented from outside, Phase II concentrates on how to obtain the suggested fleet V_{pt}. Recall that we can either purchase trailers at the beginning of planning horizon, or reallocate or rent trailers at the beginning of each time period.

For the Phase II model, the objective (3.22) includes the fleet acquisition and adjustment costs, which tend to be less than the total objective value in (3.1). Therefore, the Phase II model can reach a much lower optimality gap and thus provide high-quality solutions. The key constraint (3.23) connects Phase I to Phase II, which ensures that the suggested fleet is composed of trailers that are either owned or rented.

Phase II model:

$$\text{Minimize} \quad \sum_{p \in P} \alpha' N_p + \sum_{t \in T} \sum_{p \in P} \beta_p Z_{pt} + \sum_{t \in T} \sum_{p \in P} \sum_{p' \in P} \gamma_p d_{pp'} Y_{pp't} \tag{3.22}$$

Subject to

$$X_{pt} + Z_{pt} \geq V_{pt} \qquad \forall p \in P, \forall t \in T \tag{3.23}$$

$$\sum_{p \in P} \alpha N_p \leq B \tag{3.24}$$

$$X_{pt} = a_p + N_p + \sum_{p' \in P} Y_{p'pt} - \sum_{p' \in P} Y_{pp't} \qquad \forall p \in P : \Delta_p = 0, t = 1 \tag{3.25}$$

$$X_{pt} = X_{p,t-1} + \sum_{p' \in P} Y_{p'pt} - \sum_{p' \in P} Y_{pp't}, \qquad \forall p \in P : \Delta_p = 0, t = 2, \cdots, N \tag{3.26}$$

$$X_{pt} = a_p + N_p - \sum_{p' \in P} Y_{pp't} \qquad \forall p \in P : \Delta_p \geq 1, t = 1 \tag{3.27}$$

$$X_{pt} = X_{p,t-1} - \sum_{p' \in P} Y_{pp't}, \qquad \forall p \in P : \Delta_p \geq 2, t = 2, \cdots, \Delta_p \tag{3.28}$$

$$X_{pt} = X_{p,t-1} + \sum_{p' \in P} Y_{p'p,t-\Delta_p} - \sum_{p' \in P} Y_{pp't}, \qquad \forall p \in P : \Delta_p \geq 1, t = \Delta_p + 1, \cdots, N \tag{3.29}$$

$$X_{pt}, Y_{pp't}, Z_{pt}, N_p \text{ integer } \geq 0 \qquad \forall p \in P, p' \in P, t \in T \tag{3.30}$$

The main advantage of the two-phase approach is that the distribution decisions and the trailer planning decisions are separated, which greatly reduces the complexity of the integrated model.

3.4.3 Sweep-based routing algorithm

An important issue for Phase I is the number and the quality of the possible delivery routes. Since it is not practical to enumerate all the candidate routes, we limit the number of routes and focus on the most promising ones. In this approach, we apply a sweep-based heuristic to generate routes. Two types of routes are utilized in the actual distribution system:

(1) Single-customer routes, where only one customer is served on each route. The costs and times of these routes can be obtained in a very straightforward way by simply duplicating the trip between the depot and customer sites.

(2) Multiple-customer routes, where multiple customers are visited on a single route. The costs and times of these routes are more difficult to obtain, since both the customer clusters and customer visiting sequence impact the route significantly.

The basic idea underlying the algorithm is that we try to make a trailer dedicated to serve each customer so that the "majority" of demands are satisfied and the "remaining" demands are clustered so that each cluster is served on a single route. Since all customers have been assigned to a particular depot, we represent the location of each customer i using polar coordinates (θ_i, ρ_i) with the assigned depot as the center, θ_i as the angle, and ρ_i as the radius. All the customers are sorted in non-decreasing order of θ. If two customers have the same θ value, then they are ordered by non-decreasing ρ values.

Customers may have demands larger than the trailer size, and each customer can be served by either single-customer routes or multiple-customers routes. The multiple-customers routes are generated using normalized customer demands. That is, all demands δ_i are normalized using *full load drops* (fld_i) and *partial load drops* (pld_i), where

$$fld_i = \left\lfloor \frac{\delta_i}{q} \right\rfloor, \text{ and } pld_i = \frac{\delta_i}{q} - fld_i \tag{3.31}$$

The sweep-based algorithm tends to combine the customers that are close in proximity based on their pld_i so as to make a good use of the trailers. Note that pld_i is only used to generate candidate routes, but the actual delivery amount is determined by the Phase I model.

Follows we describe the sweep-based algorithm to generate *master routes*, which are primarily based on customers' partial load drop and trailer capacity. A maximum route time is used to limit the duration of a route. These master routes are further partitioned to small routes. Let N be the number of customers assigned to the depot and k be the index for master routes. Each customer can lead the sweep procedure, which is indicated by i_{start}.

Step 0: (initial) Set $i_{start} = 1$ and $k = 1$.

Step 1: (sort) Let the depot be the origin and calculate polar coordinates of each customer. Sort all customers in non-decreasing order of the polar angles θ. If two customers have the same θ value, then order them according to non-decreasing values of ρ.

Step 2: (master route start) Consider customers beginning with i_{start} on the ordered list. Construct master routes in the clockwise direction. That is, consider the list as a circle and go through the list in the clockwise direction.

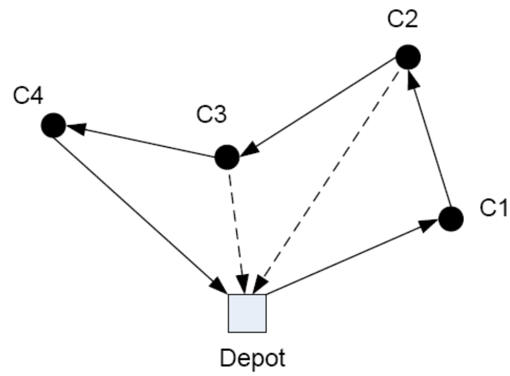

Figure 3.2: Partition of a master route

Step 3: (sweep) After selecting the initial customer, subsequent customers are sequentially added to a route as long as the total partial load drops are less then or equal to 1, the unit of trailer's capacity, and the maximum route time is not exceeded. If the total route time is exceeded, complete master route k, set $k \leftarrow k+1$, and go to Step 2 to start a new master route. If the total route time is not exceeded but the remaining capacity can not serve the next customer, skip that customer and consider the next customer on the ordered list.

Step 4: (change leading customer) If $i_{start} < N$, set $i_{start} \leftarrow i_{start} + 1$ and go back to Step 2 starting from a new customer.

To increase the size of route population, we apply a partition procedure to break down the master routes before sending them to the Phase I model. That is, the trailer may go back to the depot after visiting each customer on the master route. Figure 3.2 provides an example of four route partitions generated from a master route Depot-C1-C2-C3-C4-Depot, where the broken routes include Depot-C1-Depot, Depot-C1-C2-Depot, Depot-C1-C2-C3-Depot, and Depot-C1-C2-C4-Depot.

In the sweep-based algorithm, customers are ordered and added in clockwise direction. To provide additional candidate routes, the customers can also be traversed in a counterclockwise direction as well. A large set of potential routes increases the computational requirements for solving the problem. Therefore, it is helpful to remove the routes which are impossible or impractical and focus on the routes which are promising. As shown above, the maximum route time is considered as a primary criteria to limit the route size. Additional reduction techniques include the following:

- Limit the number of customers on a route. Due to the unloading time and the driver's workload, we restrict the number of customers consecutively visited on a route.
- Limit the number of skipped customers. In Step 3, when the remaining trailer capacity cannot serve the next customer but is still below the maximum route time, the vehicle would skip the customer and consider the following ones. However, it is impractical to skip many customers and connect two customers that are far from each other. Therefore, we also restrict the number of customers that we skip consecutively.

3.5. Numerical Studies

In this section, we evaluate the effectiveness of the proposed two-phase approach using actual data from Air Liquide. The solution approach has been implemented in Visual Studio 2008 (C#) and OPL 6.3 / CPLEX 12.3 with the input data and output results organized using Excel. All experiments were conducted on a 64-bit Intel Xeon X5570 Workstation (CPU 2.93GHz, 8 cores, RAM 24G).

Data for 382 liquid nitrogen (LN2) customers at three depots are used in this analysis. The characteristics of the three depots are summarized in Table 4.11. Customer demands for four periods are also provided, ranging from 500 kg to more than 3,800,000 kg per period. We observe that some customer demands vary substantially by period. For example, the largest quarterly demands for 154 customers are more then twice their lowest demands.

One type of trailer is considered in this study. Air Liquide provided the information about

Table 3.9: General information

Depot ID	No. of Customers	No. of Trailers Allocated
Depot 1	58	2
Depot 2	128	2
Depot 3	196	3

the current allocation of trailers and the associated trailer costs including purchase, rental, and relocation cost. We also consider trailer availability due to maintenance by including the percentage of time that each trailer can work in a time period.

The output of Phase I is summarized in Table 3.10 with the distribution costs and the number of trailers needed for each depot and each period. In this analysis, all costs are converted to U.S. dollar using the latest currency exchange rate. All the Phase I models are solved to 0.5% optimality gap within 1 minutes, for a total computational time of 2.68 minutes. As shown, depot 1 requires one trailer for each time period. Depot 2 requires two trailers for most of the time periods, except for period 3 when three trailers are needed due to the high demand for many customers. Similarly, depot 3 requires three trailers for most of the time periods and four trailers for period 3. Phase II uses the number of suggested trailers as input and provides decisions regarding purchase, relocation, and rental, as shown in Table 3.11. In this scenario, no new trailers are needed. One relocation occurs between depot 1 and depot 2 at the beginning of period 1, since depot 1 has an extra trailer. A short-term trailer rental is required for depot 3 during period 3 to address the high customer demands.

3.5.1 Two-phase approach compared to integrated model

To compare the two-phase approach and the integrated model, we solve both methods for 30, 150, and 300 customer data sets based on generated candidate routes. To further evaluate

Table 3.10: Results of Phase I

Objectives (USD)				
	Period 1	Period 2	Period 3	Period 4
Objective Value	1,429,576	1,495,496	1,787,923	1,499,045
Distribution Cost	1,399,576	1,465496	1,747923	1,469,044
Suggested Trailer Allocation				
Depot ID	Period 1	Period 2	Period 3	Period 4
Depot 1	1	1	1	1
Depot 2	2	2	3	2
Depot 3	3	3	4	3
Statistics				
	Period 1	Period 2	Period 3	Period 4
MIP Gap Tolerance	0.5%	0.5%	0.5%	0.5%
Wall Clock Time (mins)	0.45	0.88	0.83	0.52

Table 3.11: Results of Phase II

	Relocation Decisions (Total Relocation Cost = $ 1,577)			
	From Depot	To Depot	At Period	No. of Trailers Moved
	Depot 1	Depot 2	Period 1	1
	Rental Decisions (Total Rental Cost = $ 5,000)			
Depot ID	Period 1	Period 2	Period 3	Period 4
Depot 1	0	0	0	0
Depot 2	0	0	0	0
Depot 3	0	0	1	0

the impact of the length of planning horizon, the time periods are extended to eight periods and twelve periods. Table 4.16 compares the objective value and computational time for the integrated model and the two-phase approach, where the objective value of two-phase approach is computed by summing up the distribution cost in Phase I and trailer operation costs in Phase II. Since Phase I and II also include MIP models, we set all the MIP gap tolerances to 0.5%.

As shown in Table 4.16, the difference in objective values between the integrated model and two-phase approach is less than 5% (except for the smallest case with 30 customers) and the difference decreases as the problem size increases. As the planning horizon increases, the computational time for the integrated model increases substantially while the two-phase approach solves the largest case in less than five minutes. When the planning horizon is extended to eight periods, the integrated model requires approximately one hour while the two-phase approach solves the same data set within 2 minutes and provides objective solutions within 3.23% of the integrated model. When the time periods are increased to

Table 3.12: Comparison between two-phase approach and complete model (MIP gap 0.5%)

Case	Objective value (USD)			Running time (mins)	
	Model	Two-phase[a]	Difference	Model	Two-phase
30 Customers - 4 Periods	414,930	448.641	8.12%	0.47	0.08
150 Customers - 4 Periods	2,489,300	2,609,876	4.84%	1.56	0.66
300 Customers - 4 Periods	3,832,300	3,962,933	3.41%	2.10	1.02
300 Customers - 8 Periods	8,527,800	8,803,473	3.23%	60.18	1.95
300 Customers - 12 Periods	11,692,400 (0.86%)	11,965,983	2.34%	> 1440	4.32

[a] Objective value of two-phase approach = distribution cost in Phase I + objective in Phase II

twelve, the integrated model does not achieve a 0.5% MIP gap after 24 hours. In contrast, the two-phase approach solves in 4 minutes and the objective value is also very close to the best-known solution (2.34%).

3.5.2 Impact of route size

In Section 4, we discussed the importance of limiting the number of customers on one route (i.e. route size) so as to reduce the number of potential routes generated. In the previous study, the route size is limited to five such that at most four customers are allowed on a candidate route. In the following we demonstrate the impact of route sizes. The Phase I model is solved for one period with the trip size adjusted from one to ten. The results are presented in Table 3.13 which summarizes the number of routes, variables, and constraints for various route sizes (number of customers on one route). The resulting distribution costs and computational times are also reported.

As shown, for the first case with the single-customer routes that limit all route sizes to be one, only 331 routes are created and the model is solved very quickly. After that, we add

Table 3.13: Impact of route size

Route limit	No. of Routes	No. of Var.	No. of Con.	Distribution Cost (USD)	Time (mins)
1	331	1005	1997	1,700,140	0
2	1649	5615	6607	1,458,829	0.02
3	2276	8750	9742	1,410,829	0.07
4	2856	12230	13222	1,404,237	0.25
5	3330	15548	16540	1,399,576	0.43
6	3660	18188	19180	1,399,361	0.54
8	3965	21030	22022	1,399,117	0.78
10	4039	21869	22861	1,399,097	0.92

the possibility to combine customers in a single route. By adjusting the route size to include both single-customer and multiple-customer routes, the problem size increases accordingly but the distribution cost is saved substantially. Eventually, the route sizes are increased to include ten customers, the model takes around one minute to solve the candidate set with over 4,000 routes. The distribution cost decreases by 17.7% compare to the cases with only single-customer routes.

These results demonstrate the tradeoff between solution times and distribution costs. Too few routes are solved quickly resulting in unsatisfying solutions, while too many routes provide better solutions with increasing computational efforts. Therefore, it is critical to find a suitable route size limits. Figure 3.3 shows that the impact of route size on distribution cost. It shows that as the route size increases, distribution costs decreases as expected but at a decreasing rate. Particularly, when the trip size increases from five to ten, the distribution cost drops from $1,399,576 to $1,399,097 (less than 0.04%), but the solution time is doubled. This fact well supports our assumption that a trailer is unlikely to visit more than five customers in a single route, which is also consistent with findings by Song and Savelsbergh

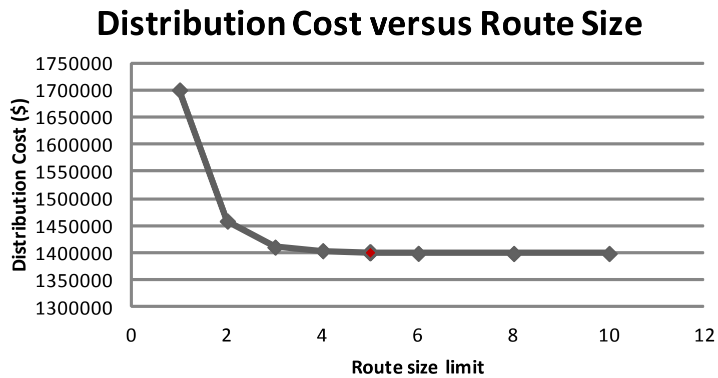

Figure 3.3: Impact of route size on distribution cost

[63] and industry practices.

3.6. Conclusions

Trailer fleet planning is an important issue in the industrial gas distribution. This paper focuses on a homogeneous trailer fleet planning problem that addresses important fleet management decisions with multiple depots and multiple periods in order to minimize total distribution costs and fleet investment costs. This problem combines both the long-term fleet acquisition decisions and the medium-term vehicle relocation and rental decisions, which are useful for increasing the flexibility when the demands are changing by periods.

The problem is formulated as a mixed-integer program for a given set of potential gas delivery routes. A two-phase approach is proposed where Phase I concentrates on distribution costs and provides the suggested fleet size, and Phase II addresses the trailer purchase, relocation, and rental decisions. Specifically, in Phase I, a sweep-based heuristic generates a set of routes that capture the operational characteristics of the actual gas delivery process, and then a reduced model is used to select routes for meeting customers' demands and determine the preferred fleet size. This decomposition approach removes the interactions between depots and periods in Phase I and reduces the complexity. Phase II provides aggregate resource plans based on outputs of Phase I.

The numerical studies, conducted using a dataset with three depots and nearly 400 customers, demonstrate the effectiveness of the approach. Compared to the integrated optimization model, the decomposition approach obtains good-quality solutions in a reasonable computational time. Additionally, the solution approach demonstrates a more robust performance as the number of time periods increases. Based on the results, Air Liquide is currently integrating the proposed solution approach in their worldwide distribution planning software.

In industry, a homogeneous fleet is easier to manage, but a heterogeneous fleet generally provides more flexible and cost-effective service for demand variations. In addition, some customers have particular requirements for a vehicle type. As future research, we plan to explore the heterogeneous trailer fleet planning problem for industrial gas provider.

Chapter 4

Heterogeneous Trailer Fleet Planning Problem

4.1. Introduction

In industry, a fleet of vehicles can be homogeneous or heterogeneous. Although a homogeneous fleet is easier to manage, organizations may prefer a diverse set of vehicles due to operational constraints and the inherent benefits of versatility. Generally, a heterogeneous fleet provides more flexible and cost-effective service for meeting demand variations, especially when customers have particular requirements for a vehicle type.

With a heterogeneous fleet, the vehicles are distinguished by the physical dimensions, the characteristics of individual vehicles, and the economic factors. Below three main factors are summarized to categorize trailer types. Physical dimensions such as the length, width, and height of a trailer determine the carrying capacity. Physical dimensions and weight sometimes also constrain access to the customer sites and road transportation system. Regarding the costs, large trailers generally have a higher fixed cost for purchase or rental, but relatively lower unit cost if capacity utilization is sufficiently high. The variable cost (i.e. distribution cost per mile) may also vary by the trailer sizes. Beyond physical dimensions, there are also various attributes that make particular types of trailers compatible of incompatible. For instance, industrial gas customers with higher working pressure often need vehicles with special equipment for loading and unloading product.

The motivation for this research is the heterogeneous trailer fleet that Air Liquide utilizes in the European markets. The heterogeneous trailer fleet planning (TFP) problem is complex due to the following considerations.

(1) **Multiple trailer types**

With multiple trailer types, the goal is not only to decide the number of trailers that should be allocated at each depot, but also the number of trailers of *each type* that should be allocated. In the reminder of the paper, we use the term *fleet composition* to include the determination of both the types of vehicles and the number of trailers of each type.

(2) **Customer-trailer restrictions**

Customers may have restrictions for trailer types that can serve on site. The restrictions may include trailer size, weight, pump type, and others.

(3) **Continuous delivery**

Besides the incorporation of multiple trailer types, more realistic distribution practices are considered in this research. In each geographic region, there is a central *depot* where the trailers park and receive maintenance. Industrial gases are stored in liquified state at *sources*, which may or may not be the same location as depots. The product is delivered by trailers to customer sites. Customers can only be served by their assigned depot(s) and source(s). In practice, a trailer leaves the assigned depot, fills at a source, and serves a sequence of customers which have been assigned to the source. When the remaining product in the trailer falls below a threshold, the trailer may return to the same source or visit a new source to refill. The trailer repeats delivering and refilling,

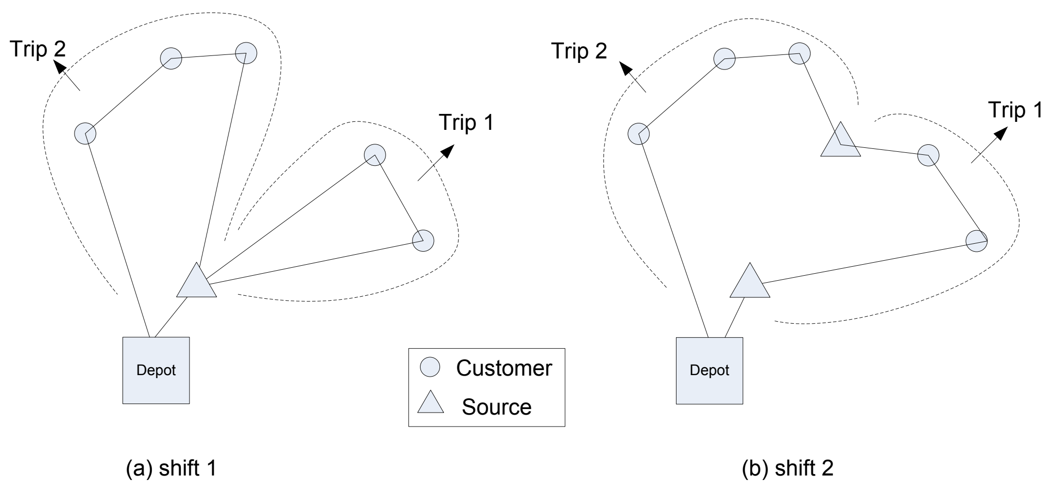

Figure 4.1: Example of shifts and trips

until the maximum working time is reached, and then returns to the depot. A complete tour is referred to as a *shift*, while the sequence of deliveries between refills is referred to as a *trip*. With *continuous delivery*, a shift may contain more than one trip. Figure 4.1 illustrates two typical shifts in the current distribution system. Both shifts contain two trips, where shift 1 refills at the same source twice and shift 2 visits two different sources. Note that the product from a source can only be delivered to the customers which have been assigned to that source.

(4) **Varying customer demand**

The heterogeneous TFP also considers customer demands that vary by time period. With varying demands, the fleet requirements can be met by either investing in new trailers, relocating trailers from different depots, or leasing trailers for a few periods.

The remainder of the chapter is organized as follows. In Section 4.2, we review the relevant literature on fleet composition problems. Section 4.3 provides a mixed integer programming model for the heterogeneous TFP problem. A modified two-phase approach is developed in Section 4.4. In Section 4.5, we demonstrate the approach using industry data, and Section

4.6 summarizes the conclusions and future research.

4.2. Literature Review

The well-known vehicle routing problem (VRP) determines optimal routes for a fleet of identical vehicles, based at a central depot, to supply customers with known demands such that each customer is visited exactly once and the variable traveling costs (or distance) is minimized. Compared to the classical VRP problem, our problem is distinct by the following characteristics:

1) There are multiple trailer types and the trailer fleet size is unknown. A similar problem is referred as the *fleet size and mix vehicle routing problem* (FSMVRP), which is a special case of *heterogeneous vehicle routing problem* (HVRP).
2) Customers may be served by more than one route, since some customer demands are greater than the largest trailer capacity. A related problem is referred as the *split delivery vehicle routing problem* (SDVRP).
3) The planning horizon is relatively long. Since the VRP generally addresses daily routing, each vehicle only operates one route. Therefore, the number of vehicles is the same as the number of routes originated from central depot. However, in a longer time, one trailer may traverse multiple routes . Related problems include the *inventory routing problem* (IRP), the *periodic vehicle routing problem* (PVRP), and the *vehicle routing problem with multiple trips* (VRPM).
4) A trailer can refill with product during a route. A related problem is referred as the *vehicle routing problem with satellite facilities* (VRPSF).

The literature on the SDVRP, IRP, PVRP, and VRP is introduced in the discussion of the homogeneous TFP (Section 3.2). In this section, we focus on the literature related to the FSMVRP and VRPSF.

4.2.1 Fleet Size and Mix Vehicle Routing Problem

The most general version of the HVRP designs a set of feasible routes to minimize total costs such that each customer is visited by exactly one route, and the number of routes performed by a particular vehicle type does not exceed its limit. In addition to the general problem, the following characteristics are often modified in the literature:

(1) The vehicle fleet is composed by an *unlimited* number of vehicles for each type;

(2) The fixed costs of the vehicles are *not considered*;

(3) The routing costs are *independent* of vehicle types.

The fleet size and mix vehicle routing problem (FSMVRP) modifies (2) and (3) and considers the fleet size as additional decision variables. Specifically, FSMVRP determines the optimal fleet composition and minimizes the total cost function that includes fixed costs for managing the vehicles in the fleet as well as variable routing costs. In one of the first papers on the FSMVRP, Golden et al. [43] propose a mixed integer formulation that assumes the variable costs are independent of the type of vehicle. To further generalize the model, Osman and Salhi [57] introduce different variable costs and time factors per distance unit for each vehicle type and include service time for customers in the model. The FSMVRP is addressed using exact approaches, constructive heuristics, and meta-heuristics.

An exact approach to the FSMVRP is proposed by Pessoa et al. [58] where robust and powerful cuts are incorporated into a branch-cut-and-price algorithm. The authors solve instances up to 75 vertices (i.e. customers) to optimality. Yaman [67] proposes another exact approach when developing formulations and valid inequalities to compute lower bounds to the problem. Six different formulations are developed and the linear programming bounds of these formulations are compared. To improve the lower bounds, the paper also derives valid inequalities and lifts some of the constraints. Baldacci et al. [14] present a MIP formulation for the FSMVRP with fixed unit running costs. They introduce new covering-type and fleet-dependent capacity inequalities to improve the resulting bounds. The authors state that the new cuts may contribute to the development of new overall algorithms for this class of problems. Choi and Tcha [27] propose a column generation (CG) based approach to design routes for a heterogeneous fleet with various capacities, fixed costs and variable costs. A tight IP model is presented and the LP relaxation is solved using a CG technique. Two dynamic programming schemes developed for the classical VRP are modified to efficiently generate feasible columns. With the tight lower bounds obtained, the branch-and-bound procedure is activated to obtain an integer solution efficiently.

Since the complexity of FSMVRP makes it difficult and impractical to implement exact approaches in practices, constructive heuristics are often employed, most of which are also successful in solving the VRP. In Golden et al. [43], several heuristics are proposed based on the savings technique for the VRP presented by Clark and Wright [28]. They also describe an approach based on a giant TSP-tour that is partitioned into subtours that satisfy the capacity of the vehicles. Desrochers and Verhoog [30] develop another savings heuristic for the FSMVRP by extending their algorithm initially designed for the classical VRP. The method is a matching based savings algorithm using fusion of successive routes. In each iteration, the best fusion is selected by solving a weighted matching problem, which provides a less myopic criteria than the usual savings heuristics. Renaud and Boctor [60] present a sweep-based heuristic for FSMVRP. The proposed algorithm first generates a large number of routes that are serviced by one or two vehicles. The routes and vehicles are then selected by solving a set partitioning problem with a special structure to optimality in polynomial time. Salhi and Rand [62] develop another efficient heuristic for FSMVRP, which incorporates a route perturbation procedure (RPERT) on existing or constructed routes to improve the vehicle utilization of the whole fleet. The algorithm starts by solving a classical VRP with a given vehicle capacity to create a starting solution for the search. Then the algorithm checks if other vehicle types can be used on the routes and whether it is economical to remove a given route from the solution and insert the customers onto other routes. Other refining procedures such as reallocation of customers from one route to another, swapping customers between routes, and combining or splitting routes are performed to check if it is possible to improve the solution.

A different strategy to solve the FSMVRP is to use neighborhood search procedures, such as the tabu search meta-heuristic, to repeatedly improve an existing solution. Osman and Salhi [57] take an initial solution produced by the enhanced constructive heuristic of Salhi and Rand [62] and apply a tabu meta-heuristic method with short term memory and with moves defined by a 1-interchange mechanism. Gendreau et al. [38] propose a tabu search method to solve for the heterogeneous fleet vehicle routing problem, which is relatively complicated and requires the use of the generalized insertion heuristic GENIUS [37]. Wassan and Osman [66] present a reactive tabu search with several neighborhoods and special data structures for efficiency. The algorithm comprises several variants obtained from the selection of different neighborhood mechanisms, tabu restrictions, and tabu tenure schemes. More recently, Brandäo [22] proposes a deterministic tabu search algorithm to solve the FSMVRP, which is based on three types of neighborhood moves: single insertion, double insertion and swap. The method uses intensification and diversification procedures and allows infeasible solutions with a penalty during the search.

4.2.2 Vehicle Routing Problem with Satellite Facilities (VRPSF)

A unique aspect of the VRPSF is the use of satellite facilities for drivers to load product while executing a route. Satellite replenishment allows the drivers to continue making deliveries without necessarily returning to the central depot. Bard et al. [16] develop a branch-and-cut methodology for solving the VRPSF subject to capacity and route time constraints. A mixed-integer linear programming formulation is provided and then a series of valid inequalities are described so as to cut off solutions to the linear programming relaxation. Several separation heuristics are then outlined that are used to generate the cuts. A VRP heuristic is embedded in the branch-and-cut scheme to find good feasible solutions at each stage of the computations. Later, Bard et al. [17] introduce three VRP-based heuristics for the inventory routing problem with satellite facilities (IRPSF). These heuristics include a randomized version of the Clarke-Wright algorithm [28], a modified greedy randomized adaptive search procedure (GRASP) [47], and a revised sweep algorithm [41].

4.2.3 Comparison to Existing Literature

A substantial portion of the VRP literature focuses on tactical and operational routing details. The length of period for a classical VRP is usually limited to several days. Our problem combines characteristics of CVRP, SDVRP, IRP, and VRPSF, which makes the problem very difficult to solve for a longer time horizon. At the strategic level, a company is more interested to know the fleet size that should be acquired, either for a long term contract or short-term adjustment, instead of the details of daily routes. Therefore, the focus of our problem is on the strategic fleet planning decisions but the characteristics of the continuous delivery practice are also considered.

In addition to continuous delivery routes, our TFP problem combines both long-term and medium-term decisions. Traditional fleet acquisition problems focus on long-term decisions (i.e purchase). Due to time varying demand, we also consider medium-term adjustment options (i.e. relocation and rental). Our solution approach is inspired by some decomposition principles and successful routing algorithms.

4.3. Formulation

Similar to the homogeneous TFP in Chapter 3, we consider the heterogeneous TFP problem across multiple depots and multiple periods. In addition to the depots, periods, and customers information, the following aspects are also considered.

- Trailer types with different capacities and costs (i.e. purchase, rental, relocation costs).
- Customer-trailer restrictions to indicate if a particular trailer type is allowed to access a customer site
- Sets of candidate shifts and trips. Each shift contains one or more trips.

The objective is to minimize the total distribution cost and the costs associated with fleet purchase, relocation, and rental. To estimate the distribution cost, we assume a set of candidate shifts is given for each depot, period, and trailer type. Each shift contains one or more trips. The model selects shifts and determines delivery amounts to each customer on each trip. Fleet decisions (purchase, relocation, and rental) are determined based on the selected shifts for each trailer type. As constraints, the suggested trailer fleet must meet customer demands while complying with the mandatory maximal working hours in each period. The conservation constraints of trailer relocation and customer-trailer restrictions also need to be considered. Below are the mathematical notation used in this formulation:

Sets and Indices:

P	set of depots, indexed by p
T	set of periods considered, indexed by t
L	set of trailer types, indexed by l
I	set of customers, indexed by i
I^p	subset of customers assigned to depot p
S	set of shifts, indexed by s
S^{lpt}	subset of shifts for trailer type l originated at depot p during period t
R	set of trips, indexed by r
R^s	subset of trips belonged to shift s

Depot Parameters:

η_p available working time for a trailer at depot p in one period (excludes the maintenance time)

Δ_p number of period(s) for incoming trailers to wait for approval at depot p

$d_{pp'}$ distance between depot p and depot p'

B budget for new trailer purchases

Customer Parameters:

δ_{it} demand requirement for customer i in period t

σ_i working capacity, which is the maximum mass that customer i can store

e_{il} indicates if customer i allows trailer type l to serve on site

Trailer Parameters:

q_l capacity of trailer type l

α_{lp} purchase cost of trailer type l at depot p

β_{lp} rental cost of trailer type l for one period at depot p

γ_{lp} relocation cost per distance to transport a trailer of type l from depot p to other depots

a_{lp} number of trailer type l which are currently at depot p

Shift and Trip Parameters:

c_s the cost of shift s, including the fixed cost to serve customers and load product as well as the variable cost to transport product

τ_s the duration of shift s, including the working time at customer sites and sources as well as the driving time on road

y_{ir} indicates if customer i is visited on trip r

Decision Variables:

Ψ_s the number of times shift s is executed

D_{ir} the amount of gas delivered to customer i on trip r

X_{lpt} the number of trailer type l allocated (owned by the company) to depot p during period t

N_lp the number of trailer type l purchased at depot p at the beginning of planning horizon

$Y_{lpp't}$ the number of trailer type l relocated from depot p to p' at the beginning of period t

Z_{lpt} the number of trailer type l rented at depot p at period t

Heterogeneous TFP Model:

$$\text{Minimize} \quad \sum_{s\in S} c_s\Psi_s + \sum_{p\in P}\sum_{l\in L}\alpha'_{lp}N_{lp} + \sum_{t\in T}\sum_{p\in P}\sum_{l\in L}\beta_{lp}Z_{lpt} + \sum_{t\in T}\sum_{p\in P}\sum_{p'\in P}\sum_{l\in L}\gamma_{lp}d_{pp'}Y_{lpp't} \tag{4.1}$$

Subject to

$$\sum_{l\in L}\sum_{s\in S^{lpt}}\sum_{r\in R^s} D_{ir}y_{ir} = \delta_{it}, \qquad \forall i\in I^p, \forall p\in P, \forall t\in T \tag{4.2}$$

$$\sum_{i\in I^p} D_{ir}y_{ir} \le q_l\Psi_s, \qquad \forall r\in R^s, \forall s\in S^{lpt}, \quad \forall l\in L, \forall p\in P, \forall t\in T \tag{4.3}$$

$$D_{ir} \le \min\{q_l,\sigma_i\}y_{ir}e_{il}\Psi_s, \qquad \forall i\in I^p, \forall r\in R^s, \forall s\in S^{lpt}, \quad \forall l\in L, \forall p\in P, \forall t\in T \tag{4.4}$$

$$\sum_{l\in L}\sum_{s\in S^{lpt}}\sum_{r\in R^s} y_{ir}\Psi_s \ge \max\left\{\left\lceil \frac{\delta_{it}}{\max\limits_{l\in L} q_l}\right\rceil, \left\lceil\frac{\delta_{it}}{\sigma_i}\right\rceil\right\}, \qquad \forall i\in I^p, \forall p\in P, \forall t\in T \tag{4.5}$$

$$\sum_{s\in S^{lpt}} \tau_s\Psi_s \le \eta_p(X_{lpt}+Z_{lpt}), \qquad \forall p\in P, \forall t\in T, \forall l\in L \tag{4.6}$$

$$X_{lpt} = a_{lp} + N_{lp} + \sum_{p'\in P} Y_{lp'pt} - \sum_{p'\in P} Y_{lpp't} \qquad \forall p\in P: \Delta_p = 0, \forall l\in L, t=1 \tag{4.7}$$

$$X_{lpt} = X_{lp,t-1} + \sum_{p'\in P} Y_{lp'pt} - \sum_{p'\in P} Y_{lpp't}, \qquad \forall p\in P: \Delta_p = 0, \forall l\in L, t=2,\cdots,N \tag{4.8}$$

$$X_{lpt} = a_{lp} + N_{lp} - \sum_{p'\in P} Y_{lpp't} \qquad \forall p\in P: \Delta_p \ge 1, \forall l\in L, t=1 \tag{4.9}$$

$$X_{lpt} = X_{lp,t-1} + \sum_{p'\in P} Y_{lp'p,t-\Delta_p} - \sum_{p'\in P} Y_{lpp't}, \qquad \forall p\in P: \Delta_p \ge 1, \forall l\in L, \quad t=\Delta_p+1,\cdots,N \tag{4.10}$$

$$X_{lpt} = X_{lp,t-1} - \sum_{p' \in P} Y_{lpp't}, \qquad \forall p \in P : \Delta_p \geq 2, \forall l \in L, t = 2, \cdots, \Delta_p \tag{4.11}$$

$$\sum_{p \in P} \sum_{l \in L} \alpha_{lp} N_{lp} \leq B \tag{4.12}$$

$$\Psi_s, X_{lpt}, Y_{lpp't}, Z_{lpt}, N_{lp} \text{ integer } \geq 0, \qquad D_{ir} \geq 0 \tag{4.13}$$

The objective function (4.1) minimizes the distribution cost and the investment costs to acquire and adjust the trailer fleet. The first term indicates the total distribution cost over all depots and all periods. The second term captures the total "effective" purchase cost. Generally, a trailer is utilized for more than 10 years, so trailers may have a *salvage value* at the end of the planning horizon. Thus, we calculate the *effective purchase cost*, α'_{lp}, by deducting the present value of the salvage value at the end of planning horizon from the new purchase cost, α_{lp}. The third term captures the total rental cost and the fourth term is the relocation cost across all the depots and depots considered.

Constraint (4.2) ensures that the amounts delivered to customer i across all trips r meet the demand of customer i in period t. Constraint (4.3) ensures that the total amount delivered to all the customers on trip r in period t does not exceed the associated trailer capacity q_l multiplied by the number of times shift s is executed, or equals to zero if the shift is not selected. For each individual customer i on trip r, constraint (4.4) ensures that the total delivery amount is less than or equal to either the associated trailer capacity or the largest capacity which the customer can hold multiplied by the number of times shift s is selected. Since trip r is associated with shift s and trailer type l, the delivery amount is forced to zero if shift s is not selected or trailer type l is not allowed to visit customer i. Constraint (4.5) strengthens the model by indicating the minimum number of visits required to customer i in period t. The minimum number of visits is determined by the demand divided by either the largest trailer capacity or the maximum working capacity. Constraint (4.6) ensures that the total time required to execute all the selected shifts associated with trailer type l cannot exceed the available working time per trailer multiplied by the number of trailer type l allocated (either trailers owned by the company or rented) in depot p during period t. Similar to the homogeneous TFP formulation, this constraint provides a lower bound for the true fleet size. Details are presented in Section 3.3. A similar estimation was also applied in Larson [52] and Larson and Webb [53] when addressing strategic inventory routing problems.

Constraints (4.7) to (4.10) ensure the trailer types are relocated correctly among depots. Recall that the purchase decision is made at the beginning of planning horizon, while the relocation and rental decisions are planned for the beginning of individual time periods. Rental trailers can be used immediately, however relocated trailers may need to wait for approval to work at a different depot. Specifically, constraint (4.7) and (4.8) demonstrate the simple case when there is no waiting period. Period 1 considers the number of trailers previously allocated, the number of new purchased trailers, and, if applicable, the number of trailers that are transferred in and transferred out of the depot. After period 1, no purchase is considered. When the waiting period is greater than one, the in coming trailer capacities are lost in period 1(4.9). After the waiting period, Constraint (4.10) includes the trailers which are transferred from other depots (the decisions to transfer trailers out have to be made Δ_p periods before). Similar to constraint (4.9), constraint (4.11) shows that when the waiting time is longer than two periods, we only consider the number of trailers in the previous period and the number of trailers transferred out for periods $t = 2, \cdots, \Delta_p(\geq 2)$. Lastly, constraint (4.12) addresses the budget constraint for new purchased trailers.

This problem is more complicated than the homogenous TFP problem due to the following aspects:

(1) The problem size increases substantially with multiple trailer types. The shifts are differentiated not only by depots and periods but also by trailer types.

(2) There are interactions between depots and periods. Based on the assignment of cus-

tomers to depots, the shifts are relatively independent for each depot and each period. The relocation decisions of trailers, however, link the shift and trailer fleet decisions together. In addition, the rental and reallocation cost are generally much lower than the distribution cost, which may result in inefficient decisions when the model is terminated within a certain optimality gap. For instance, it may be suggested to move one trailer from depot A to B and one trailer from depot B to C instead of the more efficient movement from A to C, since the difference of these two relocation costs is negligible in the objective value.

(3) Customer-trailer restriction indicator, e_{il}, forces part of the delivery amount on a trip to be zero, which may result in "inefficient" delivery in practice. Trailer utilizations are also impacted.

Therefore, we need an effective and efficient approach to solve this problem.

4.4. Modified Two-Phase Approach

To solve the heterogeneous TFP problem within acceptable limits, we develop a solution approach that extends the approach for the homogeneous TFP. The problem is decomposed into two phases as illustrated in Figure 4.2. The first phase mainly focuses on generating candidate shifts and optimizing distribution cost for each period, while the second phase addresses the fleet purchase, relocation, and rental decisions. The main difference between this approach and the approach for homogeneous TFP is that, with multiple trailer types, we consider the allocation in the previous periods and tend to favor the use of existing trailer types. Therefore, the two-phase approach is modified as follows:

(1) The route generation is based on different trailer types (Phase I);

(2) A rolling horizon scheme is applied to consider *previous allocations* of trailers. *Retention factors* are updated in each period to adjust the cost of adding a trailer and the

benefit of removing a trailer (Phase I).

(3) The Phase II model is extended to solve fleet adjustment decisions for each trailer type (Phase II);

4.4.1 Phase I

The first phase concentrates on distribution costs, which are a significant component of the objective function (4.1). The Phase I model determines the number of trailers *required* for each depot and each period. Since the route selection and delivery amount decisions are independent of depots and periods, the distribution decisions can be separated by by depots and periods, which greatly reduces the complexity of the original problem. Thus, the Phase I model is designed to run for each period and each depot. For conciseness, we remove the indices of period t and depot p in the notation. Customer set I and shift set S are associated with a particular depot and period.

To generate candidate shifts, a routing algorithm is developed for each depot, period, and each trailer type. If a customer-trailer restriction exists, then the customer is not considered when shifts for the restricted trailers are generated. Hence, the indicator for customer-trailer restriction, e_{il}, is not needed in the model. Details of the route generation algorithm are provided in Section 4.4.3.

The Phase I model is a reduced version of the heterogeneous TFP model. The decisions include selecting shifts to execute in a time period and determining the delivery amount to each customer on each trip. Based on the selected shifts, the number of trailers of each type needed, V_l, is estimated. In Phase I we do not distinguish between trailers that are owned by the company and rented from the third party. A *fixed cost*, h_l, is applied to represent the cost to use one trailer of type l in one period. It can be calculated as a weighted rental cost or a weighted purchase cost. The updated parameters are summarized as follows.

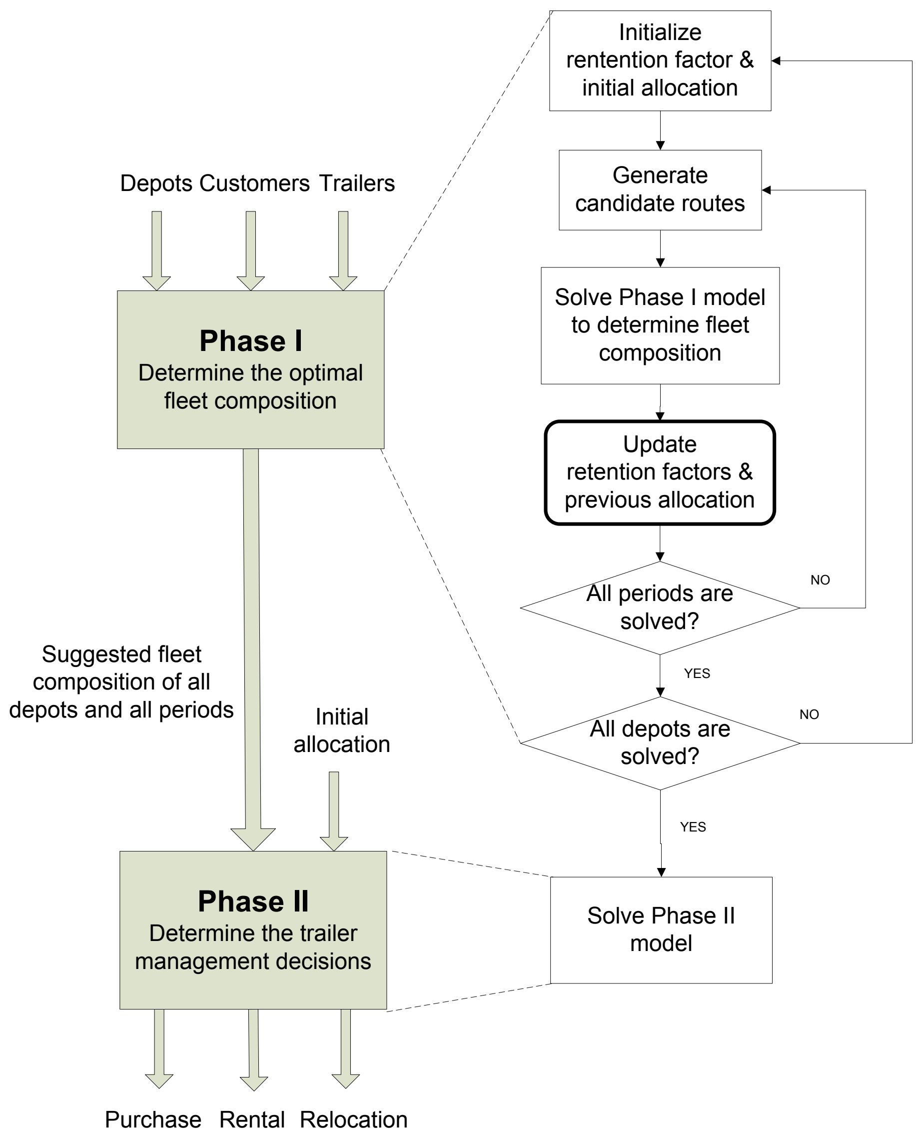

Figure 4.2: Two-Phase Approach for Heterogeneous Trailer Fleet Planning Problem

Table 4.8: Example of sub-optimal result without considering existing trailers

Trailer Type	**No. of Trailers Allocated**			
	Period 1	Period 2	Period 3	Period 4
T1	1	0	1	0
T2	0	1	0	1

Updated parameters:

I set of customers assigned to the depot

S set of shifts originated from the depot in the period

S^l subset of shifts operated by trailer type l

δ_i demand requirement for customer i

η available working time of a trailer at the depot in one period (exclude the maintenance time)

h_l fixed cost for adding one trailer of type l

Since the Phase I models consider each period independently, the existing trailers are not considered which may result in suboptimal. For instance, Table 4.8 shows the trailer allocations for four periods where each period is optimized individually. Trailer type T1 and T2 are suggested alternately, thus relocation or rental costs are required. However, when considering all periods together, an organization may prefer use one trailer type which may result in slightly higher distribution cost but lower fleet adjustment cost. Therefore, we include a retention factor to weight the value of the existing resources so that these trailer types can be prioritized. Since the Phase I models are solved in a rolling scheme, the priority of a trailer type is based on the frequency that it is acquired in the previous periods. In the Phase I model, the fleet composition at the end of the previous period, PA_l, and a retention factor, RF_l, are defined when suggesting the new fleet composition in the current period. Specifically, the retention factor has a value between 0 and 1 and indicates the frequency that trailer type l is used during previous periods. In this research, the retention factor for trailer type l, RF_l, is based on the frequency that it has been suggested in previous time periods. The retention factor can easily be modified based on other factors.

Information from previous periods:

PA_l the number of trailer type l allocated in the previous period

RF_l retention factor of trailer type l, which can be obtained by

$$RF_l = \frac{\text{No. of periods that used trailer type } l}{\text{No. of periods consided in total} + 1}$$

The decision variables and the Phase I model are presented below. With the allocated trailer fleet, we can either acquire more trailers V_l^+ or remove existing trailers V_l^-. Hence, the objective function (4.14) contains both the cost for adding trailers and the benefit of removing trailers. Recall that h_l is the cost of adding one trailer of type l, the retention factor RF_l is used together with fixed cost h_l to represent the *benefit* of removing a trailer. Specifically, we use $(1 - RF_l)h_l V_l^-$ to compute the benefit of removing a trailer type l from this depot. A larger RF_l means that this trailer type is used more often and less benefit is gained by removing a trailer of this type, so we tend to retain trailers of this type. Constraints (4.21) and (4.22) define the relationships between V_l, V_l^+, V_l^-, and PA_l.

Decision Variables:

Ψ_s the number of times shift s is executed

D_{ir} the amount of gas delivered to customer i on trip r

V_l the number of trailer type l allocated

V_l^+ the number of trailer type l added

V_l^- the number of trailer type l removed

Phase I model:

$$\text{Minimize} \quad \sum_{l\in L}\sum_{s\in S^l} c_s\Psi_s + \sum_{l\in L} h_l V_l^+ - \sum_{l\in L}(1-RF_{lp})h_l V_l^- \tag{4.14}$$

Subject to

$$\sum_{l\in L}\sum_{s\in S^l}\sum_{r\in R^s} D_{ir}y_{ir} = \delta_i, \qquad \forall i \in I, \tag{4.15}$$

$$D_{ir} \leq \min\{q_l, \sigma_i\} y_{ir}\Psi_s, \qquad \forall i \in I, \forall r \in R^s, \forall s \in S^l, \forall l \in L \tag{4.16}$$

$$\sum_{i\in I} D_{ir}y_{ir} \leq q_l\Psi_s, \qquad \forall r \in R^s, \forall s \in S^l, \forall l \in L \tag{4.17}$$

$$\sum_{l\in L}\sum_{s\in S^l}\sum_{r\in R^s} y_{ir}\Psi_{lst} \geq \max\left\{\left\lceil \frac{\delta_i}{\max\limits_{l\in L} q_l}\right\rceil, \left\lceil\frac{\delta_i}{\sigma_i}\right\rceil\right\}, \qquad \forall i \in I \tag{4.18}$$

$$\sum_{s\in S^l} \tau_s\Psi_s \leq \eta V_l, \qquad \forall l \in L \tag{4.19}$$

$$V_l^+ \leq V_l - PA_l, \qquad \forall l \in L \tag{4.20}$$

$$V_l^- = PA_l - V_l + V_l^+, \qquad \forall l \in L \tag{4.21}$$

$$\Psi_s \text{ integer } \geq 0, D_{ir} \geq 0 \tag{4.22}$$

4.4.2 Phase II

The Phase I model provides decisions for the fleet composition, V_l, for each depot and each period. After solving all periods and depots, we obtain the number of trailers of type l required for depot p and period t, V_{lpt}. The organization may use their own trailers or lease trailers from outside. Recall that trailers are purchased at the beginning of the planning horizon, and can also be reallocated or rented at the beginning of each time period.

The Phase II model is a straightforward extension of the Phase II model for the homogeneous TFP problem. The objective (4.23) includes the fleet purchase, rental, and relocation costs for all trailer types, and the relocation conservation constraints are extended for each trailer type. The key constraint (4.24) connects Phase I to Phase II and ensures that the optimal fleet is achieved by trailers owned or rented. As mentioned before, the Phase II objective is less than the objective value of the integrated heterogeneous TFP model, so the model can reach a much lower optimality gap and thus provide more meaningful solutions.

Phase II model:

$$\text{Minimize} \quad \sum_{p\in P}\sum_{l\in L}\alpha_l' N_{lp} + \sum_{t\in T}\sum_{p\in P}\sum_{l\in L}\beta_{lp} Z_{lpt} + \sum_{t\in T}\sum_{p\in P}\sum_{p'\in P}\sum_{l\in L}\gamma_{lp} d_{pp'} Y_{lpp't} \tag{4.23}$$

Subject to

$$X_{lpt} + Z_{lpt} \geq V_{lpt} \qquad \forall p \in P, \forall t \in T, \forall l \in L \tag{4.24}$$

$$\sum_{p\in P}\sum_{l\in L}\alpha_l N_{lp} \leq B \tag{4.25}$$

$$X_{lpt} = a_{lp} + N_{lp} + \sum_{p'\in P} Y_{lp'pt} - \sum_{p'\in P} Y_{lpp't} \qquad \forall p \in P : \Delta_p = 0, t = 1, \forall l \in L \tag{4.26}$$

$$X_{lpt} = X_{lp,t-1} + \sum_{p'\in P} Y_{lp'pt} - \sum_{p'\in P} Y_{lpp't}, \qquad \forall p \in P : \Delta_p = 0, t = 2, \cdots, N, \forall l \in L \tag{4.27}$$

$$X_{lpt} = a_{lp} + N_{lp} - \sum_{p'\in P} Y_{lpp't} \qquad \forall p \in P : \Delta_p \geq 1, t = 1, \forall l \in L \tag{4.28}$$

$$X_{lpt} = X_{lp,t-1} - \sum_{p'\in P} Y_{lpp't}, \qquad \forall p \in P : \Delta_p \geq 2, t = 2, \cdots, \Delta_p, \forall l \in L \tag{4.29}$$

$$X_{lpt} = X_{lp,t-1} + \sum_{p'\in P} Y_{lp'p,t-\Delta_p} - \sum_{p'\in P} Y_{lpp't}, \qquad \forall p \in P : \Delta_p \geq 1, t = \Delta_p + 1, \cdots, N, \forall l \in L \tag{4.30}$$

$$X_{lpt}, Y_{lpp't}, Z_{lpt}, N_{lp} \text{ integer } \geq 0 \qquad \forall p \in P, p' \in P, t \in T, \forall l \in L \tag{4.31}$$

The main advantage of the two-phase approach is that the distribution decisions and the fleet planning decisions are considered separately, which greatly reduces the complexity of the integrated model and guarantees the quality of solutions at different levels.

4.4.3 Modified Sweep-based Algorithm

An important issue impacting Phase I is the substantial number of possible delivery shifts. Since it is not practical to enumerate all the candidate shifts, we want to limit the number of shifts and focus on the most promising ones. This section modifies the sweep-based heuristic algorithm which is introduced in Chapter 3 so that the continuous delivery practice is addressed.

The basic idea underlying the algorithm is to make a trailer dedicated to serve each customer so that the "majority" of demands are satisfied and the "remaining" demands are clustered so that each cluster is served on one route. Since all customers are assigned to a particular source, the location of each customer i is represented by polar coordinates (θ_i, ρ_i) with the assigned source as the center, θ_i as the angle, and ρ_i as the radius. All the customers are sorted in non-decreasing order of θ. If two customers have the same θ-value, then they are ordered by non-decreasing ρ values.

In this algorithm, the shifts are generated by trailer types since (1) trailer capacity directly impacts the carrying amount on a trip and (2) some trailer types are restricted at some customer sites.

When a customer has demands larger than the trailer size, the customer can be served by either dedicated routes where only the customer is involved, or by multiple-customers routes where other customers are also served. The dedicated routes are straightforward to obtain, while multiple-customers routes are generated based on normalized customer demands. Specifically, all demands δ_i are normalized by *full load drops* (fld_i) and *partial load drops* (pld_i), where:

$$fld_i = \left\lfloor \frac{\delta_i}{q_l} \right\rfloor, \text{ and } pld_i = \frac{\delta_i}{q_l} - fld_i \tag{4.32}$$

This algorithm tends combine the customers that are close in proximity based on their associated pld_is so that the trailers are utilized effectively. Note that pld_i is only used to generate potential routes and the actual delivery amounts are determined in the Phase I model.

The routing algorithm is summarized which generates shifts for trailer type l starting from source a. Let N be the number of customers assigned to the source and k be the index for shifts. Each customer can lead the sweep procedure, which is indicated by i_{start}.

Step 0: (Initialize) Set $i_{start} = 1$ and $k = 1$.

Step 1: (Prepare customer data)

Step 1.1: (Extract) Extract customers which are assigned to source a and can accomodate trailer type l on site;
Step 1.2: (Normalize) Calculate pld_i for each customer i by Equation (4.32);
Step 1.3: (Sort) Let the source be the origin and calculate polar coordinates of each customer. Sort all customers in non-decreasing order of the polar angles θ. If two customers have the same θ-value, then order them according to non-decreasing values of ρ.

Step 2: (Shift start) Consider customers from the i_{start} position on the ordered list. Construct shift k in the clockwise direction. That is, consider the list as a circle and go through the list in the clockwise direction.

Step 3: (Sweep) Add customers one by one as long as both the capacity and maximum shift time are not exceeded. If the total shift time is exceeded, complete shift k, set $k \leftarrow k + 1$, and go back to Step 2 to start a new shift. If the remaining capacity is below some threshold, go to Step 4 to refill. If the remaining capacity is above the refill threshold but can not serve the next customer, skip the next customer and consider the following customer on the ordered list.

Step 4: (Refill) Find the nearest refill source from the current location. Add the source if the maximum shift time is not violated and go to step 3 to continue adding customers. Otherwise, complete shift k, set $k \leftarrow k + 1$, and go back to Step 2 to start a new shift.

Step 5: (Change leading customer) If $i_{start} < N$, set $i_{start} \leftarrow i_{start} + 1$ and go back to Step 2 starting from a new customer.

The algorithm is executed for each source and each trailer type. Note that in Step 4, the refill source may or may not be the same as the origin source, which depends on the current location of the vehicle. If the trailer goes back to the same source, then we can continue adding customers based on the original ordered list. However, if the vehicle travels to a different source, then we need to go through a new ordered list involving customers who have assigned to that new source.

In the algorithm shown above, customers are ordered and added in clockwise direction. To provide additional candidate shifts, the customers can also be traversed in a counterclockwise direction. However, a huge number of potential shifts would make the model too difficult to solve. Therefore, it is helpful to remove the shifts which are impractical and focus on the shifts which are promising. As shown above, the maximum shift time is considered as a primary limitation for the size of shift. Additional reduction techniques include the following:

- Limit the trip size. Due to the considerations of the unloading time and the driver's workload, we restrict the number of customers consecutively visited on a trip (between refills).
- Limit the number of refills. Too many refills would generate a long shift which may be impractical for actual delivery, so the maximum number of refills in a shift is applied.
- Limit the number of skipped customers. In Step 3, when the remaining capacity cannot serve the next customer but is still above the refill threshold, the vehicle would skip the customer and consider the following ones. However, if too many customers are skipped,

Table 4.11: General information

Depot ID	No. of Customers	Trailers Allocated
Depot 1	92	2 LARGE
Depot 2	167	2 MEDIUM 1 SMALL
Depot 3	198	2 LARGE 2 SMALL

a long trip may be obtained where customers are not close in proximity. Therefore, we also restrict the number of customers that we skip consecutively.

- Limit the search range for customers after refilling at a different source. In Step 4, after refilling at a new source, the vehicle may be more likely to go towards the depot due to the work time limits. So we limit the searching angle in the general direction towards the depot (details are presented in Appendix C).

4.5. Numerical Studies

We evaluate the effectiveness and efficiency of the proposed two-phase approach using actual data from Air Liquide. The solution approach has been implemented in Visual Studio 2008 (C#) and OPL 6.3 / CPLEX 12.3 with the data organized using Excel. All experiments were conducted on a 64-bit Intel Xeon X5570 Workstation (CPU 2.93GHz, 8 cores, RAM 24G).

Data for 475 customers at three depots are used in this analysis. As shown in Table 4.11, Depot 1 contains 98 customers and Depots 2 and 3 contain more than 100 customers. Customer demands are provided for four periods and the average periodic demands varying from 500 kg to more than 3,800,000 kg.

Three trailer types are used in this study. They are classified as SMALL, MEDIUM, and LARGE based on the capacities. Air Liquide provided information about the associated

Table 4.12: Trailer information

Trailer Type	Purchase Cost (USD)	Salvage Value (USD)	Rental Cost (USD/period)	Relocation Cost (USD/mile)
LARGE	350,000	280,000	20,000	2
MEDIUM	320,000	260,000	18,000	2
SMALL	300,000	250,000	15,000	2

costs[1] for trailer purchase, rental, and relocation, which are summarized in Table 4.12. The current allocation of trailers are presented in Table 4.11. We also consider the trailer availability due to maintenance, i.e. percentage of time that each trailer can work in a time period.

Table 4.13 presents the results where no trailers are currently allocated (green field case). The distribution costs for each depot and period are provided based on the output of Phase I, and the number of trailers of each type are also determined. In Phase II, the effective purchase cost, relocation cost, and rental cost are determined. In this case, eight trailers purchased by the company are allocated across the three depots. A total of \$470,000 is required for new trailers (deducting the salvage value at the end of the planning horizon from the new purchase cost). In addition, a MEDIUM trailer is rented for depot 1 for one period and \$18,000 is required for this rental.

For regions where trailer fleets already exist, this TFP approach can be also used to support fleet re-allocation decision. To show the benefit of changing the fleet, the distribution costs for the current trailer allocation are estimated. The new allocated fleet, V_l, in the Phase I model is *fixed* for any depot p and period t as the initial allocation a_{lp} . This case is referred as the *fixed mode*, while the *optimization mode* implies the initial allocation can be changed. Table 4.14 shows the results of the fixed mode and optimization mode for the three depot

[1] All the costs in this section are converted to U.S. Dollar with latest currency exchange rate.

Table 4.13: Decision support for green field

	Depot 1	Depot 2	Depot 3
Distribution Costs (USD)	1,037,210	1,535,666	1,848,143
Trailers allocated	1 LARGE 1 SMALL	1 LARGE 1 SMALL	1 LARGE 1 MEDIUM 2 SMALL
Trailers rented	1 MEDIUM	-	-
Effective purchase cost (USD)[a]	470,000		
Relocation cost(USD)	-		
Rental cost(USD)	18,000		
Total cost (USD)	4,909,019		

[a] Effective purchase cost = New purchase cost - present value of the salvage value at the end of the planning horizon.

data set. All three depots achieve better distribution costs by allowing the initial fleet to be optimized. Specifically, distribution costs are reduced by 3.81%, 5.04%, and 2.90% for the three depots respectively.

When it comes to fleet adjustment costs, we need to move a LARGE trailer from Depot 1 to 2, a MEDIUM trailer from Depot 2 to 3, and a SMALL trailer from Depot 2 to 1. In total, a SMALL trailer is saved in Depot 3 and only $2,696 is incurred for trailer relocation in the optimization mode. Considering the total cost, the optimization mode provides a solution which is $ 173,579 (3.78%) less than the fixed mode. Therefore, it is worthwhile to make this trailer fleet adjustment.

4.5.1 Two-phase approach compared to integrated model

To compare the two-phase approach and the integrated model, we use both methods to solve one depot, two depots, and three depots data sets based on generated candidate shifts.

Table 4.14: Decision support for trailer re-allocation case

Statistics		**Fixed mode**	**Optimization mode**	**Saving**
Distribution cost (USD)	Depot 1	1,077,556	1,036,488	3.81%
	Depot 2	1,616,348	1,534,939	5.04%
	Depot 3	1,902,556	1,847,410	2.90%
Trailer allocation	Depot 1	2 Large	1 Large 1 Small	
	Depot 2	2 Medium 1 Small	1 Large 1 Medium	
	Depot 3	2 Large 2 Small	2 Large 1 Medium 1 Small	
Effective purchase cost[a](USD)		-	-	
Relocation cost(USD)		-	2,696	
Rental cost(USD)		-	-	
Total cost(USD)		4,596,460	4,421,533	3.81%

[a] Effective purchase cost = New purchase cost - Present value of the salvage value at the end of planning horizon.

Table 4.15: Comparison between integrated model and two-phase approach (MIP gap 2.0%)

Case	**Objective value (USD)**			**Running time (mins)**	
	Integrated	Two-phase[a]	Difference	Integrated	Two-phase
1 Depot - 1 Period	260,534	263,904	1.29%	0.03	0.03
1 Depot - 2 Periods	522,389	528,592	1.19%	0.07	0.06
1 Depot - 3 Periods	785,486	793,150	0.98%	0.09	0.08
1 Depot - 4 Periods	1,040,641	1,054,469	1.33%	0.10	0.10
2 Depots - 1 Period	647,504	652,679	0.80%	0.14	0.12
2 Depots - 2 Periods	1,304,563	1,300,492	0.31%	2.41	0.24
2 Depots - 3 Periods	1,952,812	1,951,894	0.05%	5.39	0.35
2 Depots - 4 Periods	2,604,689	2,596,656	0.31%	10.45	0.46
3 Depots - 1 Period	1,116,678	1,116,229	0.04%	1.01	0.83
3 Depots - 2 Periods	2,244,248	2,227,686	0.74%	8.92	1.90
3 Depots - 3 Periods	3,348,671	3,336,865	0.35%	24.96	2.25
3 Depots - 4 Periods	4,557,509	4,443,272	2.51%	> 1440(2.37%)	2.54

[a] Objective value of two-phase approach = distribution cost in Phase I + objective in Phase II

To evaluate the impact of the length of planning horizon, the time periods are also changed from one period to four periods. Table 4.15 compares the objective values and computational times from the integrated model and the two-phase heuristic. The objective value of two-phase approach includes the distribution costs in Phase I and trailer management costs in Phase II. All the MIP gap tolerances are set to 2%.

As shown in Table 4.15, both the integrated model and two-phase approach solve the cases with one depot very quickly, since no relocation is considered. When two depots are considered, the interactions between the depots make the integrated model more difficult to solve. The computational times increase substantially from two minutes to over ten minutes when the planning horizon increases from two periods to four periods. On the other hand, the two-phase approach provides more robust performance by removing the interactions between depots in Phase I. The cases with two depots take less than one minute to solve and the objective values are very close to the integrated model (less than 1%). The results of three depots further demonstrate the advantages of the two-phase approach. Considering three depots and three periods, the integrated model requires over 20 minutes, while the two-phase approach solves the same data set in less than three minutes and provides objective values within 0.35% of the integrated model. As the number of time periods increases to four, the integrated model cannot reach 2% of optimality gap after 24 hours. In contrast, the two-phase approach solves within three minutes.

When the MIP gap is further reduced to 1%, five out of the twelve cases above cannot be solved in 24 hours by the integrated model. However, the two-phase approach solves the largest data set, including three depots and four periods, in approximately eight minutes. Detailed results are provided in Table 4.16.

4.5.2 Effect of multiple trailer types

Lastly, we are interested in the comparison between heterogeneous fleet and homogeneous fleet. To compare the results of three trailer types, we create three homogeneous TFP cases by modifying Depot 2 to include only SMALL, MEDIUM, or LARGE trailer type. Table 4.17 presents the trailer allocation, distribution cost, trailer effective purchase cost, and total cost for these cases. When considering only one trailer type, LARGE trailers provide the most efficient service resulting in the lowest distribution cost. Although the cost for each LARGE trailer is the highest, the number of trailers is saved by one and the total trailer investment is still better than the SMALL and MEDIUM cases.

As shown in Table 4.17, the heterogeneous fleet case reduces all the cost components even more compared to the LARGE trailer case. Specifically, the combination of one LARGE trailer and one SMALL trailer saves $72,441 in distribution cost and $20,000 in trailer in-

Table 4.16: Comparison between two-phase approach and complete model (MIP gap 1.0%)

Case	Objective value (USD)			Running time (mins)	
	Integrated	Two-phase[a]	Difference	Integrated	Two-phase
1 Depot - 1 Period	258,830	262,250	1.32%	0.03	0.03
1 Depot - 2 Periods	518,422	524,925	1.65%	0.13	0.12
1 Depot - 3 Periods	777,476	787,765	1.32%	0.40	0.21
1 Depot - 4 Periods	1,036,830	1,050,488	1.32%	1.43	0.25
2 Depots - 1 Period	644,722	649,516	0.74%	0.87	0.12
2 Depots - 2 Periods	1,292,413	1,294,772	0.18%	6.23	1.12
2 Depots - 3 Periods	1,947,554	1,937,426	0.52%	> 1440(1.62%)	1.91
2 Depots - 4 Periods	2,603,491	2,580,622	0.88%	> 1440(1.82%)	2.52
3 Depots - 1 Period	1,105,784	1,114,446	0.78%	4.05	0.83
3 Depots - 2 Periods	2,232,449	2,223,122	0.42%	> 1440(1.30%)	4.91
3 Depots - 3 Periods	3,337,559	3,327,620	0.30%	> 1440(1.46%)	7.31
3 Depots - 4 Periods	4,557,509	4,432,680	2.74%	> 1440(2.37%)	8.66

[a] Objective value of two-phase approach = distribution cost in Phase I + objective in Phase II

Table 4.17: Impact of multiple trailer types

Statistics	SMALL	MEDIUM	LARGE	Heterogeneous
Trailer Allocation	3 SMALL	3 MEDIUM	2 LARGE	1 LARGE 1 SMALL
Distribution Cost (USD)	1,696,255	1,687,096	1,611,677	1,539,236
Effective Purchase Cost[a] (USD)	150,000	180,000	140,000	120,000
Total Cost (USD)	1,846,255	1,867,096	1,751,677	1,659,236
No. of Shifts	1,514	1,405	1,471	4,390
Computational Time (mins)	0.06	0.08	0.08	2.46

[a] Effective purchase cost = New purchase cost - present value of the salvage value at the end of planning horizon.

vestment. This distribution cost is obtained from the shifts generated for each trailer type, hence more candidate shifts are considered with heterogeneous fleet. The computational time is increased to 2.46 minutes, which is still reasonable for a depot with 167 customers.

4.6. Conclusions

A heterogeneous fleet generally provides more flexible and cost-effective service for an organizaiton. This paper explores a heterogeneous trailer fleet planning problem where multiple trailers with various capacities and costs are considered. In addition, customers may have restrictions on the trailer types that can access sites. In this problem, both the long-term fleet acquirement decisions and the medium-term vehicle relocation and rental decisions are addressed. The objective is to minimize the total distribution costs and fleet investment costs over multiple depots and multiple periods.

The problem is formulated as a mixed-integer program for multiple depots and multiple periods based on sets of potential gas delivery routes. However, the relationship between depots and periods makes the problem very complicated to solve. Therefore, a modified two-phase approach is proposed where Phase I concentrates on distribution costs and provides the suggested fleet composition, and Phase II addresses the trailer acquisition decisions. This decomposition approach removes the interactions between depots and periods in Phase I, which greatly reduces the complexity of the original problem. Phase II provides aggregate plans based on outputs from Phase I.

Specifically, in Phase I a reduced model is solved for each period and depot based on a set of potential shifts. The trailer allocation in the previous period and the frequency that each trailer type has been used are also considered so that the trailer types that are used for longer time are prioritized. To generate candidate shifts of high quality, a modified sweep-based heuristic is developed to capture the characteristics of the continuous delivery practice in the industrial gas distribution.

The numerical studies, conducted using a data set of three depots and three trailer types, demonstrate the effectiveness of the approach. This approach successfully provides decisions support for both allocating trailers in a green field scenario, and reallocating trailers for regions where a fleet exists. Compared to the integrated optimization model, the decomposition approach obtains good-quality solutions in a reasonable computational time. Additionally, the solution approach demonstrates a more robust performance as the problem size increases. Finally, we compare the results of heterogeneous fleet and homogeneous fleet, and analyze the effect of multiple trailer types. The comparison shows that heterogeneous TFP problem suggests more efficient solutions but requires longer computational time. Air Liquide is currently integrating the proposed solution approach in their worldwide distribution planning software.

The proposed two-phase approach provides a very good scheme that combines constructive heuristics and mathematical programs. Since the number and quality of candidate routes play significant role in this approach, it is worthwhile to apply different route generation algorithms.

Chapter 5

Conclusions and Future Work

Rail and truck are two of the most important modes of for freight transportation today. Rail is generally used to transport heavy and bulky commodities over long distance, while trucks provide fast and flexible service for small and high-value products. In this dissertation, we have studied two different distribution planning problems that arise in rail and truck transportation systems:

1. Time and capacity constrained routing problem in railroad planning, and
2. Trailer fleet planning problem in industrial gas distribution

In this chapter, we review the main contributions and discuss areas for future work.

5.1. Railroad Trip Planning

In the railroad industry, three optimization problems address important tactical decisions: (1) the blocking problem, which determines how to aggregate shipments to reduce the impact of reclassification; (2) the train scheduling problem, which determines train origins, destinations, routes, and weekly schedules; and (3) the block-to-train (BTA) problem, which determines the train paths of blocks. Based on these tactical results, the time and capacity constrained routing (TCCR) problem assigns multiple shipments to blocks and train-runs so that the total transportation cost is minimized while train capacities are not exceeded and the due dates of the shipments are met.

The TCCR problem is defined on a time-space-train-block (TS-TB) network which includes the information about train routes, time schedules, and blocks. Based on the TS-TB network, two integer programming models are formulated. The first arc-based formulation is a conventional network flow model, which selects arcs in the TS-TB network so as to find a connected path for each shipment from its origin-node to the destination-node while satisfying all side constraints. The second path-based formulation considers a set of feasible paths with respect to the TS-TB network and decides the optimal path for each shipment.

Due to the large size of TCCR formulations, two heuristic algorithms are proposed to solve the problem more efficiently. The sequential algorithm sends shipments one by one based on priority and each shipment is guaranteed to meet its due date and not to violate the existing train capacities. In this approach, Yen's algorithm is used to determine the K shortest train-blocking paths in the train-block network which only contains associated costs, and then the fastest train-run path is determined to meet the due time by only considering the trips which have enough remaining capacity. The sequential algorithm solves the TCCR problem primarily based on a relatively small network, but involves many repetitive operations when solving for a large number of shipments. Therefore, a bump-shipment algorithm is proposed to further improve the computational time. In this algorithm, all shipments are scheduled without considering train capacity limits, and then shipments with lower priorities are bumped and rearranged so as to satisfy the capacity constraints.

In the numerical study, both the sequential and bump-shipment algorithms solve the TCCR problem in a reasonable amount of computational time for representative data with 500,000 shipments provided by a major U.S. railroad. Moreover, the dynamic trip plans solved by our algorithms demonstrate greater flexibility to send shipments and result in better performance for every key cost factor than the static trip plan currently used by railroads. On average, the total transportation cost is reduced by 12% for shipments which are assigned to a different trip plan than the static plan. To gain more insights of the dynamic trip planning problem, we further analyze the impact of distance factor and capacity limits. Larger distance factors reduce travel miles but increase station costs, while larger train capacities decrease both the travel time and total cost.

5.2. Trailer Fleet Planning

Trailer fleet planning (TFP) is an important issue for industrial gas distribution. In this research, we consider both the long-term fleet acquisition decisions and the medium-term vehicle relocation and rental decisions. The goal is to minimize the total distribution costs and fleet management costs across multiple depots and multiple periods with time varying demands. In industry, a homogeneous fleet is generally easier to manage, but a heterogeneous fleet provides more flexible and cost-effective service for demand variations. Therefore, both the homogeneous TFP and heterogeneous TFP problems are investigated in this dissertation.

For the homogeneous TFP problem, a mixed-integer program is developed for multiple depots and multiple periods. The objective is to minimize the total distribution costs and associated trailer purchase, rental, and relocation costs. To estimate the distribution cost, we assume that a set of candidate replenishment routes is given for each period and each depot so that the model selects optimal routes and determines delivery amounts. The models are difficult to solve due to the large number of decision variables and the interaction between depots and periods. Therefore, a two-phase approach is proposed to solve the problem efficiently. Phase I concentrates on distribution costs and provides the suggested fleet size, while Phase II addresses the fleet purchase, relocation, and rental decisions. Specifically, to generate candidate routes of good quality, a sweep-based heuristic is applied in Phase I. Then a reduced model is used to select routes for meeting customer demands and determine the preferred fleet size. Phase II provides aggregate plans based on outputs from Phase I. This decomposition approach removes the interaction between depots and periods in Phase I, which greatly reduces the complexity of the original problem.

The heterogeneous TFP problem is more complicated than the homogeneous problem. In the integer programming model, various trailer capacities and associated costs are introduced and candidate routes for each trailer type are assumed. The two-phase approach is modified, where a rolling horizon scheme is applied in Phase I. When determining fleet compositions for the current period, trailer allocations in previous periods are considered so that existing trailer types take advantage. If a trailer type is used more frequently in previous periods, it is more likely to be retained. Additionally, the sweep-based heuristic is improved to capture the characteristics of continuous delivery where trailers are allowed to refill products during a route. This heuristic generates routes for each trailer type, hence the customer-trailer restrictions are considered.

The numerical studies, conducted using a data set with three depots and more than 400 customers, demonstrate the effectiveness of the two-phase approaches. The approaches successfully support decisions for both allocating trailers in green field where no fleet exists, and reallocating trailers for regions where fleets exist. Compared to the integrated optimization models, the decomposition approaches obtain good-quality solutions in a reasonable computational time. As the problem size increases, the two-phase approach demonstrates more robust performance. To gain more insights, the impact of trip size limit and the effect of multiple trailer types are analyzed. Results show that longer trip sizes increase the number of candidate routes and thus provides reduced distribution cost. A heterogeneous trailer fleet provides more efficient service than a homogeneous fleet but the solution approach requires more computational time. Based on these results, Air Liquide is currently integrating the proposed solution approaches in their worldwide distribution planning software.

5.3. Future Work

Potential future work includes extending the TCCR problem and the TFP problem and addressing intermodal transportation that combines rail and truck.

5.3.1 Relaxation of existing tactical plans

In the TCCR problem, the blocking plan, the train schedule, and the BTA are all considered as given. This assumption restricts the size of the trip planning problem for planning purposes. In practice, however, railroad companies also desire the capability to adjust their current blocking plan, train schedules, or BTA if appropriate to save costs. Therefore, an area for future research is considering the situation when these pre-determined plans can be partially changed. This problem provides greater flexibility to transport shipments but is also more complex than the problem with fixed tactical plans.

5.3.2 Extension for dynamic trip planning problem

The TCCR problem is a special version of dynamic trip planning problem, which addresses shipment due time and train capacity constraints. In reality, additional constraints can to be considered when transporting shipments, such as block volume and shipment width, height, and weight restrictions. Also, some shipments may have priority on the existing blocks. Hence, other constraints could be explored in the dynamic trip planning problem.

5.3.3 Exploration of different routing algorithms

The two-phase approach developed in the TFP problem effectively combines constructive heuristics and mathematical programs. Since the number and quality of candidate routes play a significant role in this approach, it is worthwhile to consider different route generation algorithms and evaluate their impact on the total performance. For instance, the saving heuristics and tabu meta-heuristics that are considered successful for solving VRP may be explored for generating candidate routes.

5.3.4 Integration of tractors and drivers

The TFP problem primarily considers the vehicle as a single unit. In truck transportation system, however, it is common to use tractor-trailer combination, where the cargo-carrying capacity depends on the size of the trailer and tractors can be swapped between different trailers quickly. In addition, driver is another important factor when scheduling the vehicles. There are many rules and regulations in place to govern the use of tractors and drivers. An area of future research is integrating tractors and drivers issues into the TFP problem. The problem aims to recommend the number of tractors and drivers together with the optimal trailer fleet for each depot and each period, while satisfying both the driver work rules and restrictions at customer sites.

5.3.5 Intermodal rail and truck transportation

Intermodal transportation of rail and trucks grew rapidly in the 1980s. Generally, intermodal rail-truck service includes one or more trucks providing the short-haul pickup and delivery service part of the trip and one or more railroads providing the long-haul or line-haul service. This intermodal transportation combines the door-to-door convenience of trucks with the high-volume, long-haul economies of railroads. Therefore, compared with either individual mode, the intermodal service offers the potential to reduce transit time and lower rates. The distribution problems related to intermodal transportation are the promising research areas in the future.

Bibliography

[1] Association of American Railroads. 1997. *Railroad Facts*. Washington, D.C.

[2] Eno Transportation Foundation. 1998. *Transportation in America, 16th ed.*. Washington, D.C.

[3] Association of American Railroads. 2006. *Railroad Facts*. Washington, D.C.

[4] American Association of State Highway and Transportation Officials. 2007. *Transportation-Invest in Our Future: Future needs of the U.S. Surface Transportation System*. Washington, D.C.

[5] Aberdeen Group. 2007 March. *Service on the Move: Driving Profitability via Fleet Management*. Boston, MA.

[6] The Council of Supply Chain Management Professionals. 2010 July. *21st Annual State of Logistics Report*. Washington, D.C.

[7] Association of American Railroads. 2011. *An Overview of America's Freight Railroads*. Washington, D.C.

[8] Ahuja, R. K., K. C. Jha, J. Liu. 2007. Solving real-life railroad blocking problems. *Interfaces* **37**(5) 404–419.

[9] Ahuja, R. K., J. Liu, J. B. Orlin, D. Sharma, L. A. Shughart. 2005. Solving real-life locomotive-scheduling problems. *Transportation Science* **39**(4) 503–517.

[10] Ahuja, R. K., J. B. Orlin, S. Pallottino, M. G. Scutell. 2003. Dynamic shortest paths minimizing travel times and costs. *Networks* **41**(4) 197–205.

[11] Araque, J. R., G. Kudva, T. L. Morin, J. F. Pekny. 1994. A branch-and-cut algorithm for vehicle routing problems. *Annals of Operations Research* **50**(1) 37–59.

[12] Archetti, C., M. Speranza, M. Savelsbergh. 2008. An optimization-based heuristic for the split delivery vehicle routing problem. *Transportation Science* **42**(1) 22–31.

[13] Avella, P., B. D'Auria, S. Salerno. 2006. A LP-based heuristic for a time-constrained routing problem. *European Journal of Operational Research* **173**(1) 120–124.

[14] Baldacci, R., M. Battarra, D. Vigo. 2009. Valid inequalities for the fleet size and mix vehicle routing problem with fixed costs. *Networks* **54**(4) 178–189.

[15] Ballou, R. H. 1999. *Business Logistics Management: Planning, Organizing and Controlling the Supply Chain*, chap. 6. Prentice Hall, 135–184.

[16] Bard, J. F., L. Huang, M. Dror, P. Jaillet. 1998. A branch and cut algorithm for the vrp with satellite facilities. *IIE Transactions* **30**(9) 821–834.

[17] Bard, J. F., L. Huang, P. Jaillet, M. Dror. 1998. A decomposition approach to the inventory routing problem with satellite facilities. *Transportation Science* **32**(2) 189–203.

[18] Barnhart, C., C. A. Hane, P. H. Vance. 2000. Using branch-and-price-and-cut to solve origin-destination integer multicommodity flow problems. *Operations Research* **48**(2) 318–326.

[19] Beltrami, E. J., L. D. Bodin. 1974. Networks and vehicle routing for municipal waste collection. *Networks* **4**(1) 65–94.

[20] Bodin, L. D., B. L. Golden, A. D. Schuster, W. Romig. 1980. A model for the blocking of trains. *Transportation Research Part B: Methodological* **14**(1-2) 115–120.

[21] Brandao, J., A. Mercer. 1998. The multi-trip vehicle routing problem. *The Journal of the Operational Research Society* **49**(8) 799–805.

[22] Brandäo, J. 2009. A deterministic tabu search algorithm for the fleet size and mix vehicle routing problem. *European Journal of Operational Research* **195**(3) 716–728.

[23] Brännlund, U., P. Lindberg, A. Nou, J. Nilsson. 1998. Railway timetabling using lagrangian relaxation. *Transportation Science* **32**(4) 358–369.

[24] Campbell, A. M., M. W. Savelsbergh. 2004. A decomposition approach for the inventory-routing problem. *Transportation Science* **38**(4) 488–502.

[25] Caprara, A., M. Fischetti, P. Toth. 2002. Modeling and solving the train timetabling problem. *Operations Research* **50**(5) 851–861.

[26] Chao, I.-M., B. L. Golden, E. Wasil. 1995. An improved heuristic for the period vehicle routing problem. *Networks* **26**(1) 25–44.

[27] Choi, E., D.-W. Tcha. 2007. A column generation approach to the heterogeneous fleet vehicle routing problem. *Comput. Oper. Res.* **34**(7) 2080–2095.

[28] Clarke, G., J. Wright. 1964. Scheduling of vehicles from a central depot to a number of delivery points. *Operations Researchs* **12**(4) 568–581.

[29] Coyle, J. J., E. J. Bardi, R. A. Novack. 2000. *Transporation 5e*, chap. 3. South-Western College Publishing, 109.

[30] Desrochers, M., T. W. Verhoog. 1991. A new heuristic for the fleet size and mix vehicle routing problem. *Computers & Operations Research* **18**(3) 263–274.

[31] Dessouky, M., Q. Lu, J. Zhao, R. Leachman. 2006. An exact solution procedure to determine the optimal dispatching times for complex rail networks. *IIE Transactions* **38**(2) 141–152.

[32] Dijkstra, E. 1959. A note on two problems in connexion with graphs. *Numerische Mathematik* **1** 269–271.

[33] Dror, M., G. Laporte, O. Trudeau. 1994. Vehicle routing with split deliveries. *Discrete Applied Mathematics* **50**(3) 239–254.

[34] Fu, L., D. Sun, L. R. Rilett. 2006. Heuristic shortest path algorithms for transportation applications: state of the art. *Computers and Operations Research* **33**(11) 3324–3343.

[35] Garvin, W. W., H. W. Crandall, J. B. John, R. A. Spellman. 1957. Applications of vehicle routing in the oil industry. *Management Science* **3**(1) 407–430.

[36] Gaudioso, M., G. Paletta. 1992. A heuristic for the periodic vehicle routing problem. *Transportation Science* **26**(2) 86–92.

[37] Gendreau, M., A. Hertz, G. Laporte. 1992. New insertion and postoptimization procedures for the traveling salesman problem. *Operations Research* **40**(6) 1086–1094.

[38] Gendreau, M., G. Laporte, C. Musaraganyi, E. D. Taillard. 1999. A tabu search heuristic for the heterogenous fleet vehicle routing problem. *Computer Operations Research* **26**(12) 1153–1173.

[39] Ghoseiri, K., F. Morshedsolouk. 2006. ACS-TS: train scheduling using ant colony system. *Journal of Applied Mathematics and Decision Sciences* **2006** 1–28.

[40] Ghoseiri, K., F. Szidarovszky, M. Asgharpour. 2004. A multi-objective train scheduling model and solution. *Transportation Research Part B: Methodological* **38**(10) 927–952.

[41] Gillett, B., L. Miller. 1974. A heuristic algorithm for the vehicle dispatch problem. *Operations Researchs* **22**(2) 340–349.

[42] Golden, B. L., E. A. Wasil, J. P. Kelly, I. M. Chao. 1998. *Fleet Management and Logistics*, chap. Metaheuristics in vehicle routing. Kluwer, Boston, MA.

[43] Golden, B. L., A. A. Assad, L. Levy, F. Gheysens. 1984. The fleet size and mix vehicle routing problem. *Computers & Operations Research* **11**(1) 49–66.

[44] Golden, B. L., S. Raghavan. 2008. *The Vehicle Routing Problem: The Latest Advances and New Challenges*, *Operations Research/Computer Science Interfaces Series*, vol. 43. Springer, New York.

[45] Jha, K., R. Ahuja, G. Şahin. 2008. New approaches for solving the block-to-train assignment problem. *Networks* **51**(1) 48–62.

[46] Keaton, M. 1992. Designing railroad operating plans: A dual adjustment method for implementing Lagrangian relaxation. *Transportation Science* **26**(4) 263–279.

[47] Kontoravdis, G., J. F. Bard. 1995. A GRASP for the Vehicle Routing Problem with Time Windows. *ORSA Journal on Computing* **7**(1) 10–23.

[48] Kwon, O., C. Martland, J. Sussman. 1998. Routing and scheduling temporal and heterogeneous freight car traffic on rail networks. *Transportation Research Part E: Logistics and Transportation Review* **34**(2) 101–115.

[49] Laporte, G., M. Gendreau, J. Y. Potvin, F. Semet. 2000. Classical and modern heuristics for the vehicle routing problem. *International Transactions in Operational Research* **7**(4-5) 285–300.

[50] Laporte, G., Y. Nobert, M. Desrochers. 1985. Optimal routing under capacity and distance restrictions. *Operations Research* **33**(1) 1058–1073.

[51] Laporte, G., F. Semet. 2000. *The Vehicle Routing Problem*, chap. Classical heuristics for the capacitated VRP. SIAM, Philadelphia, PA.

[52] Larson, R. C. 1988. Transportation sludge to the 106-mile site - an inventory routing model for fleet sizing and logistics system design. *Transportation Science* **22**(3) 186–198.

[53] Larson, R. C., I. R. Webb. 1995. Period and phase of customer replenishment - a new approach to the strategic inventory routing problem. *European journal of Operational Research* **85**(1) 132–148.

[54] Martins, E., M. Pascoal. 2003. A new implementation of Yens ranking loopless paths algorithm. *4OR: A Quarterly Journal of Operations Research* **1**(2) 121–133.

[55] Mu, S., M. Dessouky. 2011. Scheduling freight trains traveling on complex networks. *Transportation Research Part B: Methodological* **45**(7) 1103–1123.

[56] Nozick, L. K., E. K. Morlok. 1997. A model for medium-term operations planning in an intermodal rail-truck service. *Transportation Research Part A: Policy and Practice* **31**(2) 91–107.

[57] Osman, I., S. Salhi. 1996. *Modern Heuristic Search Methods*, chap. Local search strategies for the vehicle fleet mix problem. John Wiley & Sons, Chichester, UK, 131–153.

[58] Pessoa, A., M. P. De Aragão, E. Uchoa. 2007. A robust branch-cut-and-price algorithm for the heterogeneous fleet vehicle routing problem. *Proceedings of the 6th international conference on Experimental algorithms*. WEA'07, Springer-Verlag, 150–160.

[59] Petch, R., S. Salhi. 2004. A multi-phase constructive heuristic for the vehicle routing problem with multiple trips. *Discrete Applied Mathematics* **133**(1-3) 69–92.

[60] Renaud, J., F. F. Boctor. 2002. A sweep-based algorithm for the fleet size and mix vehicle routing problem. *European Journal of Operational Research* **140**(3) 618–628.

[61] Rochat, Y., R. D. Taillard. 1995. Probabilistic diversification and intensification in local search for vehicle routing. *Journal of Heuristics* **1** 147–167.

[62] Salhi, S., G. K. Rand. 1993. Incorporating vehicle routing into the vehicle fleet composition problem. *European Journal of Operational Research* **66**(3) 313–330.

[63] Song, J. H., M. Savelsbergh. 2007. Performance measurement for inventory routing. *Transportation Science* **41**(1) 44–54.

[64] Taillard, E. D., G. Laporte, M. Gendreau. 1996. Vehicle routeing with multiple use of vehicles. *The Journal of the Operational Research Society* **47**(8) 1065–1070.

[65] Toth, P., D. Vigo. 2002. *The vehicle routing problem*, vol. 9. Society for Industrial Mathematics.

[66] Wassan, N. A., I. H. Osman. 2002. Tabu search variants for the mix fleet vehicle routing problem. *Journal of The Operational Research Society* **53** 768–782.

[67] Yaman, H. 2006. Formulations and valid inequalities for the heterogeneous vehicle routing problem. *Mathematical Programming* **106**(2) 365–390.

[68] Yen, J. 1971. Finding the K shortest loopless paths in a network. *Management Science* **17**(11) 712–716.

Appendix A

Static Trip Planning Problem

The static trip planning problem is given the *shipment-block assignment*. That is, each shipment has been pre-assigned to a sequence of blocks. A given block, however, may include multiple train-blocks. The goal is to determine both the train-blocking path and train-run path. The sequential algorithm, described in Section 2.5.1, is appropriate to solve the static trip planning problem since each shipment is restricted to a small network. We discuss two cases in the following:

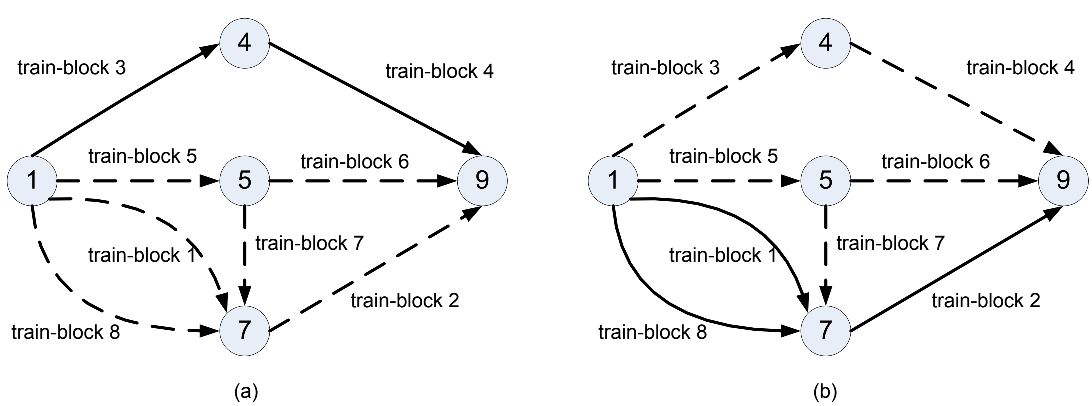

Figure A.1: Train-block network illustration of static trip plan

(a) Each block only contains one train-block.

Take Figure 2.1 for example, we consider one shipment is assigned to a sequence from block 3 to block 4. Since both block 3 and block 4 are associated with only one train path, we can tell the train-blocking path directly, which is train-block 3 to train-block 4. Figure A.1(a) shows the corresponding train-block network with solid lines.

(b) Any block contains more than one train-blocks.

A more complex case is when blocks are separated into multiple train-blocks. In Figure 2.1, block 1 is associated 2 train paths where block 1 can travel on either Train A or Train D. In this case, the block is considered as two train-blocks: train-block 1 and train-block 8 in the train-block network (Figure 2.2). For a shipment with a block sequence of block 1 followed by block 2, two candidate train-blocking paths include: train-block 1 to train-block 2, or train-block 8 to train-block 2. The corresponding train-block network is shown in Figure A.1(b). The optimal train-blocking path can be determined within a much smaller network.

In both scenarios, we can obtain the optimal train-blocking path relatively quickly. Our next step is to decide the fastest train-run path, which is the same as the procedure shown in Algorithm 3.

Compared to the TCCR problem, the static trip planning problem is solved using a smaller train-block network for each shipment, instead of dealing with the whole train-block network. In this sense, our sequential algorithm is able to solve the static trip planning problem in a relatively short amount of computational time.

Appendix B

Implementation Issues in the TCCR Algorithms

This railroad trip planning research was conducted in collaboration with Innovative Scheduling, Inc. They desired an efficient transportation planning tool to solve a large data set as quickly as possible. Therefore, much effort was dedicated to improve the computational time and reduce the storage space. We introduce three approaches used in this research to implement the dynamic trip planning algorithms.

- Check the shortest path from previous search

  A rail company typically deals with many shipments every week, and many of them may have the same origin and/or same destination. Since we apply a labeling algorithm to solve the shortest path problem for each shipment, it saves computational time if we can use the shortest paths obtained from previous searches. In the implementation, we check the shortest paths which have been found when sending a new shipment. If the new shipment has the same origin and destination as before, then we just duplicate the obtained shortest blocking path. If the new shipment has the same origin but different destination, then we check if the new destination has already been included in the existing shortest blocking path. Due to the characteristic of the shortest path, any sub-path within the shortest path must be optimal as well. Hence, if the new destination has already been searched, we can duplicate the sub-path from the obtained shortest path. As a result, the procedure for determining the shortest train-blocking paths improves significantly.

- Limit the search area

  To improve the traditional approach of finding the shortest path, Fu et al. [34] summarized various heuristic shortest path algorithms that have been successful in practice. One useful method is to limit the search area, which is also applied in our algorithm. Specifically, when a node is selected for expansion, we skip the examination of the links that have a low probability of being either on the shortest path or used in practice. In the implementation, an origin-destination table is created to show if it is possible to go from a origin to a destination. In this way, the number of search links is greatly reduced for each shipment and the computational time is improved as well.

- Cross linked list structure

  In the implementation of the bump-shipment algorithm, one of the challenges is to implement an effective data structure so the algorithm executes conveniently. Since the problem considers the due time of each shipment and the capacity utilization of each trip, two major issues arise frequently: the sequence of trips a shipment is taking and the sequence of shipments the trip is carrying. To quickly determine these solutions, a cross linked list structure is implemented as shown in Figure B.1. Following the down arrow, we can easily find the train-run path of each shipment. Following the right arrow, we can also find the shipments that are riding on certain train segment as needed to reschedule. The cross linked list is helpful to implement the bump-shipment algorithm and reduces the storage requirements.

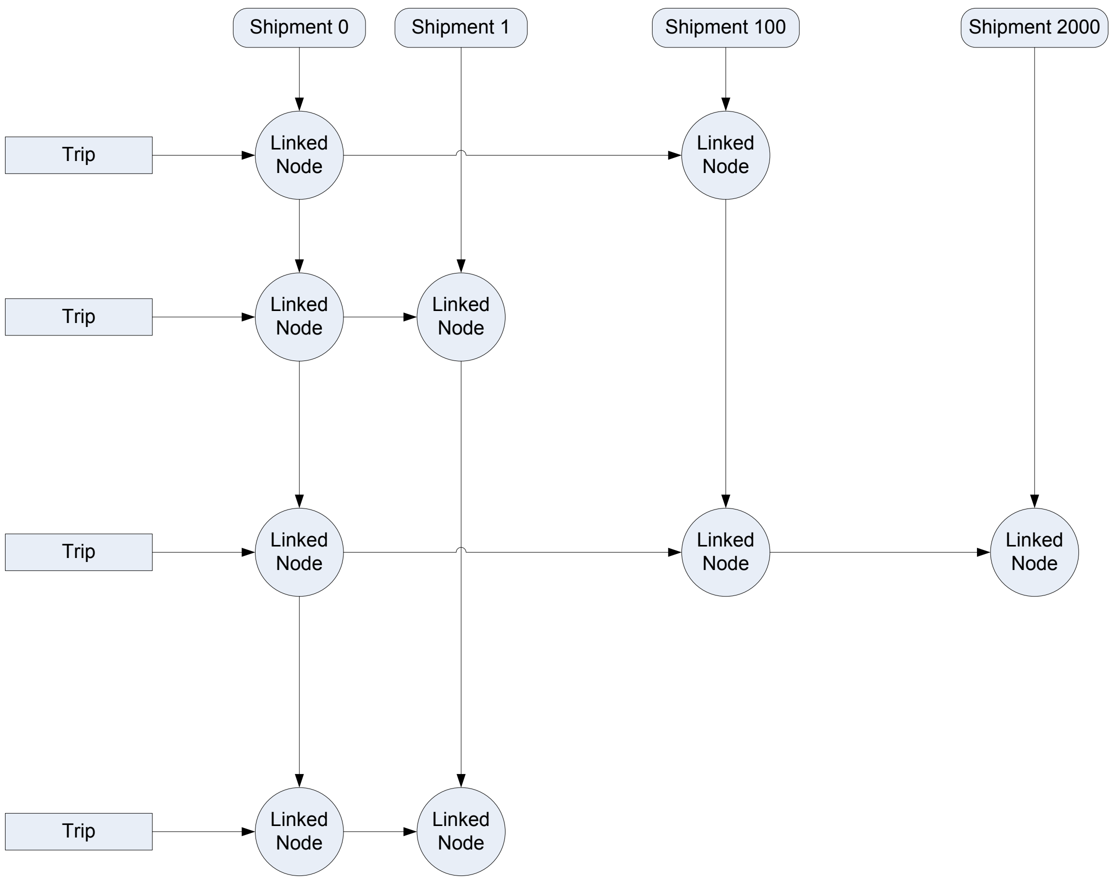

Figure B.1: Cross linked list illustration

Appendix C

Implementation of Shift Generation after Visiting a New Source

During the process of shift generation, it is possible for a trailer to visit multiple sources on a single shift. After loading product from a source, a trailer delivers product to customers based on a sweep procedure. When the remaining product is below a certain threshold, the trailer may consider refilling product. The algorithm then selects the source that is nearest to the trailer's current location. If this source is different from the last one, the trailer refills product and serve customers assigned to this new source. That is, the trailer is moving to serve a new set of customers.

After refilling product at a new source, which customer to consider next is an arbitrary decision. To avoid a long shift, we assume the trailer most likely moves towards the depot after visiting a new source. Therefore, a reduced set of customers that are generally between the source and the depot is considered after the trailer visits a different source. The procedures to determine the search area is summarized in the following:

(1) Recall that all the customer locations are represented by (ρ_i, θ_i) with the assigned source as the center. We convert all the customer location (ρ_i, θ_i) to new coordinates

(ρ_i, θ_i^1), where

$$\theta_i^1 = \theta_i - \theta_D, \quad \forall i \tag{C.1}$$

and θ_D is the location of depot with the new source as center. Therefore, the new coordinate for depot θ_D^1 is 0.

(2) Define customer C_f as the last customer visited before refilling and Define θ_f^1 as the location of customer C_f.

(3) Determine the search area based on θ_f^1. The basic idea is to keep the direction of θ_f^1 and move towards the depot. (The ending search point can be either the depot or $\theta_f^1 \pm \frac{\pi}{2}$, whichever occurs later.)

(3.a) if $\theta_f^1 \leq 0$, select customers in increasing order between $\left[\theta_f^1, \max\{0, \theta_f^1 + \frac{\pi}{2}\}\right]$, as shown in Figure C.1 (a) and (b).

(3.b) if $\theta_f^1 > 0$, select customers in decreasing order between $\left[\min\{0, \theta_f^1 - \frac{\pi}{2}\}, \theta_f^1\right]$, as shown in Figure C.1 (c) and (d).

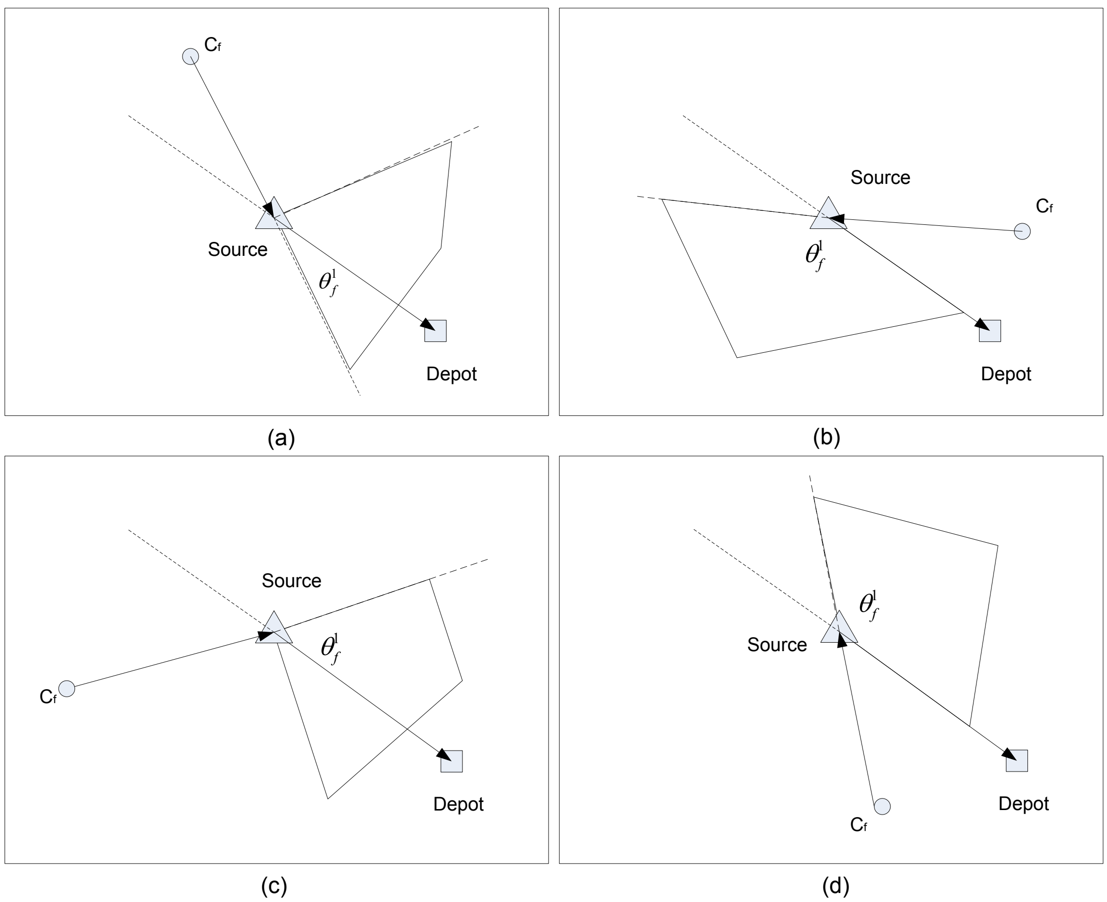

Figure C.1: Customer selection after visiting a new source