2015

The Benefits of Information Sharing in Carrier-Client Collaboration

Hossein Zolfagharinia
Wilfrid Laurier University, zolf1380@mylaurier.ca

The Benefits of Information Sharing in Carrier-Client Collaboration

Author:
Hossein Zolfagharinia

Under the Supervision of Dr. Michael Haughton

DISSERTATION

Submitted to the School of Business and Economics, Wilfrid Laurier University
In Partial Fulfillment of the Requirements for Doctorate of Philosophy in
Management, Operations and Supply Chain Management

ABSTRACT

In the trucking industry, many transport providers face highly variable demands from clients as well as other challenging issues reported by the American Trucking Associations. The Ontario Trucking Association also reports that Canadians face similar concerns. Despite several attempts in the literature, the need for operational improvements by incorporating simple and effective methods is still felt.

This dissertation includes three related papers to investigate different methods that can help transport providers improve their operational efficiency. The first paper models and measures the profit improvement trucking companies can achieve by collaborating with their clients to obtain advance load information (ALI). The main approach is to formulate a comprehensive and flexible mixed integer mathematical model and implement it in a dynamic rolling horizon context. The findings illustrate that access to the second day ALI can improve the profit by an average of 22%. Moreover, increasing ALI from two to three days improves the profit by a further 6%. We also found that the impact of ALI depends on radius of service and trip length but statistically independent of load density and fleet size.

The second paper investigates the following question of relevance to truckload dispatchers striving for profitable decisions in the context of dynamic pick-up and delivery problems: "since not all future pick-up/delivery requests are known with certainty (i.e., advance load information (ALI) is incomplete), how effective are alternative methods for guiding those decisions?" We propose a simple intuitive policy and integrate it into a new two-index mixed integer programming formulation, which we implement using the rolling horizon approach. On average, in one of the practical transportation network settings studied, the proposed policy can, with just

second-day ALI, yield an optimality ratio equal to almost 90% of profits in the static optimal solution (i.e., the solution with asymptotically complete ALI). We also observe from studying the policy that second-day load information is essential when a carrier operates in a large service area. We enhance the proposed policy by adopting the idea of a multiple scenario approach. With only one-day load information, the enhanced policy improves the ratio of optimality by an average of 6 percentage points. That improvement declines with more ALI. In comparison to other dispatching methods, our proposed policy and the enhanced version we developed were found to be very competitive in terms of solution quality and computational efficiency.

Finally, inspired by a real-life third party logistic provider, this study addresses a dynamic pickup and delivery problem with full truckload (DPDFL) for local operators. The main purpose of this work is to investigate the impact of potential factors on the carriers' operational efficiency. These factors, which are usually under managerial influence, are vehicle diversion capability, the DPDFL decision interval, and how far in advance the carrier knows of clients' shipment requirements; i.e., advance load information (ALI). Through comprehensive numerical experiments and statistical analysis, we found that the ALI and decision interval significantly influence the total cost, but diversion capability does not. The findings also reveal that the impact of the re-optimization interval depends on the subcontracting cost and level of ALI. A major contribution of this work is that we develop an efficient benchmark solution for the static version of the DPDFL by discretization of time windows. We observed that three-day ALI and an appropriate decision interval can reduce deviation from the benchmark solution to less than 8%.

To my sweetheart, Maryam

ACKNOWLEDGMENTS

This is the last and I believe the most difficult part of my thesis because it is hard to find the right words to thank those who were tremendously helpful and kind to you over more than half a decade. When I joined the PhD program in 2009, I would never imagine such a long and adventurous journey. This life changing experience would have never lead to success if I was not surrounded by positive, influential, and helpful people.

I would like to first say a huge thank you to my advisor, Dr. Michael Haughton for being an incredible mentor. I cannot thank him enough for his continued support and helping me to grow. Aside from his technical advices, he was a true role model who taught me to be patient, hardworking, optimistic, and working with integrity. Without his constant feedback, guidance, and encouragement this PhD would not have been achievable.

I am also very grateful to my dissertation committee members, Drs James Bookbinder, Sapna Isotupa, Paul Iyogun for their considerable level of involvement in my work. They raised the bar high and constantly provided constructive comments that helped remarkably in refining my work quality. Thus, once more, thank you Jim for treating me like your own student at the University of Waterloo, Sapna for spending so many hours to guide me and verify the mathematical models, and Paul for always being very kind, but asking thorough questions.

A special thank to the director of the PhD program, Dr. Hamid Noori, who is leading this program to success. There were many ups and downs, but he was always available to listen to my concerns. Furthermore, I would like to extend my appreciation to all other faculty members at the Operations and Decision Sciences (ODS) department who evidently show their support to have a successful PhD program especially two junior faculty members: Drs Mike Pavlin and Mojtaba Araghi.

The strong administrative team also deserves a high level of appreciation and I want to take this opportunity to express my gratitude to all of them in the PhD office (Mrs. Amanda Kristensen), the graduate office (Mrs. Helen Paret, Mrs. Jennifer Williams, Mrs. Vanessa McMackin, and Mrs. Meghan Delaney), and the international office (Mrs. Anna Choudhury and Mr. Peter Donahue). I only remember pleasant experiences whenever I approached them.

I am also thankful to those who supported me both mentally and emotionally: my family and friends. Words cannot express how grateful I am to my parents, my younger brother, and my in-laws for all the sacrifices they made on my behalf. Their unconditional love and support were the best encouragement to move on in hard times. In addition, I would like to thank all my friends in my office (Maryam Hafezi, Adrian Tan, Alireza Azimian, Yan Jin, Justin Mindzak, Wensi Zhang, Abiodun Isiaka) and around the world who supported and incented me to strive toward my goals. I would also like to acknowledge the editorial help of Kelsey Koebel (one of my best former students) and my research assistant during the last year.

Finally, I have the most special thanks to the better half of myself and my partner in crime. The first person who encouraged me to continue my education at the PhD level and stood by me all the way since the beginning. I am very blessed to have her and hope this end starts a new chapter in our life for all good to come.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

In the trucking industry, many transport providers face highly variable demands from clients as well as other challenging issues as reported by the American Trucking Associations. The Ontario Trucking Association also reports that Canadians face similar concerns. Despite several attempts in the literature, the need for operational improvements by incorporating simple and effective methods is still felt.

Asset repositioning and driver turnover are among the most challenging issues that trucking companies (carriers) encounter. Asset repositioning, which has been studied by, e.g., Crainic (2000); and Wieberneit (2008), is due to natural characteristics of truckload transportation networks such as demand dynamism and network imbalance between supply and demand. Ergun et al. (2007a) report that empty movement of trucks costs U.S. carriers nearly 165 billion dollars annually. Based on the American Trucking Association (ATA) 2013, the ratio of empty to total mileage is usually higher for small carriers (22%) with a sparser network of lanes than larger ones with a more sophisticated lane network (17%). Since empty repositioning of trucks does not generate any positive contribution, it will lower different measures of performance (e.g., carrier's profit).

The issue of driver turnover is strongly influenced by drivers' dissatisfaction with work schedules requiring overly long periods away from home. Studies confirming this include Rodriguez and Griffin (1990), Shaw et al. (1998), Keller (2002), and Suzuki et al. (2009). The driver turnover problem is significant (according to the Council of Supply Chain Management Professionals (2006), it can reach 130% in a year) and costly: the replacement cost of a driver (e.g., including training and loss of experience) is estimated to cost between $2,200 to over $20,000 with an average of $8000 (e.g., Rodriguez et al., 2000). Given the size of the U.S.

trucking industry, driver turnover translates to approximately three billion dollars a year (Suzuki et al., 2009).

In the second chapter, titled “the benefit of advance load information for truckload carriers”, we consider relatively small trucking companies (with 20 trucks and fewer). Given the highly fragmented trucking industry in North America, these small companies form the major part of the trucking industry. This problem can be placed under pickup and delivery categories with full truckload where clients’ requests (loads) are gradually received by the carrier. Among various methodologies in the literature, we use the most appropriate method. This method is based on developing a mathematical formulation followed by implementation in the dynamic context using a rolling horizon approach. The mathematical formulation is flexible enough to be easily implemented in the dynamic context.

The contributions of this chapter can be categorized in two broad categories. First, we explicitly model the notion of a home base in designing a dispatching method. This is very crucial because truck drivers need to regularly visit their home due to human related considerations. Missing this consideration adversely impacts the driver turnover rate. Second, we statistically examine the benefits of collaboration (via advance load information sharing) between a carrier and its clients through comprehensive numerical experiments. The statistical analysis reveals that a majority of benefits are achievable by acquiring the second day load information. Although obtaining additional information still improves the profit, the marginal benefit reduces significantly. Moreover, the level of improvement in profit depends on the radius of service and the average trip length of loads. For example, there is more incentive for carriers to improve their relationship with their clients if they operate in a larger service area (larger service radius).

The third chapter, titled "effective truckload dispatch decision methods with incomplete advance load information", considers a similar dynamic problem which was investigated in the previous chapter. Although collaborating through sharing advance load information helps trucking companies to extend the knowledge window of dispatchers and improve the operational efficiency, there is always uncertainty after the knowledge window (i.e., advance load information is not complete). In the absence of exact information about future loads beyond the knowledge window, the dispatcher's range of decisions (load acceptance/rejection, load sequencing, etc.) is influenced by the matter of where the truck will be positioned for serving future (unknown) loads. In this situation, one might choose a more conservative policy that prefers to serve loads that take the truck close to its domicile; i.e., to avoid large empty truck repositioning costs to the domicile (called deadheading costs in this study) when the truck must eventually return deadhead to the domicile.

The main contribution of this chapter is to design a simple and intuitive policy for improving the thin profit margin of carriers. In order to evaluate the performance of the proposed policy under various transportation network settings, the static optimal solution is used as a benchmark. The static optimal solution is unrealistically good because it solves the problem when all load information is available in advance, but it still can be used as a fair benchmark. Another contribution of this work is to reformulate the problem using a two-index mixed integer programming that helps us to solve the model to optimality in the static version.

In one of the most practical settings, we found that the proposed simple policy can generate almost 90% of the static optimal solution with only two days of advance load information. However, the proposed policy does not have an acceptable performance in some specific settings. To improve the performance of the algorithm, we develop the enhanced version of that policy by incorporating a multiple scenario approach (from the vehicle routing problem literature). The enhanced version of the policy significantly produces a higher quality solution when advance load information is limited. To further illustrate the performance of our proposed policies, they are examined against two other dispatching methods (policies).

Unlike the first two chapters, the last study, titled "operational flexibility in the truckload trucking industry", targets small carriers that generally operate in a smaller service area (local operators). In this setting, the notion of home base becomes less important. The main focus of this work is to identify and test the impact of potentially important strategies in decreasing the operational costs of a transport provider. To the best of our knowledge, this is the work that moves beyond sharing advance load information by including diversion capability and a re-optimization interval as two other factors (strategies). The main inspiration of this work is a small third party logistic provider (Logikor Inc.) located in Ontario, Canada. This company accepts all load requests and serves them using either the company owned trucks or subcontracts them to other carriers. Through a comprehensive numerical study and applying a regression model, we found out that advance load information and the re-optimization interval significantly reduces the total cost but that diversion capability does not.

The next step of this study is to introduce different policies (based on significant strategies). They can be compared against each other based on their deviation from a benchmark solution. The benchmark solution (similar to the previous chapter) is the static optimal solution. However, solving the developed mathematical model is taking too much time even for small instances (e.g., >48hrs for 6 trucks and 50 loads). Thus, we design an efficient algorithm based on the idea of time window partitioning (proposed by Wang and Regan, 2002). We prove that the proposed algorithm converges to the minimum total cost as the number of iterations increases.

CHAPTER 2

THE BENEFIT OF ADVANCE LOAD INFORMATION FOR TRUCKLOAD CARRIERS

2.1. Introduction and Literature Review

Asset repositioning is one of the important issues in truckload transportation (Crainic, 2000; Wieberneit, 2008). A recent estimate that 18% of trucks move empty every day translates to more than 165 billion dollars annually in the US market (Ergun et al., 2007a). This is a natural result of imbalance between supply and demand at different cities. To correct for this issue, strategies such as collaborative transportation (CT) are used to ensure that trucks are repositioned in a way that efficiently fulfills future demand.

In CT, logistics participants (i.e., shippers/consignees and carriers) collaborate with each other to improve the performance of transportation planning. Examples of collaborative transportation networks are Nistevo (www.nistevo.com) and Transplace (www.transplace.com). They are non-asset based companies that provide modular software under common web-based network to create connectivity and encourage collaboration. These fairly young companies (Nestivo founded in 1997; Transplace founded in 2000) focus on finding new opportunities which cannot be achieved within the internal company scope. One of the best examples is empty repositioning of trucks. The shipper lacks information on how its shipment requests might impact the empty repositioning of trucks. However, the carriers implicitly charge the shipper for this cost component. This issue can be resolved by connecting shippers and carriers to their partners through visibility of orders. For example, two members of the Nistivo network could save 19% over the cost of one-way rates and their shippers experience a more routine schedule and lower empty repositioning cost (Lynch, 2001).

In general, CT helps to reduce total transportation costs, increase trucks utilization and lower driver turnover (Ergun et al., 2007b). Collaboration could be among transportation clients (e.g., Ergun et al., 2007a), among carriers (e.g., Özener et al., 2011), or between client(s) and carrier(s) (e.g., Tjokroamidjojo et al., 2006) or all the above scenarios. Collaboration between a carrier and its clients is the focus of this study. One of the least costly methods when freight transportation service clients and carriers collaborate with each other is to communicate timely load information (from clients to carriers) and pickup and delivery plans (from carriers to clients). The benefit of information sharing has been extensively examined in several contexts such as inventory management or production planning. See, for example, Bourland et al. (1996), Lewis and Talalayevsky (1997), Gavirneni et al. (1999), Frohlich and Westbrook (2001), Patterson et al. (2003), Helper et al. (2010), and Zolfagharinia and Haughton (2012). However, such attempts in the transportation field remain limited. These studies include the works by Mitrović-Minić et al. (2004), Jaillet and Wanger (2006), Tjokroamidjojo et al. (2006), Angelelli et al. (2009), and Özener et al. (2011).

We distinguish between less than truckload (e.g., Mitrović-Minić et al., 2004; Angelelli et al., 2009) and full truckload literature. Mitrović-Minić et al. (2004) developed a double-horizon heuristic algorithm for the same-day dynamic pick-up delivery problems with time windows. The heuristic solved the problem with short-term (minimizing total distance) and long-term goals (efficiently serving future requests). The benefit of advance load information was found to be positive but smaller for larger instances. Jaillet and Wanger (2006) addressed the benefit of advance information for two variations of the traveling salesman problem. By defining the notion of disclosure dates for incoming requests, they analytically showed how advance load information helps to improve competitive ratios. Angelelli et al. (2009) examined different short terms strategies for dynamic multiple-period routing problems where requests can be postponed for the next day. They also analyzed the impact of the short-term strategies on the long term objective. The obtained results suggested that 2-day look-ahead policy was definitely superior to

1-day look-ahead policy. Since the problem under consideration in this chapter is a full truckload one, the rest of the review only focuses on relevant full truckload studies.

A recent work which addressed the benefit of information sharing is by Özener et al. (2011). The focus of their study was to answer this question: how does information sharing help carriers to collaborate with each other? Since each carrier has the full information about its demand and cost structure, different lane exchange mechanisms were proposed with and without information sharing. The obtained results showed that information sharing with side payments helped carriers significantly to improve their performances.

One of the relevant studies to the current work is by Tjokroamidjojo et al. (2006). They studied the dynamic load assignment problem (DLAP) in a full truckload industry. They modified the model developed by Keskinocak and Tayur (1998) in the aircraft scheduling problem. Comparing their work with traditional DLAP (e.g., White, 1972; Powell, 1996), the model has a tour building capability. The ultimate goal of their study was to evaluate the benefit of advance load information (ALI) in the dynamic load assignment problem. Tjokroamidjojo et al. (2006) modeled the problem's time dimension implicitly by using a preprocessing approach. Their optimization-based computational analyses illustrated that ALI does not help the carrier to reduce its costs if the truck dispatching decision is fixed as soon as load information is realized.

The closest typical problems to DLAP are full truckload dynamic pickup and delivery problems. They are also called dynamic stacker crane problems (Berbeglia et al., 2010). Dynamic pickup and delivery problems with full truckload (DPDFL) have received much less attention in comparison to the static version. However, the input data are often revealed through time when a client requests transportation services. Thus, it is crucial to assign drivers (or equivalently trucks) to requests (i.e., loads) on a real time basis. The studies by White and

Bomberault (1969) and White (1972) are probably the first attempts show how the load assignment problem can be handled in a dynamic setting in which each node represents a region with demands at the particular point of time. The problem is reduced to a simple transshipment problem if future forecast is known.

The more realistic model appeared in the work by Powell (1986) since it considered two types of vehicle movement between regions. The model does not let trucks move between regions unless there is an actual demand for them. Thus, if the realized demand in a particular lane is less than the number of assigned trucks, extra trucks are held at their current locations for future demands. Powell (1987) extended his previous work by presenting the network flow problem. Similar to the previous works each node represents a region at a particular time. Two types of arcs were considered in the model, one represents deterministic information and the other for stochastic ones. Following the same approach, Powell et al. (1988) proposed a model called LOADMAP which combines the real-time load assignment with sophisticated future forecast to maximize the truckload profit and service level. Running the model four times a day could help company to increase its annual profit by 2.5 million US dollars.

In another work, Powell (1996) proposed a stochastic DLAP formulation. He showed that when some stochastic information about future demand is available, the proposed model outperforms the deterministic one, which is updated as new information arrives. The model was evaluated under three conditions: fleet size density, demand uncertainty and ALI. Not surprisingly, the stochastic model is superior with more fleet density, higher uncertainty but not with more advance load information.

Yang et al. (1998) proposed a mixed integer programming with rolling horizon framework for DPDFL in which requests arise continuously. A model was designed for a static case and rerun at each decision epoch. The proposed mathematical model was compared with three simple heuristics. Obtained results with only four vehicles showed that the optimal myopic method produces high-quality solution but, by being computationally inefficient, it was slower than the heuristics. They unified re-sequencing, reassigning, and diversion in their proposed model to minimize empty travel costs, delay costs, and lost revenue as a result of job rejection.

Powell et al. (2000) took a comprehensive simulation-based approach for tackling DLAP. The approach was to design an offline algorithm for the static version and put it into practice for a dynamic problem when demands were gradually realized as the time elapses and there was no information on future demand. They questioned the practical value of optimal myopic solutions in comparison to a greedy solution over a long run given that there was no guarantee of user compliance with the model's solution. User non-compliance often exists in practice, since the model cannot capture all available system information. That is why truckload companies reported that suggested solutions by commercial software are implementable in less than 60% or 70% of time. The result suggested that the greedy approach can be superior in long run in comparison to optimal myopic solution in the presence of uncertainty in customer demands and travel times.

Yang et al. (2004) extended their previous work (Yang et al., 1998) by introducing two mixed integer programming formulations and comparing them with three heuristic decision rules. The objective function components were similar to Yang et al. (1998) but the time windows are soft (i.e., deviation from pickup and delivery times are allowed but penalized). The main contribution was to develop an advanced policy which can use the probability of future demand in repositioning of vehicles to improve the system performance. Using numerical examples, they illustrated that the proposed advanced policy is superior to the one not utilizing probabilistic information of future loads. However, the problem complexity limits them to small-size problems in which only ten vehicles with a thousand of loads are taken into account.

Our proposed model is a comprehensive DPDFL in which several operational factors in the truckload industry are taken into account. To highlight the novelty of this study, we carefully point out the limitations of relevant works in the literature (summarized in Table 2.1). Although all of these papers addressed truckload problems, the key factor that remarkably influences the choice of modeling approach is tour capability (i.e., designing continuous truckload routes). This feature becomes less important when the average time of serving a load is very long (between two to four days) which is the case for large trucking companies working in nationwide or international markets. Powell and colleagues investigated this type of problem which is simplified to different versions of assignment problems. The other stream of relevant works focus on smaller trucking companies that view tour capability as essential. These studies used mixed integer programming to formulate the problem and rolling horizon approach for implementation (Yang et al, 1998; Yang et al, 2004; Gronalt et al., 2003; Tjokroamidjojo et al., 2006).

The defined problem was the same in the studies by Yang et al. (1998) and Yang et al. (2004). The objective was to minimize the total cost (including delay, empty movement, and load rejection costs). They used their models to develop tours with the capability of diverting trucks based on the arrival of new information into the system. As defined by Regan et al. (1995), diversion is a model capability that can divert a vehicle moving empty toward a pickup point to take another request. However, it is not allowed to divert loaded-moving vehicles while updating the decision. Ichoua et al. (2006) estimated that diversion in dynamic vehicle routing problems by improved system performance up to 4.3% despite its operational difficulty.

However, dwelling cost which is one of the important components of costs structure was not part of their model. Another limitation of that work is that trucks moved continuously between different cities which means than a truck may never return to its home base.

Unlike the previous studies, the work of Gronalt et al. (2003) addressed tour length to force trucks return home after a predefined interval. The approach was based on generating tours with a very restrictive assumption that there is no limit on number of available trucks. Their model did not capture the cost of delay and dwelling in designing tours. The proposed policy was also very restrictive in the sense that no loads could be rejected and no trucks could be diverted.

Tjokroamidjojo et al. (2006) addressed a full truckload pickup and delivery problem in which empty movements, dwell, and subcontracting costs were taken into account. They also investigated how much a trucking company can reduce cost by obtaining additional information further in advance. However, their proposed mathematical model was subject to some limitations. For example, similar to Yang et al. (1998) and Yang et al (2004), there was no home base for the trucks.

Addressing the limitation of related studies, we can put the contributions of this chapter in two broad categories:

- To the best of our knowledge, this is the first study of its kind that explicitly considers the notion of home base (domicile/depot) for trucks in designing dispatching rules. This is essential from humanity-related considerations because drivers need to come back home to visit their families. The point is quite important since it is well-known that insufficient time at home adversely affects driver's job satisfaction, leading to high turnover (which, according to the Council of Supply Chain Management Professionals (2006) can be as high as 130% in a year).

Models not addressing this issue overestimate the capacity of transportation network. Moreover, the proposed model can handle load rejection, truck diversion, and advance load information.

- Managerial insights through a comprehensive simulation study:

  In this work, using an advanced load dispatching policy, we gauge the benefit of advance load information for a truckload carrier and test the moderating impact of other transportation network settings.

The remainder of this chapter is organized as follows. In section 2.2, we briefly review common mathematical models and dynamic policy in solving full truckload dynamic pickup and delivery problems. In section 2.3, the problem is defined and formulated as a mixed integer programming (MIP) problem. Section 2.4 briefly describes how the numerical experiments are designed and the proposed MIP is implemented in a dynamic environment by using a rolling horizon approach. Section 2.5 discusses the numerical experiments, statistical analyses of the results, and the ensuing managerial insights. The conclusion and future research directions are provided at the end.

Table 2.1. Summarizing the most related studies to the current study

Author(s) and publication year	Tour Capability	Demand Pro. Information	Decision Interval	Objective Function	Modeling Approach	Diversion Capability	Job Rejection Option	Benefit of ALI
Powell (1987)	X	√	Discrete	Max. Profit	Stochastic Formulation	X	√	X
Powell et al. (1988)	X	X	Discrete	Max. Profit	Stochastic Formulation	X	√	X
Powell (1996)	X	√	Discrete	Max. Profit	Stochastic Formulation	X	√	√
Yang et al. (1998)	√	X	Continuous	Min. Cost	Mixed Integer Programming	√	√	X
Powell et al. (2000)	X	X	Discrete	Min. Cost	Integer Programming	X	√	X
Gronalt et al. (2003)	√	X	Discrete	Min. Cost	Mixed Integer Programming	X	X	X
Yang et al. (2004)	√	√	Continuous	Min. Cost	Mixed Integer Programming	√	√	X
Tjokroamidjojo et al. (2006)	√	X	Discrete	Min. Cost	Integer Programming	X	X	√
The present study	√	X	Discrete	Max. Profit	Mixed Integer Programming	√	√	√

2.2. Review of Common Mathematical Models and Dynamic Policy

2.2.1 Common Mathematical Models

There are two common ways to formulate a DPDFL problem. The first one uses an extended version of the assignment problem (e.g., assignment with timing constraints) to exploit the problem's characteristics. This is the most common approach in the literature (see Yang et al. 1998; Powell et al., 2000; Yang et al., 2004; Tjokroamidjojo et al., 2006). In the second one, the problem can be formulated as a variant of capacitated arc routing problems (CARP) in which each directed arc represents one load with designated origin and destination. Recent works by Liu et al. (2010a, b) proposed an integer-programming model to formulate CARP for truckload industries and a quality lower bound. They also developed a heuristic method based on graph theory to solve the proposed model since the exact method is incapable of handling large problem instances. However, they did not capture time windows for fulfilling demands. Comparing the different approaches in the literature, the former is shown to be more promising to use because the dimensionality of the model grows quickly in the latter case. Among the related studies, the one by Tjokroamidjojo et al. (2006) used an effective approach to handle DPDFL. The utilized approach consists of two parts, a preprocessing part for time-based restrictions and an assignment problem afterwards. Since time-based restrictions are explicitly handled outside the mathematical model, the approach performs well by reducing the number of constraints and decision variables. Although our approach is similar to Tjokroamidjojo et al. (2006), we must handle some of the time-based constraints inside the MIP because most of the loads and trucks attributes are determined after solving the model.

2.2.2 Dynamic Policy

Before formulating the abovementioned problem, it is worthwhile to briefly review the most common dynamic strategy used in DPDFL. In dynamic models for general freight transportation, Powell et al. (2007) proposed different algorithmic strategies based on the information classes. When the information class is the data explaining the current status (i.e., no information available about future demand), the algorithmic strategy is classical deterministic programming. However, the classical deterministic program can be also used to handle situations in which there is some probabilistic information available about future loads (see, Yang et al., 2004). To develop a model with tour making capability, the commonly utilized strategy is to formulate a static version of the model and apply it into a dynamic environment by using a rolling horizon framework (e.g. Yang et al., 1998; Powell et al., 2000; Yang et al., 2004; Tjokroamidjojo et al., 2006). To apply the static version, the deterministic mathematical formulation is called at each decision epoch. It has been shown that the solution quality of this strategy is superior to simple heuristics rules, e.g., adding the new load to the end of the current job sequence of a vehicle with smallest marginal cost (Regan et al., 1998; Yang et al., 1998; Yang et al. 2004). However, there is no guarantee that solving a series of sub-problems optimally will always result in a higher quality solution. Tjokroamidjojo et al. (2006) used the same strategy to evaluate the benefit of advance load information when timing of preplanning was addressed.

A rolling horizon approach has been widely used for modeling dynamic problems in the areas of inventory management and production planning (e.g., Bookbinder and H’ng, 1986; Anupindi et al., 1996; Cheevaprawatdomrong and Smith, 2004). In these studies, three time fences are usually defined, namely frozen interval, re-planning interval and forecasting interval/window. The first two terms are self-explanatory. The term *forecasting interval* illustrates how far in advance the data, either stochastic or deterministic, are included in the model (Kern and Wei, 1996). Interested readers are referred to a comprehensive paper by Chand et al. (2002) who reviewed more than two hundred studies in inventory management and production planning.

Using the most common strategy, we formulate the static version of the defined problem and then re-optimize it in small discrete intervals as the new information arrives into the system. Moreover, as mentioned in the literature review, maximum flexibility is incorporated in formulating the model by considering reassignment, re-sequencing of loads, and even diversion of empty vehicles, as defined by Regan et al. (1995). Given the abovementioned points, interval freezing is not considered in fixing the future plan and the re-planning interval is a relatively short constant duration (i.e., each period). Forecasting interval in the other research fields is equivalent to how far in advance loads information is passed from clients to the carrier in truckload trucking. Varying that interval in our model yields answers to this study's main research question: How significant are the benefits from acquiring load information further in advance?

2.3. Problem Definition

As mentioned earlier, the problem under study is called dynamic pickup and delivery truckload. There is a fixed fleet of trucks in the transportation network. The customers' demands (loads) are known gradually as time elapses. We retain the literature's standard assumption that each trip is executed without a break. Loads and trucks have their own attributes. The truck attributes are home domicile, hours away from home, the maximum allowed hours away from home, determined by a carrier or federal department of transportation (for drivers), and the current location. The load attributes are the earliest and latest pickup time, the maximum permissible delay time, the pickup location and the delivery location. Taking all the attributes of loads and trucks into account, the optimal DPDFL solution specifies the carrier's profit maximizing decisions concerning (i) whether to accept or reject a new load, (ii) the sequence of accepted loads that each truck will serve. The major assumptions are as follows:

- Vehicles can be homogenous or heterogeneous regarding their capacity and capability for handing different load types.
- The shipment cost is a linear function of travel time which itself is a linear function of distance.
- Similar to what is common in the literature (e.g., Powell et al., 1988; Powell, 1996), the gained revenue is proportional to the trip length, i.e., the distance/time between pickup and delivery points.
- The length of each tour (i.e. tour time span) has to be less than the maximum hours that a driver can be away from home.
- Full truckload transportation is considered (i.e., each vehicle can handle one load at a time).
- Given long haul transportation, loading and unloading times are a negligible part of the total time to serve a load and can therefore be ignored.
- There is a hard time-window to serve a load. Thus, the load will be rejected if it cannot be served within the predefined time interval.
- Depot is the home domicile of drivers. A truck is returned to the depot if it is not scheduled to serve any load at that decision epoch. This is a common practice if the dispatcher has access to advance load information (e.g., knowing that there is no request arriving for the rest of the day). The logic is simple because the average

repositioning is typically shorter from the depot (if it is located at the center) and dwelling cost is negligible at the driver's home domicile.

2.3.1. The Model Inputs

To formulate the proposed model, notations, parameters, and decision variables are presented below.

- **Notation**

I: set of all available trucks, indexed by i

J: set of loads, indexed by r, j, k

L: set of depots

- **Parameters**

a_j: departure location of load j

b_j: destination location of load j

α_j: the earliest departure time of load j

$D(.,.)$: travel time between any two points in the service area. Traveling time between two locations can be described as function of distance.

h_i:home domicile of truck i, (i.e.$h_i \epsilon$L)

N: maximum hours that a driver can be away from home

U_k: maximum permissible delay for serving customer k

n_i: maximum hours left for truck i to be away from its home at the decision epoch

e: the revenue earned per hour while moving loads

c: the traveling cost (empty or loaded) per hour of driving

l: the penalty cost per hour for late pickup

w: the penalty cost per hour for a truck being idle at any load location (dwelling cost)

τ: time at the decision epoch

The current location of each truck is important at each decision epoch because of the problem's dynamic nature. If the current location of truck i is denoted with η_i, $D(\eta_i, q)$ shows the traveling time from current location of truck i to the location q. Dwell time is the waiting time experienced by a driver/truck if the truck must wait at the pickup location (i.e., it reaches the pickup location of load j earlier than α_j). Although we consider the same dwell cost for all clients' locations in this computational study, the model is flexible enough to address varying dwelling costs across client locations. Still, our study does reflect that dwelling costs at truck/driver domicile is significantly smaller than at client locations. This is because there is no extra facility usage cost for, say, a driver to dwell at his/her home or at accommodations provided by the carrier (e.g., Challenger Motor Freight's well-equipped rest facility for drivers at its Cambridge depot, more detail about this trucking company can be found at its official website: http://www.challenger.com).

Since the model is flexible enough to allow reassignment and re-sequencing of loads and diversion of empty moving trucks, the decision made at the previous decision epoch can be modified at the current decision epoch for all the loads which have not received service yet. To acknowledge this assumption, we first define TST(i) as the status of truck i at the decision epoch τ. TST(i) can take three values 1, -1, 0 meaning truck i is moving loaded, empty (either moving or idle at any location other than the depot), or sitting idle at its own depot, respectively. If truck i is serving load j at the decision epoch τ, it will be available at the later time, $\tau + D(\eta_i, b_j)$ at the destination location of load j. If a truck is idle or empty, $\text{TST}(i) \leq 0$, then truck i will be available for scheduling at time τ at its current location. There is also a need to keep track of load status which is denoted with LST(*j*). There are four possible load statuses. If the load is being served at the decision epoch, LST(*j*) is equal to 2. The other loads which were already rejected never enter the model (i.e., LST(*j*) =0). The loads which are accepted but have not received service yet (i.e., LST(*j*) =1) enter the model for possible reassigning and re-sequencing. In order to distinguish new loads (i.e., the loads for which acceptance is not finalized yet) from the current ones, their statuses will be LST(*j*) =3. We also define ST(i,*j*) as a binary parameter to address the status of truck and load together. If truck *i* is serving load *j* at the decision time, then ST(i,*j*) takes 1 otherwise 0.

Another important time-dependent attribute is the number of hours left for the drivers to return home. Two situations can be considered for them: sitting idle at their home domicile (i.e., $n_i = N$) or on duty away from their home ($n_i < N$). It will be explained how these features are incorporated in the proposed model.

Since there is no type of uncertainty considered in traveling time, it is enough to calculate lateness at the load pickup locations. Based on abovementioned assumption, there are two lateness types defined as follows.

$\mathrm{DL0}(i,j)$: the lateness duration at the load pickup location $\mathrm{a_j}$ if truck *i* serves load *j* first. There is no difference if the truck is heading off from its depot or the previous delivery location of a load. The only consideration is whether there is enough time to reach to the pickup location of load *j* or not.

For $\mathrm{TST}(i) < 1$, $\mathrm{DL0}(i,j)$ modified as $\mathrm{DL0_D}(i,j)$=max $\left(0, \mathrm{D}\left(\mathrm{\eta_i}, \mathrm{a_j}\right) + \tau - \mathrm{\alpha_j}\right)$. If the truck is moving loaded, $\mathrm{TST}(i) = 1$, toward the destination of a load (e.g., load *k*), $\mathrm{DL0_D}(i,j) = \max\left(0, \tau + \mathrm{D}(\mathrm{\eta_i}, \mathrm{b_k}) + \mathrm{D}\left(\mathrm{b_k}, \mathrm{a_j}\right) - \mathrm{\alpha_j}\right)$. If the maximum traveling time for the driver is approaching, the truck lateness for an empty truck at load *j* pickup location will be $\mathrm{DL0_T}(i,j) =$

$\max\left(0, \tau + \mathrm{D}(\eta_i, h_i) + \mathrm{D}(h_i, a_j) - \alpha_j\right)$ and for a loaded truck (e.g. while serving load *k*) will be $\mathrm{DL0_T}(i,j) = \max\left(0, \tau + \mathrm{D}(\eta_i, b_k) + \mathrm{D}(b_k, h_i) + \mathrm{D}(h_i, a_j) - \alpha_j\right)$.

$\mathrm{DL1}(j,k)$: the minimum lateness at the load pickup location *k* if the same truck serves load *k* immediately (or through its depot) after load *j*. Load *k* will experience some lateness if there is not enough time to reach the pickup location of load *k* immediately after serving load *j*. It is denoted with $\mathrm{DL1_D}(j,k) = \max\left(0, \left(\alpha_j + \mathrm{D}(a_j, b_j) + \mathrm{D}(b_j, a_k)\right) - \alpha_k\right)$. However, the minimum lateness of load *k* if it is served after load *j* via depot of truck *i* will be $\mathrm{DL1_T}(i,j,k) = \max\left(0, \left(\alpha_j + \mathrm{D}(a_j, b_j) + \mathrm{D}(b_j, h_i) + \mathrm{D}(h_i, a_k)\right) - \alpha_k\right)$.

2.3.2 Preprocessing Stage

As mentioned earlier in section 2.2.1, we tackle the static version of problem in two stages. In the first stage, the preprocessing stage, time considerations are explicitly taken into account. In this stage, the following two tasks are performed: 1) updating all dynamic attributes of trucks (e.g., hours away from home and current truck location) and loads (e.g., a load is waiting to be served or being served) 2) identifying infeasible combinations of loads and trucks and infeasible combination of loads when they are served by the same truck.

Given the current status of the trucks, it is checked to see whether a particular truck is eligible for serving a certain load. This must be done for all available truck-load combinations. It is obvious that certain truck-load combinations are not feasible if the truck cannot be available at the pickup location of the load without violating the maximum delay. To check for feasibility, a set of binary parameters will be defined as TL_{ik}^0 and TL_{ik}^1. If it is feasible for truck *i* to serve load *k* directly (i.e., $\mathrm{DL0_D}(i,j) \le U_j$), TL_{ik}^0 takes 1 otherwise 0. As defined earlier, U_j is maximum permissible delay for serving customers. Thus, $\mathrm{TL}_{ik}^0 = 0$ means that truck *i* (based on its current attributes) cannot be available at pickup location of load k without violating its time window. Similarly if it is possible for truck i to serve load k via its home depot ($\mathrm{DL0_T}(i,k) \leq U_k$), $\mathrm{TL}_{\mathrm{ik}}^{1}$ is set equal to 1 otherwise 0.

Similar to what is done for truck-load combinations; we define another set of binary parameters (called LL_{jk}^{0} and LL_{jk}^{i}) to check the feasibility of serving load k immediately (via depot of truck i) after load j. Here, it is checked the best possible situation for load combinations. For example, load k cannot be served directly (or via depot of truck i) after load j when $\mathrm{DL1_D}(j,k) > U_k$ (or $\mathrm{DL1_T}(i,j,k) > U_k$), i.e. LL_{jk}^{0}=0 (or LL_{jk}^{i}=0). On the other hand, if the minimum lateness is smaller or equal than the maximum allowable delay (i.e.,$\mathrm{DL1_D}(j,k) \leq U_k$ or $\mathrm{DL1_T}(i,j,k)$)$\leq U_k$), the combination is not conclusively infeasible LL_{jk}^{0}, LL_{jk}^{i}=1. It is extremely important to note that having LL_{jk}^{0} or LL_{jk}^{i}=1 does not guarantee the load feasibility at the end since the decision at this stage is made based on the minimum lateness not the actual lateness. Considering different possible assignment decisions, some load combinations with LL_{jk}^{0} or LL_{jk}^{i}=1 may or may not be feasible but the one with LL_{jk}^{0} or LL_{jk}^{i}=0 is infeasible with certainty. This exactly explains why we need to have time components in the second phase (i.e., mathematical model). For easier reading of the proposed mathematical model, parameters quantified at the preprocessing stage are summarized in Table 2.2.

Table 2.2. Parameters of the preprocessing stage

Symbols	Definition of parameters obtained from preprocessing stage
TST(i)	Status of truck i, it takes values of: -1,0, 1
LST(j)	Status of load j, it takes values of: 0,1, 2, 3
ST(i,j)	Binary parameter indicating if truck i is serving load j at the decision epoch
$\mathrm{TL}_{\mathrm{ik}}^{0}$	Binary parameter checking the feasibility of serving load k by truck i directly
$\mathrm{TL}_{\mathrm{ik}}^{1}$	Binary parameter checking the feasibility of serving load k by truck i through its depot
LL_{jk}^{0}	Binary parameter checking the feasibility of serving load k immediately after load j
LL_{jk}^{i}	Binary parameter checking the feasibility of serving load k after load j through the depot of truck i

2.3.3 Mathematical Model

Having defined all parameters and dynamic aspects of the model in the preprocessing stage, it is time to define decision variables and formulate the conceptual model.

$$X_{ijk}^{0}: \begin{cases} 1 & \text{If load } k \text{ is served immediately after load } j \text{ by truck } i \text{ and } \mathrm{TL}_{ij}^{0} = \mathrm{TL}_{ik}^{0} = \mathrm{LL}_{jk}^{0} = 1,\ k \neq j \\ 0 & \text{Otherwise} \end{cases}$$

$$X_{ijk}^{1}: \begin{cases} 1 & \text{If load } k \text{ is served through the depot after load } j \text{ by truck } i \text{ and } \mathrm{TL}_{ij}^{0} = \mathrm{TL}_{ik}^{1} = \mathrm{LL}_{jk}^{i} = 1, k \neq j \\ 0 & \text{Otherwise} \end{cases}$$

$$Y_{ik}^{0}: \begin{cases} 1 & \text{If truck } i \text{ serves load } k \text{ at the first stop and } \mathrm{TL}_{ik}^{0}{=}1 \\ 0 & \text{Otherwise} \end{cases}$$

$$Y_{ik}^{1}: \begin{cases} 1 & \text{If truck } i \text{ serves load } k \text{ through its own depot at the first stop and } \mathrm{TL}_{ik}^{1} = 1 \\ 0 & \text{Otherwise} \end{cases}$$

O_k: arrival time at the pickup location of load k

N_{ij}: the remaining allowable time for the driver of truck i when it is at the pickup location of load j.

Before formulating the proposed model, it is important to check which loads enter the model and their notations. As defined earlier, J represents the set of all loads entered the model. However, we are required to differentiate them in order to have a neat mathematical formulation. To do so, the set of new jobs are denoted with $\bar{J}$ (LST(j) =3), the set of accepted jobs waiting for service, $\bar{\bar{J}}$ (LST(j) =1), and the set of jobs being served at the decision epoch, $\hat{J}$ (LST(j) =2). Thus, $J = \bar{J} \cup \bar{\bar{J}} \cup \hat{J}$. Having the parameters and decision variable defined, the model will be formulated as follows. To have a better understanding of the model, we break it down into smaller components and explain them one by one. The objective function to be maximized is the profit which includes the revenue and the relevant costs. It is also worth to note that all the nonlinear terms in the objective function and constraints are written in linear form before implementation.

- **Revenue;** the revenue depends on trip-length of the accepted loads:

$$e\sum_{i\epsilon I}\sum_{j\epsilon \overline{J}\cup\bar{\bar{J}}}\sum_{k\epsilon \overline{J}\cup\bar{\bar{J}}} D(a_k,b_k)\left(X_{ijk}^0+X_{ijk}^1+Y_{ik}^0+Y_{ik}^1\right) \tag{2.1}$$

- **Cost of moving loaded trucks**;

$$c\sum_{i\epsilon I}\sum_{j\epsilon \overline{J}\cup\bar{\bar{J}}}\sum_{k\epsilon \overline{J}\cup\bar{\bar{J}}} D(a_k,b_k)\left(X_{ijk}^0+X_{ijk}^1+Y_{ik}^0+Y_{ik}^1\right) \tag{2.2}$$

- **Cost of moving empty trucks;** empty traveling cost can be as a result of moving trucks from the delivery location of one load to the pickup location of the next load:

$$c\sum_{i\epsilon I}\sum_{j\epsilon \overline{J}\cup\bar{\bar{J}}}\sum_{k\epsilon \overline{J}\cup\bar{\bar{J}}} D\left(b_j,a_k\right)X_{ijk}^0 + c\sum_{i\epsilon I}\sum_{j\epsilon \overline{J}\cup\bar{\bar{J}}}\sum_{k\epsilon \overline{J}\cup\bar{\bar{J}}}\left[D\left(b_j,h_i\right)+D(h_i,a_k)\right]X_{ijk}^1 \tag{2.3}$$

- The empty traveling cost occurs for repositioning empty, idle or loaded trucks to the pickup location of the first load in the sequence:

$$c\sum_{i\epsilon I,TST(i)<1}\sum_{k\epsilon \overline{J}\cup\bar{\bar{J}}} D(\eta_i,a_k)\,Y_{ik}^0 + c\sum_{i\epsilon I,TST(i)<1}\sum_{k\epsilon \overline{J}\cup\bar{\bar{J}}}\left[D(\eta_i,h_i)+D(h_i,a_k)\right]Y_{ik}^1 +$$

$$c\sum_{i\epsilon I}\sum_{j\in\hat{J},ST(i,j)=1}\sum_{k\epsilon \overline{J}\cup\bar{\bar{J}}} D\left(b_j,a_k\right)Y_{ik}^0 + c\sum_{i\epsilon I}\sum_{j\in\hat{J},ST(i,j)=1}\sum_{k\epsilon \overline{J}\cup\bar{\bar{J}}}\left[D\left(b_j,h_i\right)+D(h_i,a_k)\right]Y_{ik}^1 \tag{2.4}$$

- The empty traveling also exists in either of following cases. First, the truck is going back to its depot after serving all its assigned loads (see term 2.5). Second, a moving truck (i.e., either empty or loaded) is not assigned to any load and so it is heading back to its depot (term 2.6).

$$c\sum_{i\epsilon I}\sum_{j\epsilon \overline{J}\cup\bar{\bar{J}}} D\left(b_j,h_i\right)\left[\left(Y_{ij}^0+Y_{ij}^1\right)+\sum_{r\epsilon \overline{J}\cup\bar{\bar{J}}}(X_{irj}^0+X_{irj}^1)-\sum_{k\epsilon \overline{J}\cup\bar{\bar{J}}}(X_{ijk}^0+X_{ijk}^1)\right] \tag{2.5}$$

$$c\sum_{i\epsilon I,TST(i)=-1} D(\eta_i,h_i)\left[1-\sum_{k\epsilon \overline{J}\cup\bar{\bar{J}}}(Y_{ik}^0+Y_{ik}^1)\right]$$

$$+\,c\sum_{i\epsilon I}\sum_{j\in\hat{J},ST(i,j)=1} D\left(b_j,h_i\right)\left[1-\sum_{k\epsilon \overline{J}\cup\bar{\bar{J}}}(Y_{ik}^0+Y_{ik}^1)\right] \tag{2.6}$$

- **Dwelling cost**; this is the cost of waiting at the load pickup location which can occur when the load is either at the beginning of the sequence or after another load.

$$w \sum_{k \epsilon \bar{J} \cup \bar{\bar{J}}} \max(0, \alpha_k - O_k) \tag{2.7}$$

- **Lateness cost;** late service occurs when the truck arrives to the load's pick-up location after its availability. Lateness cost is incurred in all the following situations. A truck (e.g., moving empty, loaded or idle) is scheduled to serve a load directly from its current location, through the truck depot or after another load:

$$l \sum_{k \epsilon \bar{J} \cup \bar{\bar{J}}} \max(0, O_k - \alpha_k) \tag{2.8}$$

Having the objective function formulated, the constraints are introduced as follows. The first and second constraint sets (2.9 and 2.10) ensure that all previous accepted loads will be served but there is no guarantee to take all new loads.

$$\sum_{i \epsilon I} \sum_{j \epsilon \bar{J} \cup \bar{\bar{J}}} (X_{ijk}^0 + X_{ijk}^1) + \sum_{i \epsilon I} (Y_{ik}^0 + Y_{ik}^1) = 1, \qquad k \epsilon \bar{\bar{J}} \tag{2.9}$$

$$\sum_{i \epsilon I} \sum_{j \epsilon \bar{J} \cup \bar{\bar{J}}} (X_{ijk}^0 + X_{ijk}^1) + \sum_{i \epsilon I} (Y_{ik}^0 + Y_{ik}^1) \leq 1, \qquad k \epsilon \bar{J} \tag{2.10}$$

- A truck can serve at most one load at the beginning of a sequence.

$$\sum_{k \epsilon \bar{J} \cup \bar{\bar{J}}} (Y_{ik}^0 + Y_{ik}^1) \leq 1, \qquad i \epsilon I \tag{2.11}$$

- Each accepted load can have only one successor.

$$\sum_{i \epsilon I} \sum_{k \epsilon \bar{J} \cup \bar{\bar{J}}} (X_{ijk}^0 + X_{ijk}^1) \leq 1, \qquad j \epsilon \bar{J} \cup \bar{\bar{J}} \tag{2.12}$$

- The next set of constraints (2.13) ensures that if truck *i* serves load *k* after load *j*, load *j* is either scheduled to be the first load of truck *i* or placed after another load *r*.

$$\sum_{k\in\bar{J}\cup\bar{\bar{J}}}[X_{ijk}^{0}+X_{ijk}^{1}]-[Y_{ij}^{0}+Y_{ij}^{1}+\sum_{r\in\bar{J}\cup\bar{\bar{J}}}X_{irj}^{0}+X_{irj}^{1}]\leq 0, \qquad i\epsilon I,.\,j\in\bar{J}\cup\bar{\bar{J}} \qquad (2.13)$$

- Altogether, constraints (2.14) through (2.17) ensure that O_k does not take on an unrealistically large or small value to prevent dwelling or lateness costs. Constraints (2.14) and (2.15) apply when a truck is serving one load after another load directly while constraints (2.16) and (2.17) are for the case of a truck serving a load at the beginning of a sequence.

$$O_k-D(a_j,b_j)-D(b_j,a_k)-\max(O_j,\alpha_j)\geq\left(\sum_{i\epsilon I}X_{ijk}^{0}-1\right)H, \qquad j,k\in\bar{J}\cup\bar{\bar{J}} \qquad (2.14)$$

$$O_k-D(a_j,b_j)-D(b_j,a_k)-\max(O_j,\alpha_j)\leq\left(1-\sum_{i\epsilon I}X_{ijk}^{0}\right)H, \qquad j,k\in\bar{J}\cup\bar{\bar{J}} \qquad (2.15)$$

$$O_k-\sum_{i\epsilon I}\sum_{j\epsilon\hat{J},ST(i,j)=1}\left[\tau+D(\eta_i,b_j)+D(b_j,a_k)\right]Y_{ik}^{0}-\sum_{i\epsilon I,TST(i)\leq 0}\left[\tau+D(\eta_i,a_k)\right]Y_{ik}^{0}\geq\left(\sum_{i\epsilon I}Y_{ik}^{0}-1\right)H,$$
$$k\in\bar{J}\cup\bar{\bar{J}}, \qquad (2.16)$$

$$O_k-\sum_{i\epsilon I}\sum_{j\epsilon\hat{J},ST(i,j)=1}\left[\tau+D(\eta_i,b_j)+D(b_j,a_k)\right]Y_{ik}^{0}-\sum_{i\epsilon I,TST(i)\leq 0}\left[\tau+D(\eta_i,a_k)\right]Y_{ik}^{0}\leq\left(1-\sum_{i\epsilon I}Y_{ik}^{0}\right)H,$$
$$k\in\bar{J}\cup\bar{\bar{J}}, \qquad (2.17)$$

- Constraints (2.18) and (2.19) ensure that a truck arrives at the pick-up location of load *k* no sooner than after serving load *j* and traveling to load *k* through the depot if such a schedule is implemented.

$$O_k-\sum_{i\epsilon I}\sum_{j\epsilon\hat{J},ST(i,j)=1}\left[\tau+D(\eta_i,b_j)+D(b_j,h_i)+D(h_i,a_k)\right]Y_{ik}^{1}$$
$$-\sum_{i\epsilon I,TST(i)\leq 0}\left[\tau+D(\eta_i,h_i)+D(h_i,a_k)\right]Y_{ik}^{1}\geq\left(\sum_{i\epsilon I}Y_{ik}^{1}-1\right)H, k\in\bar{J}\cup\bar{\bar{J}} \qquad (2.18)$$

$$O_k-D(a_j,b_j)-D(b_j,h_i)-D(h_i,a_k)-\max(O_j,\alpha_j)\geq(X_{ijk}^{1}-1)H, \quad i\epsilon I,j,k\in\bar{J}\cup\bar{\bar{J}} \qquad (2.19)$$

- Constraints (2.20) guarantee that accepted loads are served within U_k hours from their earliest availabilities.

$$O_k - \alpha_k - U_k \leq 0, \qquad k \in \bar{J} \cup \bar{\bar{J}} \qquad (2.20)$$

- Constraints (2.21) impose an upper bound for a driver's allowable time while visiting the first load of the sequence. In this constraint set, M_{ik}^0 and M_{ik}^1 represent the remaining allowable time for the driver of truck *i* when serving load *k* at the beginning of the sequence either directly or through the depot. These two parameters are obtained from the preprocessing stage for all truck-load combinations.

$$N_{ik} - \left(M_{ik}^0 Y_{ik}^0 + M_{ik}^1 Y_{ik}^1\right) \leq \left(\sum_{j \in \bar{J} \cup \bar{\bar{J}}} X_{ijk}^1 + X_{ijk}^0\right) H, \qquad k \in \bar{J} \cup \bar{\bar{J}},\ i \epsilon I \qquad (2.21)$$

- Constraints (2.22) introduce an upper bound for a driver's allowable time when serving load *k* immediately after load *j*. Constraints (2.23) perform similarly for the case that the driver returns to the depot in-between visits.

$$N_{ik} - \left[N_{ij} - (\alpha_j - \min(O_j, \alpha_j) - D(a_j, b_j) - D(b_j, a_k)\right] \leq (1 - X_{ijk}^0)H, \qquad j, k \in \bar{J} \cup \bar{\bar{J}}, i \epsilon I \quad (2.22)$$

$$N_{ik} - [N - D(h_i, a_k)] \leq \left(1 - \sum_{j \in \bar{J} \cup \bar{\bar{J}}} X_{ijk}^1\right) H, \qquad k \in \bar{J} \cup \bar{\bar{J}},\ i \epsilon I \qquad (2.23)$$

- Finally, constraints (2.24) guarantee that all drivers return to the home domicile (i.e., the depot) without violating the predefined time limit.

$$N_{ik} \geq (\alpha_k - \min(O_k, \alpha_k)) + D(a_k, b_k) + D(b_k, h_i) - \left(1 - \left(\sum_{j \in \bar{J} \cup \bar{\bar{J}}} X_{ijk}^1 + X_{ijk}^0 + Y_{ik}^0 + Y_{ik}^1\right)\right) H$$

$$k \in \bar{J} \cup \bar{\bar{J}},\ i \epsilon I \qquad (2.24)$$

2.4. Designing the Experiments and Dynamic Implementation

In this section, we first explain how the numerical study is designed and model's parameters are generated to have useful managerial insights. We then illustrate how the static MIP model is implemented in a dynamic environment with a very simple example.

2.4.1 Experimental Designs

Our investigation of the academic literature and empirical reports suggested the potential influence of the following factors on a carrier's profitability: radius of service, trip length, load density, fleet size, and advance load information. We examine each of these factors at two levels, and the factor of primary interest (advance load information) at three levels. The main reason to focus on ALI is that other factors either are not directly under full control of the carrier (e.g., trip length) or require some capital investments (e.g., fleet size). In such circumstance, ALI is considered because one of the least costly methods when freight transportation service clients and carriers collaborate with each other is to communicate timely load information (from clients to carriers) and pickup and delivery plans (from carriers to clients).

Radius of service: defined as the furthest distance from the depot that a truckload carrier is willing to carry a load. The low of 18 hours (driving) and high of 36 hours (driving) are taken into account.

Trip length: measured as a travel time between a load's origin and destination. The test problems are generated in two categories called short and long trip-length groups. In the former, the majority of loads (80%) are shorter than the radius of service while in the latter the majority (80%) of loads are longer than the radius.

Fleet size: the most recent released statistics from American Trucking Association in 2013 shows that 90.5% of carriers operate with fewer than 6 trucks and only 2.5% of them run their business with more than 20 trucks. The Canadian statistics also show that majority of carriers have fewer than 20 trucks. Thus, the fleet size will be considered at two levels: 6 trucks and 20 trucks.

Load density: number of loads entering to the system per truck per week. Load density is inversely related to the average length of loads (Powell, 1996) which usually ranges between 2 to 2.5 loads (per truck per week) for large companies with the average load length between two to four days. Since this study targets carriers with fewer than 20 trucks with shorter trip length, the load density are studied at two levels, 2.5 (low load density) and 5 (high load density) loads per truck per week.

Advance load information (ALI): it is called knowledge window (KW) by Tjokroamidjojo et al. (2006) who define it as number of hours that loads' information are available in advance. Since the trucking industry is identified with excess capacity and a high level of competitiveness, last-minute call for transportation services is very common in the industry. It is also unusual for a shipper to book a load more than two or three days in advance (Frantzeskakis and Powell, 1990). Thus, acquiring load information very far in advance (e.g., a week or so) does not provide practical managerial insights. That is why we focus our attention on the three ALI levels: 24, 48, and 72 hours.

The result of the abovementioned five factors at different levels becomes 48 combinations. Each test problem (observation) is generated as follows. Since observations should be independent within combinations, the experiments' stochastic conditions are randomly generated. First, the locations of cities are randomly selected. City locations (X-Y coordinates) are randomly selected from the 20x20 grid within a service area in which a depot is located at the center. Second, loads are created by randomly selecting their pick-up and drop-off locations.

Third, the earliest load availability is assigned by using an exponential distribution function (the mean inter-arrival time depends on load density). Fourth, the dynamic attributes of trucks (i.e., the initial location and remaining service hours) are randomly generated for each factor combination. The five test problem by the 48 combinations yielded 240 observations. For each test problem, the parameters are generated as follows:

- The underlying transportation network is considered to have 50 cities across all test problems.
- To generate each load, an origin-destination pair is selected randomly from a 50-city network. The loads are generated according to the previously discussed specifications for the trip length factor.
- The earliest loads availability is generated from exponential distribution in which the average inter-arrival is determined based on load density.
- The initial location of trucks is determined by placing them randomly among the 50 cities. The maximum number of hours left for truck i (n_i) is generated from uniform distribution [radius of service, N] to guarantee that each truck has enough time to return home before the predefined limit.
- The average operating speed in highways is used since the majority of cities are connected to each other via highways. The average operating speed is set to 55 mph, which is typical on US highways (refer to the recent report by the US department of energy (2011)).
- Following Tjokroamidjojo et al. (2006), hourly dwelling and lateness cost are set to be $25 per hour. The maximum permissible delay for serving customers is drawn from a discrete uniform distribution with maximum of 5 hours.

- In trucking industry where drivers can be simply away from home between one to four weeks, most carriers try to have the drivers back home every fortnight (Powell, 1996). To be consistent with these statistics, this study sets the maximum number of hours that a driver can be away from home equal to 240 hours (i.e., N=240 hours in numerical experiments).
- Fuel cost and driver wages are the major portion of the operational cost. However, there are other miscellaneous cost components such as insurance premiums and maintenance. Given that we consider dwelling and lateness cost separately, it is fair to set the operational cost equal to $1.10 per mile (empty/loaded) and revenue to $2.25 per loaded miles. The earned revenue per mile also conforms to the TRANSCORE survey in 2011 from 600 small carriers. The 2:1 ratio of revenue to cost is also supported by the work of Gregory and Powell (2002).
- The overall length of the planning horizon highly depends on the average speed of the transportation mode. The slower mode of transportation usually requires a longer overall planning horizon. For example, Choong et al. (2002) considered 15-day and 30-day planning horizon in empty container management in which a barge was one of the transportation modes. A shorter planning horizon (20-day) was considered in the truckload industry by Tjokroamidjojo et al. (2006). We consider a three-week planning horizon with each period length equals to 12 hours. Illustration of time elements of the model is depicted in Figure 2.1. For example, period 2 starts at τ=12 and ends at τ=24. ALI/ knowledge window represents how far in advance the dispatcher accesses to load information including the earliest availability, pick-up, and delivery locations. If we assume that the current decision epoch is the beginning first period (τ=0) and the

knowledge window is set to two periods, the dispatcher decides about the assignment of the first five loads to the trucks given their current attributes. The next decision epoch is the beginning of the second period when the carrier receives the loads information until the end of period 3 (i.e., loads 6 and 7). The only issue that we need to fix is the beginning and the end of horizon anomalies. For example, in Figure 2.1, the decision about loads 1 and 2 must have been made before period 1. Moreover, there is no load information for period 43 when we are in period 41. To deal with this issue, the problem is handled for the entire planning horizon while only the middle two weeks solution (Day 4 to Day 17) is considered for further analysis in this chapter.

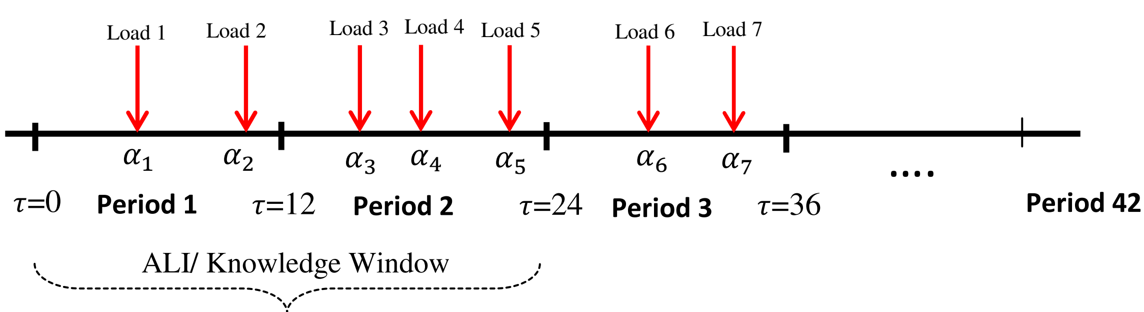


Figure 2.1. Illustration of the model's time elements

2.4.2 Dynamic Implementation of the Model

We used AIMMS modeling language and Gurobi 5.1 as a solver to run the test problems. The whole algorithm, which was explained through the simple example in the previous section, was programmed in MATLAB 2012b and run on a 2.8 GHz computer. As shown in Figure 2.2, the algorithm starts with setting ALI or knowledge window. The clock is set equal to zero and the preprocessing engine is called to update truck and load status and exclude infeasible schedules. Then, the loads with status 1, 2, and 3 are entered the model. In other words, the loads that have been already delivered and the ones that are far in future (i.e., beyond the knowledge window) are not included in the model. The next step is to call the solver to handle the proposed MIP model to optimality. After the model is solved, the obtained schedule is implemented up to the next interval and checked for the termination condition (i.e., whether all the loads are considered during the overall planning horizon). It is important to note that we need to record all the movement of trucks during the planning horizon since diversion of empty trucks is allowed. After the stopping criterion is satisfied, a simple algorithm tracks each truck's contribution to compute the system total profit for the middle two weeks of the study.

The simplified example in Figure 2.3 and Tables 2.3 and 2.4 clarifies how the static MIP is implemented in a dynamic context. In this example, radius of service, trip-length, and fleet size are at the low level while load density is high. For ease of exposition, we assume a 10-city network in which the depot is located at the center. Table 2.3 provides information about loads including their pick-up location (origin), delivery location (destination), the earliest availability, trip length (expressed in hours), and their status at each decision epoch. In this example ALI is set equal to 4 periods (48hrs); this means that the decision maker has information of the first eight loads at τ=0. Table 2.4 shows the trucks' attributes at the beginning of the first period.

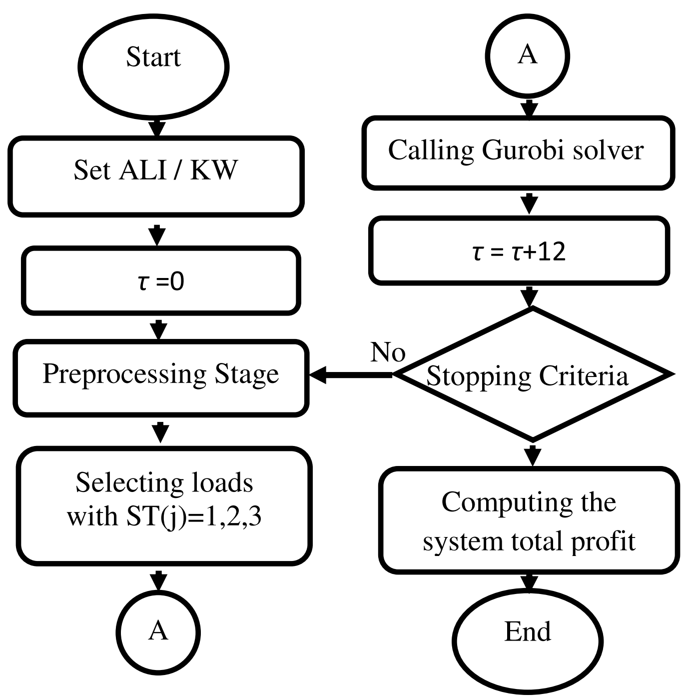


Figure 2.2. The detail of the dynamic implementation

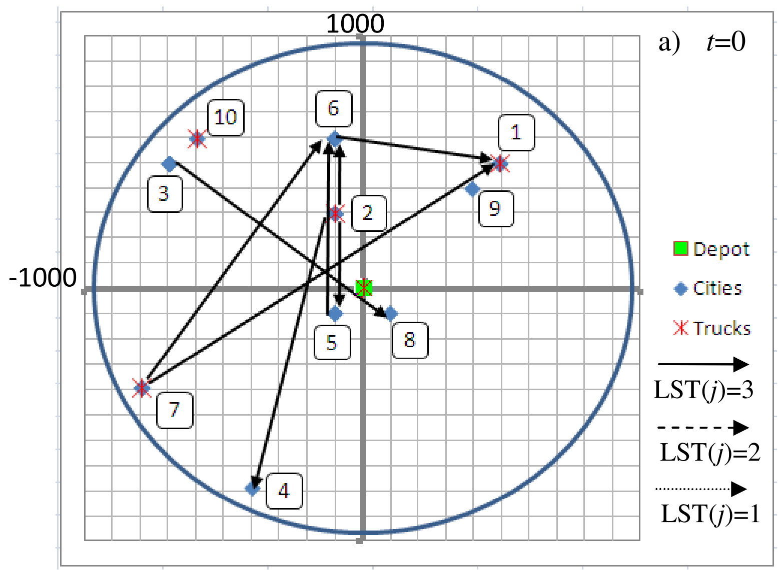
Figure 2.3 (a) shows the location of cities and initial location of trucks in a service area with radius of almost 1000 miles (18 hours of driving with 55 mph speed). Table 2.4 and Figure 2.3 show that the first two trucks are at the depot and trucks 3, 4, 5, and 6 are at cities 2, 10, 1, and 7, respectively. After preprocessing, the model is solved with 6 trucks and 8 loads at the beginning of period one (τ=0). As explained earlier, infeasible schedules can be excluded at preprocessing stage. It is important to note that rejection decisions are not made at this stage. Instead, the focus is on easy identification and removal of infeasible schedules before calling the solver. For example, it is easy to check which load(s) cannot be scheduled after load 1 on a same truck. Since the earliest availability of this load is 5.1 and its trip length is 28.5, the earliest drop off would be at time 33.6 in City 1. Given the earliest availability of other loads (i.e., loads 2 to 8), their time windows, and the traveling times between their pick up locations and City 1, none of the current loads can be scheduled right after load 1 (i.e., the truck does not go back to the depot before serving the next load). Following a similar approach, infeasible truck-load combinations can be identified. For example, truck 6 (currently is available at City 7) cannot serve load 2 because it is too far from the pickup location of that load (i.e., City 2). In section 2.5.1, we will examine the efficiency of preprocessing stage in finding the optimal solution at each decision epoch. After updating the status of loads and trucks and excluding infeasible loads, the solver is called to solve the model. Based on the structure of mixed-inter programming model, the loads are rejected because they are either non-profitable or infeasible. By solving the model, loads 4 and 5 are rejected. The truck 3 is scheduled to pick up load 3 and then load 8 directly. Truck 2, which is at the depot, is scheduled to pick up load 7. Trucks 4, 5, and 6 are scheduled to directly pick up loads 2, 6, and 1, respectively.

Table 2.3. Loads' attributes for the first 5 periods (60 hours) with ALI=48hrs

Load No.	Ori.	Dest.	Earliest Availability	Trip Length	LST(j) at t=0	LST(j) at t=12
1	7	1	5.1	28.5	3	2
2	2	5	6.4	7.2	3	2
3	2	4	12.1	20.5	3	1
4	6	1	21.5	11.0	3	0
5	2	6	24.6	5.4	3	0
6	3	8	28.7	18.0	3	1
7	5	6	33.5	12.6	3	1
8	7	6	39.9	25.8	3	1
9	3	2	48.1	11.4	No Info.	3
10	3	9	49.6	19.9		3
11	2	1	50.8	11.4		3
12	1	10	56.6	19.9		3
.	.	.	.	.		No Info.

Table 2.4. Trucks' attributes for the first 2 decision epochs

Truck No.	Remaining Hours (n_i)	TST(i)	(X-Y) Truck Position	
Decision epoch t=0				
1	240	0	0	0
2	240	0	0	0
3	125	-1	-108	324
4	37	-1	-648	648
5	174	-1	540	540
6	180	-1	-864	-432
Decision epoch t=12				
1	240	0	0	0
2	240	0	0	0
3	113	-1	-108	324
4	25	1	-108	234
5	162	-1	247	540
6	168	1	-519	-193



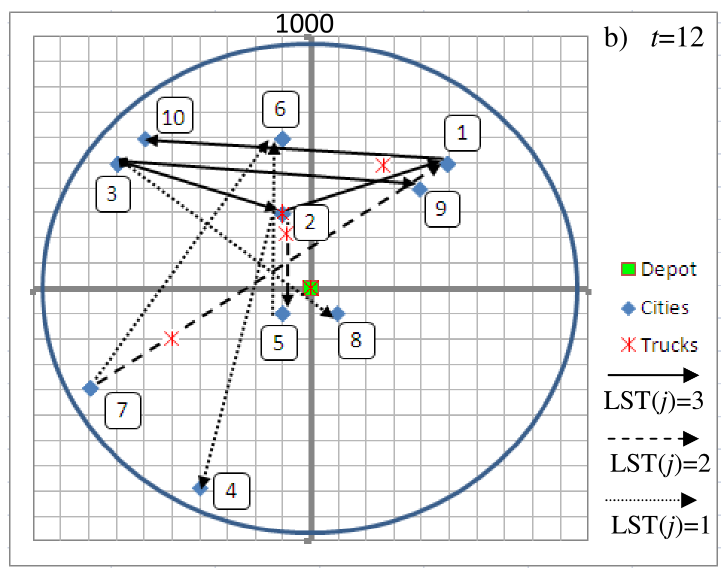


Figure 2.3. Loads status and trucks locations at the first and second decision epochs

After running the MIP model, it is the time to implement the obtained schedule until reaching to the next decision epoch. Based on the defined MIP, dwelling time is the time spent by a driver at the pick-up location of a load. It is worthwhile to note that since the dwelling cost is the same at all cities, the exact same schedule can be implemented with trucks waiting at the delivery location of loads. In static case, there is no difference between waiting at pick-up or delivery location of loads but in dynamic environment waiting at the delivery location helps to do the future re-sequencing and re-assigning more efficiently (Pillac et al., 2013). For example, if truck 5 at city 1 is dispatched immediately to pickup load 6 at τ=0, it needs to wait at city 3. In implementation, truck 5 is kept at city 1 and dispatched in a way that reaches to load 6 at its earliest availability. The next decision epoch is the beginning of period two (τ=12hrs) when four new loads enter the system. Before calling the MIP model, a preprocessing is required to exclude the infeasible assignments and update the status of trucks and loads. Seen from Table 2.3, the status of the first two loads is equal to 2 (depicted with broken lines in Figure 2.3 (b)) meaning that they are being served at the decision time (τ=12hrs) while loads 3, 6, 7, and 8 are waiting for service (i.e., their status is equal to 1 and depicted with dotted lines). The status of new loads is set to 3 and shown with solid lines. The locations of trucks are also illustrated in Figure 2.3 (b) at the decision time. The same procedure is repeated for the entire planning horizon.

2.5. Conducting Numerical Experiments and Statistical Analysis

In this section, details of numerical experiments are presented. We first point out some intuitive results. Second, an in-depth analysis is done by conducting statistical tests. Finally, further analyses are carried out regarding some factors that remain the same throughout all numerical experiments.

2.5.1 Numerical Results

Table 2.5 provides the detail of numerical experiments for all 48 combinations. The first three columns show all possible combinations of numerical design. The first column divides the test problems based on number of trucks. The second one represents other characteristics of test instances based on radius of service, trip length, and load density (L for Low, and H for high) and the third column refers to ALI values. The average profits per truck and rejection rates (over a two-week period) are also included. As seen from this table, the maximum profit belongs to a 20-truck company operating, with high level of advance load information, in a large service area where the majority of loads are long (i.e., high revenue loads) and load density is high. Not surprisingly, the lowest profit is obtained on the other extreme side where all factors have low values.

Rejection rate will be at the highest level when a small trucking company, with low-revenue load (i.e., short trip length) and low load density, operates in a large geographic area (i.e., large radius of service). This is where access to the second-day load information results in a remarkable improvement in average profit and rejection rate. The result is intuitive because any mistake in decision making due to lack of information is most likely to cost the company a considerable amount of money because of huge empty repositioning miles. As most of the useful insights are not readily evident in Table 2.5, an in-depth statistical analysis is discussed in the next sub-section.

Before moving to that discussion, it is worthwhile to have a quick look on the performance of the proposed algorithm. The sixth column (PP time) represents the CPU time to exclude infeasible schedules at the preprocessing stage. Since we target trucking companies with 20 trucks and fewer, the preprocessing stage at each decision epoch will not exceed fraction of a second. To have a better understanding of preprocessing effectiveness, we turn our attention to the second and the third last columns of the table. The first observation is that CPU times at each decision epoch are very tiny for the smallest problems sizes (i.e., 6 trucks with low load density). This is where the preprocessing stage not only lacks efficiency in finding the optimal solution but also may negatively impact the CPU time by a very small amount. However, the importance of the preprocessing stage becomes more evident when the size of the problem grows (e.g., 6 trucks with high load density or 20 trucks). The largest observed improvement is 52.9% in presence of 72 hrs advance load information for a 20-truck company with a low service radius, high trip length, and high load density network. It is also interesting to note that the test problems with a low radius of service, low trip length, high load density, and larger fleet size are the most difficult problems to solve when knowledge window is 6 periods (72hrs). It takes on average 604 seconds (or 1120 without preprocessing) to solve each problem to optimality at each decision epoch which translates to an average of 6 hours (or over 11 hours without preprocessing) to solve one test problem in such a setting.

Table 2.5. Details of numerical studies

# of Trucks	Code	ALI	Profit	Load Rejection	PP time (Sec)	CPU time with PP	CPU time without PP	% of Improvement
6 trucks	HHH	24	3624.60	76.5%	0.026	0.39	0.42	7.1%
		48	4331.58	64.3%	0.028	0.49	0.52	6.6%
		72	5496.75	63.2%	0.032	0.56	0.63	11.9%
	LHH	24	4483.08	36.6%	0.031	0.44	0.44	0.8%
		48	5039.62	34.5%	0.032	0.51	0.51	0.6%
		72	5218.25	33.6%	0.036	0.61	0.62	1.6%
	LLH	24	1165.39	71.8%	0.031	0.36	0.41	11.4%
		48	1251.08	68.1%	0.032	0.51	0.63	19.5%
		72	1298.13	68.0%	0.037	0.73	0.90	18.6%
	HLH	24	1326.35	74.8%	0.027	0.29	0.30	3.3%
		48	2283.13	67.0%	0.031	0.46	0.48	4.2%
		72	2356.86	66.8%	0.035	0.60	0.67	10.4%
	LHL	24	1493.74	45.4%	0.017	0.34	0.32	-
		48	1847.94	37.8%	0.021	0.53	0.50	-
		72	2061.22	37.3%	0.030	0.65	0.59	-
	LLL	24	145.40	77.4%	0.018	0.11	0.09	-
		48	236.36	76.4%	0.023	0.14	0.12	-
		72	266.38	75.8%	0.034	0.22	0.19	-
	HLL	24	497.88	85.6%	0.018	0.09	0.07	-
		48	1047.53	72.2%	0.023	0.10	0.08	-
		72	1274.63	70.9%	0.035	0.14	0.11	-
	HHL	24	1706.86	70.8%	0.021	0.11	0.09	-
		48	2449.78	57.7%	0.023	0.13	0.11	-
		72	2699.10	57.1%	0.027	0.14	0.11	-
20 trucks	HHH	24	7113.92	65.8%	0.209	3.52	3.68	4.2%
		48	8547.43	58.3%	0.226	5.49	5.84	6.1%
		72	8787.16	58.8%	0.254	8.96	9.84	8.9%
	LHH	24	7795.84	23.7%	0.197	3.50	4.52	22.6%
		48	8144.84	20.3%	0.214	4.31	6.26	31.2%
		72	8427.86	19.1%	0.239	176.00	374.00	52.9%
	LLH	24	2124.02	60.0%	0.205	4.13	4.31	4.2%
		48	2323.53	55.9%	0.220	68.41	70.92	3.5%
		72	2442.62	55.8%	0.245	604.00	1120.00	46.1%
	HLH	24	3461.38	69.6%	0.195	3.59	4.00	10.1%
		48	3883.94	64.8%	0.209	4.49	5.10	12.0%
		72	4190.88	67.2%	0.228	18.14	20.74	12.5%
	LHL	24	3505.83	22.5%	0.190	4.19	4.41	5.1%
		48	3831.57	17.2%	0.210	5.42	5.78	6.2%
		72	3847.20	16.3%	0.233	102.00	112.67	9.5%
	LLL	24	516.34	66.7%	0.170	3.98	4.06	2.0%
		48	522.21	63.8%	0.184	5.12	5.24	2.3%
		72	530.69	63.0%	0.203	9.47	9.80	3.4%
	HLL	24	1192.00	72.5%	0.165	3.91	3.95	1.0%
		48	2069.43	60.0%	0.176	4.16	4.30	3.3%
		72	2077.91	58.8%	0.191	4.62	4.78	3.4%
	HHL	24	3075.21	64.1%	0.190	3.69	3.76	2.0%
		48	5216.20	37.1%	0.206	5.38	5.55	3.2%
		72	5296.12	36.0%	0.227	6.03	6.23	3.2%

2.5.2 Statistical Analysis

After solving all the test problems, a linear regression model is used to statistically test the impact of advance load information on a carrier's profit. The dependent variable is average profit per truck during the middle two weeks of the planning horizon. The model comprises the five factors in the experiments. For categorical variables, the effect coding is used to make them appropriate to be used in the linear regression. As the main focus of our study is to examine the benefit of ALI, we control the impact of other factors (radius of service, trip length, density, and fleet size) and their interactions. Thus, they are first entered in the model (refer to model 1 of Table 2.6). Then, at the next step, the ALI factor and its interactions with the other factors are included into the regression model (model 2 of Table 2.6). The obtained results illustrate that a great portion of variation in profit (around 95%) is explained by the main factors and their interactions. Moreover, it indicates that ALI and its interactions can explain the variations in profit by 2.7% over and above all other transportation factors.

Table 2.6. Model summary (dependent variable is average profit per truck)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.960[a]	.923	.917	706.57
2	.974[b]	.949	.944	582.36

a. Predictors: (Constant), Density, Trip Length, Radius, fleet size and all the interactions

b. Predictors: (Constant), the predictors of the first model plus ALI and all the 2-way interactions with the other factors

The details of statistical tests are depicted in Table 2.7. It shows that ALI and its interactions with radius of service and trip length are statistically significant at the 5% level. Figure 2.4 clarifies the main impact of acquiring advance load information and its interactions with the other factors on a carrier's profit. In Figure 2.4, Y-axis of each chart represents the average profit per truck during the two weeks when data are collected. Figure 2.4 (a) shows that access to the second-day load information boosts the profit by an average of 3314 – 2702 (by 22%) compared to only having first-day load information. This suggests that considerable savings can be obtained by accessing the first two days of load information. Obtaining an additional day's worth of ALI (i.e., from a two-day knowledge window to a three-day knowledge window) improves profit further by a much smaller margin of almost 6%, indicating that increases in the knowledge window yields a decreasing rate of profit improvement. Since the interaction effects between ALI and radius of service and trip length are statistically significant, we now turn our attention to those effects.

Table 2.7. Coefficients test of the regression model for the average profit per truck

Regression Model	Standardized Coefficients	t	Sig.
	Beta		
(Constant)		23.761	.000
Radius	-.040	-.975	.331
Trip Length	.525	12.940	.000
Density	.463	11.430	.000
Size	.390	9.610	.000
Size* Radius	.057	3.736	.000
Size* Trip Length	.181	11.804	.000
Size* Density	.115	7.513	.000
Radius* Trip Length	-.084	-5.482	.000
Radius* Density	-.034	-2.227	.027
Trip Length* Density	.192	12.556	.000
Size* Radius* Trip Length	-.011	-.697	.486
Size* Radius* Density	.012	.787	.432
Size* Trip Length* Density	.026	1.715	.088
Radius* Trip Length* Density	-.048	-3.149	.002
Size* Radius* Trip Length* Density	-.001	-.087	.931
ALI	.136	8.847	.000
Radius*ALI	.190	4.681	.000
Trip Length*ALI	.129	3.171	.002
Size*ALI	-.001	-.013	.990
Density*ALI	.029	.717	.474

Figure 2.4 (b) depicts the benefit of advance load information at different radii of service. It is intuitive to see the benefit grows as the radius of service becomes larger. Since the nodal density is lower in a larger service area, the repositioning of trucks becomes more important not only to reduce relevant costs but also to take advantage of upcoming loads. Thus, it is extremely helpful for the dispatcher to have advance load information while service radius grows. Although it is still beneficial to acquire more than one-day load information, there is very little benefit to obtaining more than two-day ALI. With respect to statistical analysis, the interaction of ALI and trip length is also significant which is also illustrated in Figure 2.4 (c). As seen from this figure, the advance load information has different levels of benefits respect to the trip length variations. It is noticeable that there is remarkable benefit to collecting more than two-day load information if average trip-length is typically long. Figure 2.4 (d) confirms the results from statistical analysis that the benefit of advance load information does not show different behavior by changing the fleet size. However, the observation may suggest the third-day load information can provide more benefit for smaller companies than larger ones since the slope of the graph is slightly steeper. Based on the statistical result in Table 2.7, the interaction effect of ALI and load density on the carrier's profit is not statistically significant. One explanation for this finding is that higher load density may signal an already profitable market for a carrier, leaving very little additional profit to be gained by acquiring ALI. However, a carrier can still benefit from advance load information even when load density is high. Thus, we need to concentrate on the main sources of benefits from accessing to ALI. These sources are: 1) accepting more profitable loads or 2) rejecting fewer loads or 3) serving loads in more efficient ways 4) or any combination of them.

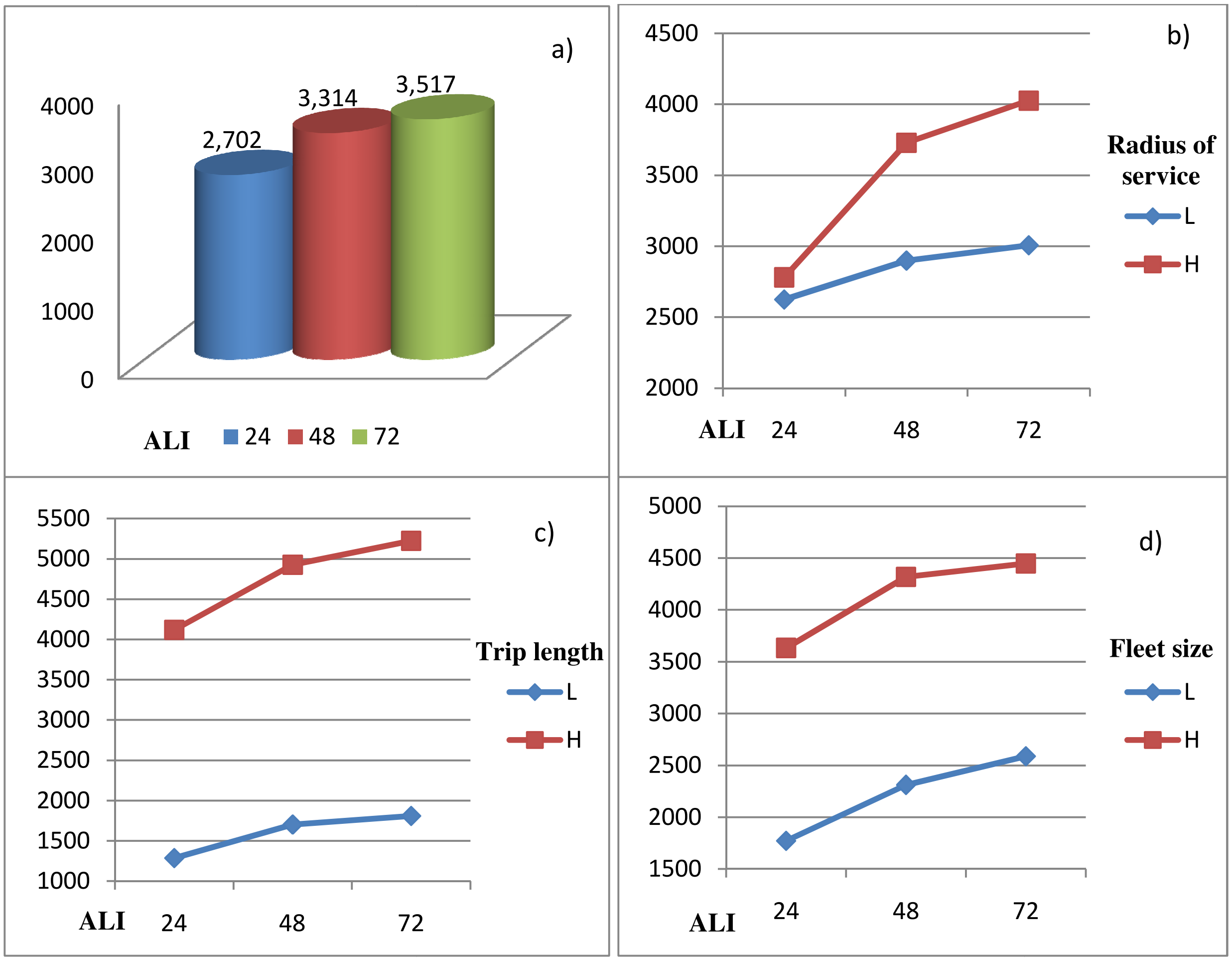


Figure 2.4. Impact of ALI and some of its interactions with the other factors on the profit

To figure out which sources contribute to the improvement of carrier's profit, a similar statistical analysis with rejection rate as the response variable is conducted (see Table 2.8). The results show that advance load information and its interactions with radius of service and load density have significant impact on the rejection rate. Regarding the ALI effects on rejection rate, a quick look to Figure 2.5(a) reveals that there is very little benefit (in terms of lowering the load rejection) by obtaining additional information if we have already access to the next two days load information. Based on the statistical analysis and our observation, we can conclude that the initial benefit of ALI (i.e., accessing to the second-day load information) is substantial because it helps with simultaneously reducing the rejection rate and serving loads more efficiently.

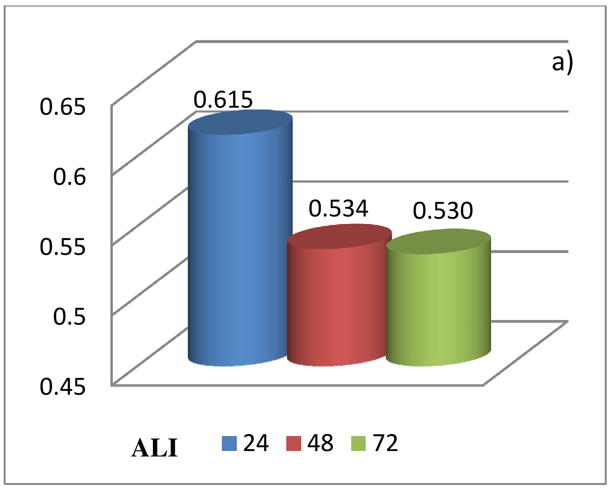
Moreover, any additional information (i.e., obtaining the third-day load information) is still significant but less than the initial impact because the major improvement comes from serving loads more efficiently, not from further lowering the rejection rate.

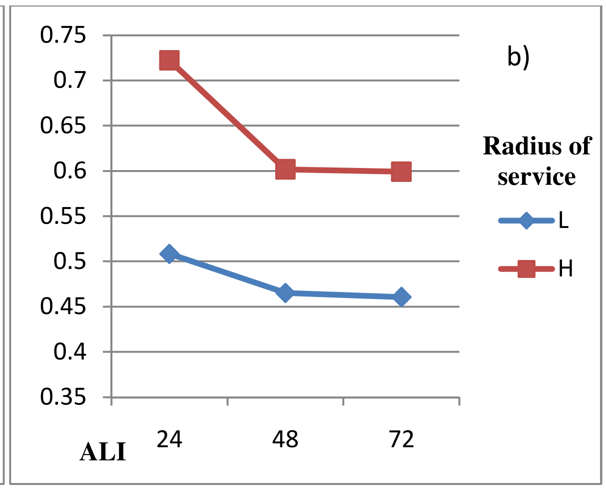
Table 2.8. Coefficients test of the regression model for rejection rate

Regression Model	Standardized Coefficients	t	Sig.
	Beta		
(Constant)		56.358	.000
Radius	.611	10.418	.000
Trip Length	-.550	-9.383	.000
Density	-.135	-2.307	.022
Size	-.330	-5.629	.000
Size* Radius	.043	1.934	.054
Size* Trip Length	-.043	-1.936	.054
Size* Density	.062	2.775	.006
Radius* Trip Length	.331	14.922	.000
Radius* Density	.139	6.278	.000
Trip Length* Density	.094	4.248	.000
Size* Radius* Trip Length	.033	1.499	.135
Size* Radius* Density	.014	.614	.540
Size* Trip Length* Density	.047	2.117	.035
Radius* Trip Length* Density	.020	.884	.378
Size* Radius* Trip Length* Density	.001	.031	.975
ALI	-.178	-8.051	.000
Radius*ALI	-.209	-3.559	.000
Trip Length*ALI	-.099	-1.680	.094
Size* ALI	-.008	-.138	.890
Density*ALI	.155	2.646	.009

Another interesting observation can be found in Figure 2.5 (b). It shows that ALI can reduce the rejection rate but the improvement margin depends on the radius of service. The impact on rejection rate is greater when the service radius is longer. Note also that the graph depicts the recurring theme that most of the improvement comes from having two days worth of ALI: having three days worth yields minimal further improvement.

Figure 2.5 (c) illustrates an intuitive result. It shows a larger impact of advance load information when load density is lower. This is justified by the fact that more knowledge about the upcoming loads typically provides more alternatives for the decision maker either with low or high load density. However, this impact is lower when load density is high because enough profitable alternatives are already available to the decision maker so the need for more advance load information is less comparing to the low density case.





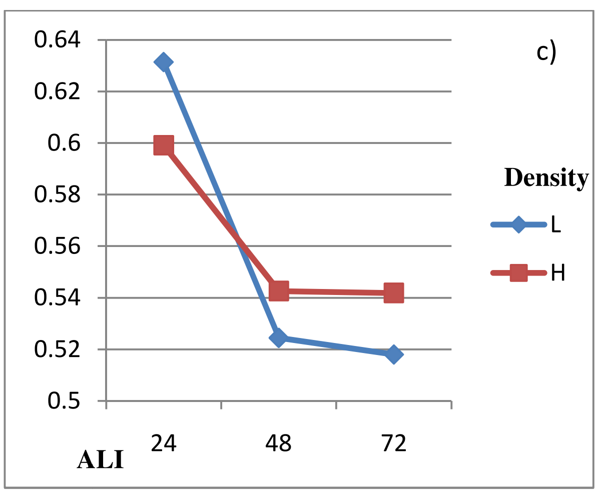


Figure 2.5. Impact of ALI and its significant interactions on rejection rate

2.5.3 Further Analyses

2.5.3.1 Network Size Impact

In all conducted numerical experiments, it was assumed that there are 50 cities within an area with a predefined radius of service. The number of cities in transportation network is referred to as network size. These potential cities are representative of loads' origins and destinations. This approach is commonly used in the literature. For simulation studies, some authors consider that the shape of the service area is square (e.g., Yang et al. 2004; Özener and Ergun, 2008). To capture radius of service, as one of the transportation network settings, we assumed that the trucking company operates in a circle-shaped area. The selection of 50 cities as potential pick-up and drop-off locations is also supported by some practical cases. Consider for example, Ontario: a major Canadian province comprised of 47 metropolitan areas (see the Statistics Canada report at http://www12.statcan.gc.ca). Those 47 metropolitan areas are where shippers are mainly located and they comprise the service region of the majority of the province's 700-plus small trucking companies (details on the services and home locations of these companies can be found at www.CanadaTrasportation.com).

Although the choice of network size is justified from both academic and empirical aspects, it is still interesting to know about the benefit of advance load information when number of potential cities in transportation network varies. In this regard, the network size was changed from 50 to 10 cities (also used by Tjokroamidjojo et al., 2006). The same approach as described in sub-section 2.4.1 was used to come up with 240 new test instances in which the underlying network has fewer cities. It is trivial to see lower nodal density for the test problems where other factors remain the same. Since nodal density directly impacts the average distance between nodes (cities), the average distance between cities will increase with fewer nodes in the transportation network. Figure 2.6 illustrates the improvement percentage of a carrier's profit by obtaining the second-day and the third-day load information. The overall trend is the same in the sense that the majority of benefit is obtained by receiving the second-day load information. However, the percentage of improvement in acquiring second-day load information is about 45% which is remarkably larger than the 22% for a network size of 50. Since a major benefit of ALI is reduction of empty vehicle repositioning, ALI will be more beneficial for carriers operating in networks where inter-city vehicle repositioning distances are longer because there are fewer cities. Although it is still beneficial for the carrier to access the third-day load information, the margin of benefit declines to almost 6% regardless of the network size.

2.5.3.2 A Real-World Case Study

As mentioned earlier, we assumed that the service area has a regular shape (i.e., circle-shape area) and cities are uniformly distributed within the service area. This may raise a question about the magnitude of advance load information benefit where those assumptions are violated. To briefly address this issue, we study the benefit of advance load information for a small trucking company with 6 tractors in Canada. This company operates in the truckload industry with the home base (depot) in Toronto. This company operates within the province of Ontario where the pick-up and drop-off locations are metropolitan areas (depicted in Figure 2.7). Seen from this figure, the cities are not uniformly distributed within the service area (i.e., majority are located in southern part of province). Among a wide range of network settings that was used in simulation study, the setting with low radius of service, low trip length, and high load density (coded as LLH) is the closet to the real case example. The performance of trucking company expressed in percentage of improvement by acquiring the second- and third-day load information is illustrated in Figure 2.8. Seen from this figure, the 6.22% profit improvement from the second day ALI is just slightly smaller than the 7.35% realized in the simulated LLH setting. When it comes to the third-day load information, the margin of benefit becomes comparatively less (3.76% for LLH compared to 1.24% for the real-world case study). The lower benefit of ALI in such a setting can be explained as follows. The general service area is the whole province but most cities are located in the southern part of the province. Consequently most transportation moves occur in a relatively much smaller service area, a phenomenon which has been shown to yield statistically significant reductions in the benefit of advance load information.

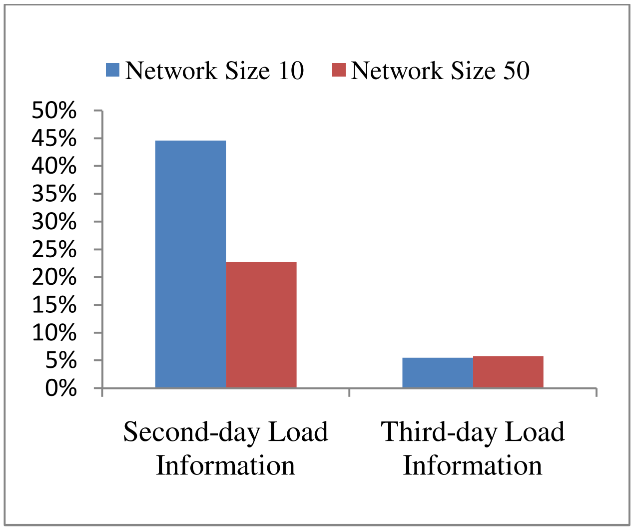


Figure 2.6. Network size impact (% of improvement)

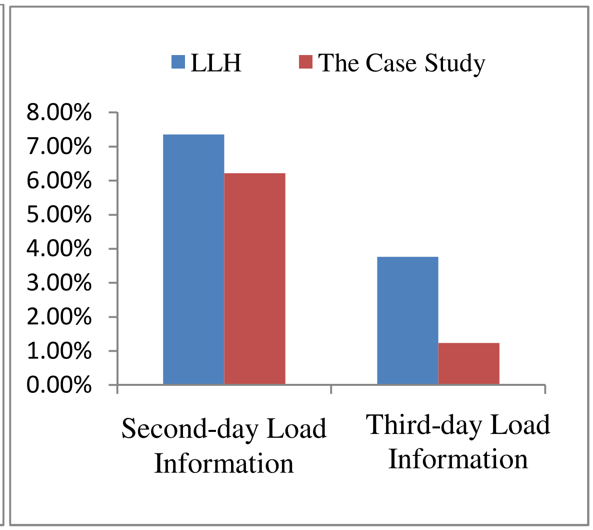


Figure 2.8. Dispersion impact (% of improvement)

Figure 2.7. The major cities of transportation network in Ontario, Canada (source: satellite view of Google map)

2.6. Conclusion and Future Research Directions

There are many research studies on long haul transportation dispatching rules that did not address the requirement of drivers and trucks to regularly return to their domiciles. This overestimates the capacity of transportation network. Moreover, the majority of them assume that all loads information is available in advance. Thus, many of these models are not suitable to be implemented in a dynamic context. One contribution of this study is that it develops a comprehensive MIP model that is flexible enough to include many operational details and can be easily implemented in a dynamic environment by using a rolling horizon approach. The model's flexibility is unaffected by our retention of the literature's standard assumption that each trip is executed without a break (e.g., Powell, 1987; Powell et al. 1988; Powell, 1996; Yang et al., 1998; Gronalt et al., 2003; Yang et al., 2004; Tjokroamidjojo et al., 2006; Özener et al., 2011). This assumption can be justified when a team of drivers is responsible for serving loads (Simao et al., 2009). Moreover, in case of a single driver, the parameters of our presented model can be modified to produce reasonable dispatching recommendations. For example, instead of computing travelling time as a linear function of distance, the traveling time between two cities can include all the rests components that a driver must have based on the driving rules and regulations.

Given the paucity of research studies on information sharing in the transportation field (as compared to the robust body of such studies in the inventory management field), this chapter's other major contribution is to examine the benefit of advance load information in the truckload industry. In this regard, a comprehensive set of numerical experiments covering five factors is designed. The obtained results illustrate that access to the second day loads information can improve the profit by an average of 22%. The benefit can be further improved by acquiring more information but the margin decreases to 6%. Moreover, the impact of ALI depends on the other transportation network settings. For example, the impact of ALI on a carrier's profit is greater when the majority of carrier's loads are long or the carrier is operating within a large service area. The rejection rate can be also reduced by accessing loads information further in advance. Most of that reduction is achieved by having the second-day loads information. The improvement becomes trivial if the carrier collects the third-day information. It is also important to note that the improvement in rejection rate depends on radius of service and load density. The benefit (in terms of lowering the rejection rate) becomes larger if the radius of service grows. The rejection rate also improves to a greater extent when the load density is lower.

The current work can be extended in various directions. For example, in the presented model, it is assumed that when a truck returns to the home domicile, it is immediately ready for the next 240 hour trip. This is true only if there is a backup driver who can take responsibility of the incoming truck. Modifying the presented MIP model to capture different real-world operating policies can be viewed as an interesting research venue.

In practice, the home depot is usually close the areas with more demands (loads). Since there is one home depot in our simulation study (it is also usual for small trucking companies (with 20 trucks and fewer) to have only one depot and loads are uniformly generated within the area of service, it is a reasonable assumption to consider the location of the depot at the center. Finding the optimum location of depot is beyond the scope of this work and can be viewed as a future research direction.

Another possible research direction is to address information uncertainty since loads information (e.g., pick-up time or cancellation) may change even after it is received by the carrier. In this study, it is also assumed a constant traveling time which can be relaxed to consider trucks breakdown or possible accidents. One of the other interesting extensions is to test the impact of loads distribution over the planning horizon because it is typical to have more requests earlier in a week than the weekends.

CHAPTER 3

EFFECTIVE TRUCKLOAD DISPATCH DECISION METHODS WITH INCOMPLETE ADVANCE LOAD INFORMATION

3.1. Introduction

Two issues loom large for carriers in the truckload industry as they undertake efforts to assure prosperity and survival in the ongoing economic recession: (i) asset repositioning and (ii) driver turnover. Asset repositioning, which has been studied by, e.g., Crainic (2000); and Wieberneit (2008), is due to natural characteristics of truckload transportation networks such as demand dynamism and network imbalance between supply and demand. Ergun et al. (2007a) reports that empty movement of trucks costs U.S. carriers nearly 165 billion dollars annually. Based on the American Trucking Association (ATA) 2013, the ratio of empty to total mileage is usually higher for small carriers (22%) with a sparser network of lanes than larger ones with a more sophisticated lane network (17%).

The issue of driver turnover is strongly influenced by drivers' dissatisfaction with work schedules requiring overly long periods away from home. Studies confirming this include Rodriguez and Griffin (1990), Shaw et al. (1998), Keller (2002), and Suzuki et al. (2009). The driver turnover problem is significant (according to the Council of Supply Chain Management Professionals (2006), it can reach 130% in a year) and costly: the replacement cost of a driver (e.g., including training and loss of experience) is estimated to cost between \$2,200 to over \$20,000 with an average of \$8000 (e.g., Rodriguez et al., 2000). Given the size of the U.S. trucking industry, driver turnover translates to approximately three billion dollars a year (Suzuki et al., 2009).

To correct for these issues, a commonly used strategy is collaborative transportation (CT); e.g., CT networks such as Nistevo (www.nistevo.com) and Transplace (www.transplace.com). In CT, logistics participants (i.e., shippers/consignees and carriers) collaborate to improve transportation performance; e.g., reduce total transportation costs and driver turnover and increase truck utilization (Ergun et al., 2007b). Collaboration could be among transportation clients (e.g., Ergun et al., 2007a), among carriers (e.g., Özener et al., 2011), or between client(s) and carrier(s) (e.g., Tjokroamidjojo et al., 2006) or all the above scenarios.

The focus of this study is the collaboration between a carrier and its clients. One of the least costly methods when freight transportation service clients and carriers collaborate with each other is to communicate timely load information (from clients to carriers). Although sharing advance load information (ALI) can improve the carrier's performance by expanding its knowledge window (KW) into the future (Powell, 1996; Tjokroamidjojo et al., 2006), there is always uncertainty after the KW (Caplice and Sheffi, 2003).

In the absence of exact information about future loads beyond the knowledge window, the dispatcher's range of decisions (load acceptance/rejection, load sequencing, etc.) is influenced by *the matter of where the truck will be positioned for serving future (unknown) loads*. Consider two extreme options open to the dispatcher in deciding which known loads the truck should be assigned to:

i. the conservative policy of preferring loads that take the truck close to its domicile; i.e., to avoid large empty truck repositioning costs to the domicile (called deadheading costs in this study) when the truck must eventually return deadhead to the domicile.
ii. the more optimistic policy of making truck-load assignments with greater risk of large deadheading costs in the hope that those assignments will put the truck in a better position to access highly profitable future (unknown) loads.

From the above, it is clear that in a given context (load density, radius of service, etc.), and for a given truck at a given instance of time (e.g., current and imminent truck location vis-à-vis its domicile), the following is true: a significant factor in what policy the dispatcher should choose is the deadhead cost. The dispatcher's dilemma is that the true deadhead costs can be known only *a posteriori* because that is the only time at which the exact information such as the locations, pick-up time windows, and trip lengths of future loads becomes known. To tackle the dilemma, we attempt getting an a priori *signal* of the efficacy of a dispatching policy by proposing the concept of a deadhead coefficient Θ ($0 \leq \Theta \leq 1$). In essence, the coefficient is only a *signal* of the extent to which the chosen dispatching policy might affect profits because at the time of decision making, the dispatcher, while knowing the revenue of serving loads and some of the cost components, has no information beyond the last known load to be served. Thus, the dispatcher's decision is directly influenced by the conservatism level of his/her policy, which can be portrayed by what we label as the ***Θ-dependent profit*** estimate (π_{Θ}). We calculate π_{Θ} as: total revenue - total known cost (including loaded and empty repositioning, dwelling, and lateness costs)- Θ×(travel cost from the destination of the last load in the sequence to the domicile/depot).

The basic intuition of the deadhead coefficient is as follows. First, consider using large Θ values for potential end of sequence loads. Those Θ values are associated with more conservative policies in that they raise the attractiveness of such loads with destinations close to the domicile. That is, based on the last term in the above expression π_{Θ}, those loads are *predicted* to have a smaller negative financial impact so they are more likely to be selected over alternatives that are distant from the domicile. Conversely, small Θ values lower the negative *predicted* financial effect of accepting end-of-sequence loads with destinations that are distant from the domicile. In other words, the dispatcher will lean towards selecting loads that, despite requiring the truck to be further from the domicile, have high values for the excess of revenue over known cost.

A small numerical example is presented in the next section to further clarify the above observations and the process of using the deadhead coefficient to tackle the dispatcher's dilemma of unavailable exact information (i.e., uncertainty) about future loads. As the example illustrates, different Θ values can yield different load selection decisions, and thereby may result in different values of profit. Thus, an obvious question of managerial interest is which Θ value yields the best attainable profit in a given transportation context (e.g., load density, radius of service, trip length, and time windows). Addressing this question is one of this chapter's major contributions.

In this study, we focus on three key points. We first develop a flexible dispatching mixed integer program (MIP) model that can incorporate important operational details of trucking companies (e.g., current location of trucks, number of hours that a truck is away from home, previous commitments) to make profitable decisions given different levels of advance load information. Second, a simple policy (based on the deadhead coefficient) is proposed to help dispatchers make load acceptance decisions in dynamic environments. The proposed deadhead coefficient policy is tuned based on different transportation network settings. Finally, the proposed policy is enhanced to improve the solution quality of the dynamic problem at the expense of a longer running time. To achieve the goals of this research, we briefly introduce the idea of the simple policy with one small example in section 3.2. Section 3.3 is devoted to reviewing the related literature for positioning this study among the existing works and highlighting its novelty. In section 3.4, the model assumptions, notations, and parameters are defined and the conceptual model is formulated as a mixed integer program. Section 3.5 explains how experiments are designed for conducting a comprehensive simulation study. In section 3.6, the proposed policy is evaluated through simulation results. In section 3.7, the proposed policy will be enhanced by applying sample scenario hedging heuristic proposed by Hvattum et al. (2006) for stochastic dynamic vehicle routing problems. We also examine our proposed policy and its enhanced version against two other dispatching methods. Conclusions and future research directions are provided at the end.

3.2. Proposed Deadhead Coefficient Policy: An Illustrative Example

For ease of exposition, we use the case of a single-truck carrier to illustrate how the proposed policy works with different Θ values. An underlying logic of the policy is that trucks not scheduled to serve any loads return to the depot. This policy is intuitive if the dispatcher has access to advance load information (e.g., knowing that there is no request available for the rest of the day). The logic is also sound because the average repositioning is typically shorter from the depot (if it is located at the center) and dwelling cost is much lower at the depot. This is because there is no extra facility usage cost for, say, a driver to dwell at his/her home or at accommodations provided by the carrier (e.g., Challenger Motor Freight's well-equipped rest facility for drivers at its Cambridge depot, more detail about this trucking company can be found at its official website: http://www.challenger.com). We label this policy as ***Deadhead Coefficient Policy*** because its success depends on selecting a proper Θ value. We will also refer to this as the Pure-Θ Policy.

In our illustrative example, the truck is idle at the depot (the driver's home domicile) at the beginning of the planning horizon, the dispatcher's knowledge window is set to two days (48 hrs), and system information is updated daily. The truck earns $130/hr for serving a load while incurring $60/hr when moving either empty or loaded. Without loss of the generality, dwell and lateness costs are not taken into account to make the example simple enough to follow. Figure 3.1 depicts how loads are distributed over time and revealed to the dispatcher. In Figure 3.1(a), the information of loads A, B, C, and D is available at the beginning of day one while load E will be realized when the system information is updated at the start of day 2 (Figure 3.1(b)). Figure 3.2(a) represents a 7-city transportation network showing all travelling times that are relevant to the example (depot-city and inter-city). Load E is also shown in Figure 3.2(a), but it is only known on the second day.

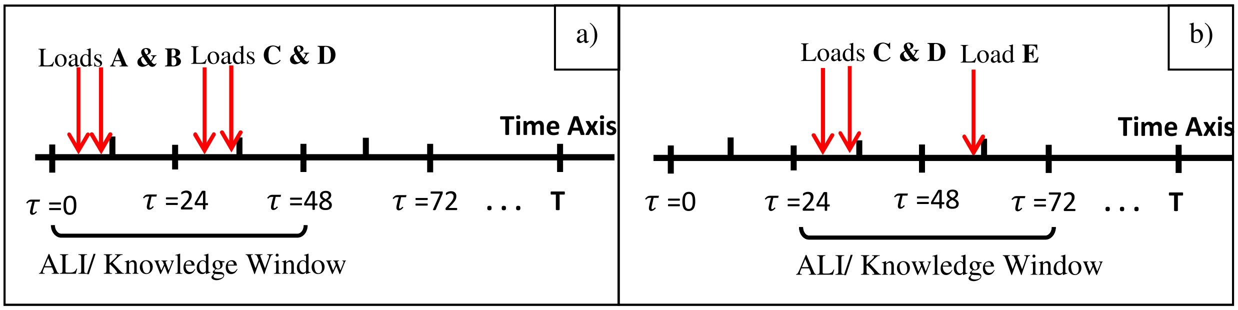


Figure 3.1. Illustration of the model's time elements at the beginning of days one and two

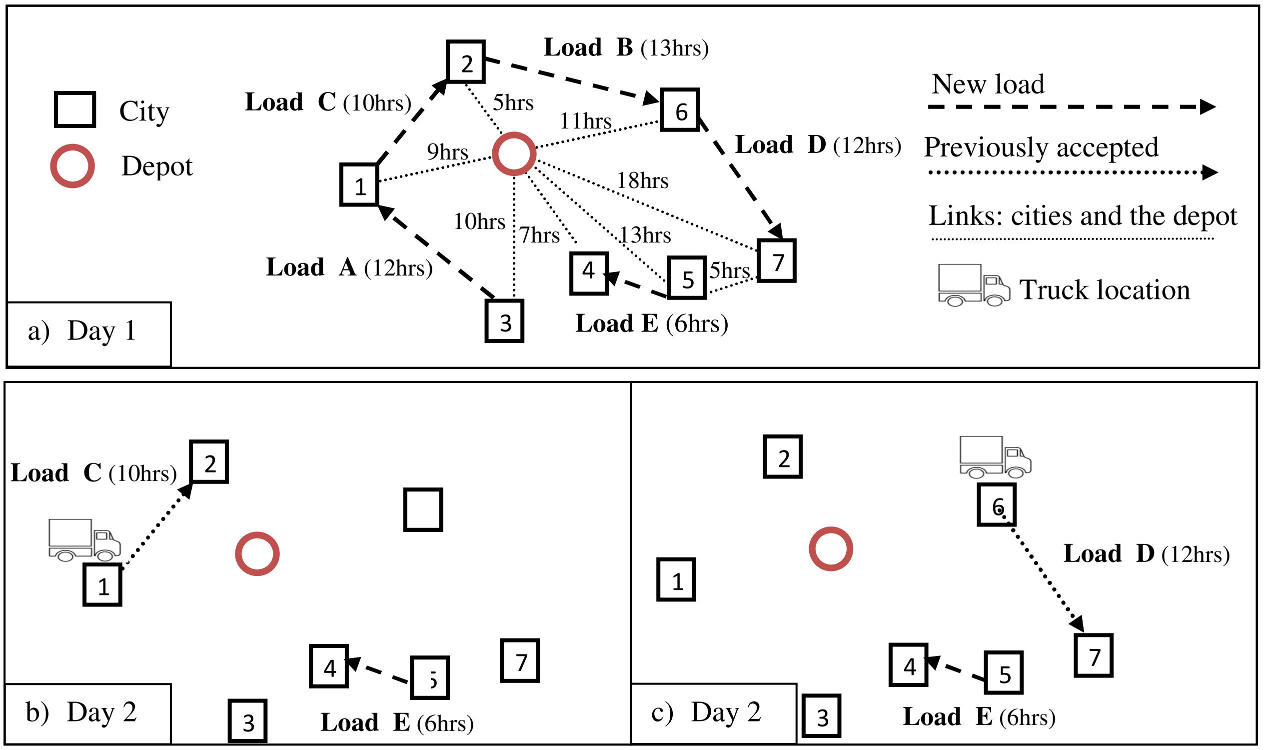


Figure 3.2. The transportation network and loads status at two decision times

Table 3.1 shows each alternative (including the sequence of loads and cities), travel time components, the ratio of weighted empty travelling time to the depot (WETTD) from the delivery location of last load, and Θ-dependent profit (π_{Θ}) for two Θ values. To briefly point out how each entry of this table was calculated, we consider the A-C sequence. In order to serve this load sequence, the truck departs the depot and visits cities 3, 1, and 2, respectively resulting in 22 hours of total loaded movement. The empty movement is 10hrs (traveling from the depot to city 3, origin of load A). WETTD is the product of the Θ value and traveling time from the delivery location of the last load (city 2) to the depot. Given the formula in section 1, $\pi_{0.8}$=130(22)-

60(22+10)-60(0.8×5)=$700. A similar approach is used to find $\pi_{0.2}$ by modifying the Θ value. As we can see from Table 3.1, different Θ values not only impact the selection of the last load of the sequence but also can change the whole sequence (i.e., Sequences B-D and A-C are selected with Θ=0.2 and 0.8, respectively). A conservative policy (Θ=0.8) results in truck repositioning at city 2, which is closer to the depot compared to a more optimistic policy (Θ=0.2) in which the truck will end up at city 7 after serving known loads.

Table 3.1. Evaluating all possible alternatives based on Θ-dependent profit criterion at τ =0

Sequence			Travel time components				WETTD/other movements		π_Θ	
PA	New	Cities	Empty (hrs)	Loaded (hrs)	WETTD (hrs)				$\pi_{0.2}$	$\pi_{0.8}$
					Θ=0.2	Θ=0.8	Θ=0.2	Θ=0.8		
-	-	-	0	0	0	0	0	0	0	0
-	**A**	Depot→C3→ C1?	10	12	1.8	7.2	0.08	0.33	132	-192
-	**B**	Depot→C2→ C6?	5	13	2.2	8.8	0.12	0.49	478	82
-	**C**	Depot→C1→ C2?	9	10	1.0	4.0	0.05	0.21	100	-80
-	**D**	Depot→C6→ C7?	11	12	3.6	14.4	0.16	0.63	-36	-684
-	**A-C**	Depot→C3→ C1→C2?	10	22	1.0	4.0	0.03	0.13	880	**700**
-	**A-D**	Depot→C3→C1→C6→C7?	30	24	3.6	14.4	0.07	0.27	-336	-984
-	**B-C**	Depot→C2→C6→C1→C2?	25	23	1.0	4.0	0.02	0.08	50	-130
-	**B-D**	Depot→C2→ C6→C7?	5	25	3.6	14.4	0.12	0.48	**1234**	586

Note: PA: Previously Accepted
WETTD: Weighted Empty travelling time to the depot

The impact of the Θ choice can be further elaborated by moving to the next decision epoch. With Θ=0.8, the truck will be at city 1 at the beginning of day 2 according to the previously designed plan to serve the B-D sequence (Figure 3.2(b)). Given the current commitment (to serve load C) and the truck location, there are only two alternatives whether to accept the new load or not. According to the computational details of Table 3.2 (the last two cells of π_Θ), the acceptance of a new load (load E) is not recommended. However, following a similar approach, the policy with Θ=0.2 will schedule load E after load D.

Having a closer look at the behavior of ***Deadhead Coefficient Policy*** in load selections at two decision epochs, two key factors are observed. First, the choice of Θ has direct impact on one of the cost components (which is depicted by the ratio of WETTD to all other movements).

This ratio is remarkable and influential with practical advance load information (e.g., two days). As an example from Table 3.1, the ratio for the load sequence B-D is 12% if Θ=0.2 but rises to a very substantial 48% if Θ= 0.8. Second, this impact is reinforced by an inherent feature of truckload transportation: economy of scope defined in Caplice (2007) as strong cost interdependency between loads (because a truck should be moved from the delivery location of one load to the pickup location of the next one), economy of scope is commonly acknowledged in works on truckload transportation (see, e.g., Chang, 2009; Berger and Bierwirth, 2010; Özener et al., 2011). In our illustrative example, this dependency, which translates to subsequent decisions being affected by earlier load selection decision, is highlighted by the following fact: selecting the load combination B-D (using Θ=0.2) leads the subsequent decision to add load E to that combination but, on the other hand, selecting the A-C combination (using Θ= 0.8) renders load E as an unprofitable addition. As mentioned earlier, our goal is to find a proper Θ value to aid carriers in improving profit over the planning horizon. The impact of Θ on a carrier's profit will be illustrated through extensive numerical experiments.

Table 3.2. Evaluating all possible alternatives based on Θ-dependent profit criterion at τ =24

Sequence			Travel time components				WETTD/other movements		π_Θ
PA	New	Cities	Empty (hrs)	Loaded (hrs)	WETTD (hrs) Θ=0.2	WETTD (hrs) Θ=0.8	Θ=0.2	Θ=0.8	
Θ=0.2									
D	**-**	C6→ C7?	0	12	3.6		0.30		624
D	**E**	C6→ C7→ C5→ C4?	5	18	1.4		0.06		**876**
Θ=0.8									
C	**-**	C1→C2?	0	10		4.0		0.40	**460**
C	**E**	C1→C2 →C5→C4?	18	16		5.6		0.16	-296

3.3. Literature Review

The present work belongs to two streams of literature: one on full truckload transportation and the other on dynamic vehicle routing problems (DVRPs). Table 3.3 summarizes the relevant studies. Since the problem falls under the category of full truckload transportation, we first position our work in that literature and highlight its novelty. Then, we briefly review the related DVRPs to adopt a solution concept to handle the proposed problem.

Our proposed model is a comprehensive Dynamic Pickup Delivery Full Truckload (DPDFL) problem in which several operational factors in the truckload industry are taken into account. To highlight the novelty of this study, we carefully point out the limitations of relevant truckload studies summarized in Table 3.3. Although all of these works addressed truckload problems, the key factor that remarkably influences our choice of modeling approach is tour capability (i.e., continuous truckload routes). This becomes less important when the average time of serving a load is very long (between two to four days) which is the case for large trucking companies working in nationwide or international markets. Powel and colleagues investigated this type of problem, which is simplified to different versions of assignment problems.

Powell (1987) extended his previous work (Powell, 1986) on the full truckload transportation problem by presenting the network flow problem. Similar to the former study, each node represents a region at a particular time. Following the same approach, Powell et al. (1988) proposed a model called LOADMAP which combines the real-time load assignment with sophisticated future forecasts to maximize the truckload profit and service level. In another work, Powell (1996) proposed a stochastic dynamic load assignment problem formulation. He showed that when some stochastic information about future demand is available, the proposed model outperforms the deterministic one. Powell et al. (2000) took a comprehensive simulation-based approach for tackling dynamic load assignment problems. The approach was to design an offline algorithm for the static version and put it into practice for a dynamic problem when demands were gradually realized as the time elapses. The result suggested that the greedy approach can be superior in the long run compared to the optimal myopic solution with the uncertainty in demands and travel times. Gregory and Powell (2002) modeled a truckload problem using a stochastic dynamic resource allocation approach. They used adaptive dynamic programming with a non-linear approximate function to solve the problem. The result showed that the algorithm based on the proposed approximation produced a near-optimum solution to deterministic problems. Finally, the work by Simao et al. (2009) motivated by Schneider National Inc. (the largest truckload motor carrier in the US) is the largest scale problem in the literature with over 6000 drivers. Their approximate dynamic programming model handled a great level of detail.

The other stream of relevant full truckload research focuses on smaller trucking companies that view tour capability as essential. These studies used mixed integer programming to formulate the problem and a rolling horizon approach for implementation (e.g., Yang et al, 1998; Yang et al, 2004; Gronalt et al., 2003; Tjokroamidjojo et al., 2006). The defined problem was the same in the studies by Yang et al. (1998) and Yang et al. (2004). The objective was to minimize the total cost. They used their models to develop tours with the capability of diverting trucks based on the arrival of new information into the system. As defined by Regan et al. (1995), diversion is a model capability that can divert a vehicle moving empty toward a pickup point to take another request. However, it is not allowed to divert loaded-moving vehicles while updating the decision. Ichoua et al. (2006) estimated that diversion in dynamic vehicle routing problems improved the system performance by up to 4.3% despite its operational difficulty. However, dwelling cost, which is one of the important components of cost structure, was not part of their model. Another limitation of that work is that trucks moved continuously between different cities, which means that a truck may never return to its home base.

Table 3.3. Summarizing the most related studies to the current study

Author(s) and publication year	Problem Type	Tour Capability	Demand Prob. Information	Decision Interval	Objective Function	Modeling Approach	Diversion Capability	Job Rejection
Powell (1987)	Truckload	X	√	Discrete	Max. Profit	Stochastic Formulation	X	√
Powell et al. (1988)	Truckload	X	X	Discrete	Max. Profit	Stochastic Formulation	X	√
Powell (1996)	Truckload	X	√	Discrete	Max. Profit	Stochastic Formulation	X	√
Yang et al. (1998)	Truckload	√	X	Continuous	Min. Cost	MIP[1]	√	√
Powell et al. (2000)	Truckload	X	X	Discrete	Min. Cost	MIP	X	√
Godfery and Powell (2002)	Truckload	X	X	Discrete	Max. Profit	Adoptive DP[2]	X	√
Gronalt et al. (2003)	Truckload	√	X	Discrete	Min. Cost	MIP	X	X
Thomas and White (2004)	Single VRP	√	X	Discrete	Min. Cost	Markov decision process	X	√
Bent and Van Hentenryck (2004)	DVRPTW[3]	√	√	Continuous	Min. Cost	Sampling Method	X	√
Yang et al. (2004)	Truckload	√	√	Continuous	Min. Cost	MIP	√	√
Hvattum et al. (2006)	DVRP	√	√	Discrete	Min. Cost	Sampling Method	X	√
Ichoua et al. (2006)	DVRPTW	√	√	Continuous	Min. Cost	Sampling Method	√	√
Tjokroamidjojo et al. (2006)	Truckload	√	X	Discrete	Min. Cost	Integer Programming	X	X
Thomas (2007)	Single DVRP	√	√	Discrete	Min. Cost	Markov decision processes	X	√
Simao et al. (2009)	Truckload	X	X	Discrete	Max. Profit	Approximate DP	X	√
Schilde et al. (2011)	Dial-a-ride problem	√	√	Discrete	Min. Total Tardiness	Sampling Method	X	X
Schmid (2012)	Ambulance relocation	X	X	Continuous	Min. Response time	Approximate DP	X	X
The present study	Truckload	√	√	Discrete	Profit Max.	MIP & Sampling Method	√	√

[1] Mixed Integer Programming
[2] Dynamic Programming
[3] Dynamic Vehicle Routing Problems with Time Windows

Unlike the previous studies, the work of Gronalt et al. (2003) addressed tour length to force trucks to return home after a predefined interval. The approach was based on generating tours with a very restrictive assumption that there is no limit on the number of available trucks. Their model did not capture the cost of delay and dwelling in designing tours. The proposed policy was also very restrictive in the sense that no loads could be rejected and no trucks could be diverted.

Tjokroamidjojo et al. (2006) addressed a full truckload pickup and delivery problem in which empty movements, dwelling time, and subcontracting costs were taken into account. They also investigated how much a trucking company can reduce cost by obtaining additional information further in advance. However, their proposed mathematical model was subject to some limitations. For example, there was no option to divert a truck when new information entered the system. Moreover, similar to Yang et al. (1998) and Yang et al (2004), there was no home base for trucks. Addressing the limitation of related studies, the contributions of this chapter are threefold:

- Proposing a new two-index mixed integer programming algorithm which is more efficient compared to three-index formulations in literature (e.g., Keskinocak and Tayur,1998; Tjokroamidjojo et al., 2006). The efficiency of the MIP model is improved by incorporating preprocessing functions, which uses characteristics of the problem.
- Designing a simple and intuitive policy that can help carriers to improve their razor-thin profit provided by the transportation network characteristics.
- The sampling method concept is adopted from the literature of DVRPs to improve our proposed Pure-Θ Policy. This Pure-Θ Policy and our proposed enhanced version of it are examined against two other dispatching methods.

The last contribution of this work requires us to review the second stream of related literature. As seen from Table 3.3, modeling the problem with dynamic programming and applying Markov decision processes are used for single DVRPs because they suffer from the curse of dimensionality (Thomas and White, 2004; Thomas, 2007). Approximate dynamic programming is an effective method to overcome the curse of dimensionality in dynamic programming (Powell et al., 2007). Despite successful implementation of approximate dynamic programming by Schmid (2012), we are not aware of any implementation of the method for a problem with tour making capabilities. The majority of research studies use sampling approach (multiple scenario generation) to solve DVRPs. These studies will be discussed in section 7 before applying the solution procedure.

3.4. Problem Definition

As mentioned earlier, the problem under study is called dynamic pickup and delivery with full truckload (DPDFL) consisting of a fixed fleet of trucks in the transportation network. The customers' demands (loads) are known gradually as time elapses. We retain the literature's standard assumption that each trip is executed without a break. Loads and trucks have their own attributes. The truck attributes are home domicile, hours away from home, the maximum allowed hours away from home, determined by a carrier or federal department of transportation (for drivers), and the current location. The load attributes are the earliest and latest pickup time, the maximum permissible delay time, the pickup location and the delivery location. Taking all the attributes of loads and trucks into account, the optimal DPDFL solution specifies the carrier's profit maximizing decisions concerning (i) whether to accept or reject new load(s), and (ii) the sequence of accepted loads that each truck will serve. The major assumptions are as follows:

- The shipment cost is a linear function of travel time which itself is a linear function of distance. Similar to what is common in the literature (e.g., Powell et al., 1988; Powell, 1996), the gained revenue is proportional to the trip length, i.e., the distance/time between pickup and delivery points.

- The length of each tour (i.e., tour time span) has to be less than the maximum hours that a driver can be away from home.
- Each truck can handle one load at a time (i.e., full truckload transportation).
- Given long haul transportation, loading and unloading times are a negligible part of the total time to serve a load and can therefore be ignored.
- There is a hard time-window to serve a load. Thus, a load will be rejected if it cannot be served within the predefined time window.
- The depot is the home domicile of drivers. A truck is returned to the depot if it is not scheduled to serve any loads at that decision epoch. This is a common practice if the dispatcher has access to advance load information (e.g., knowing that there is no request arriving for the rest of the day). The logic is simple because the average repositioning is typically shorter from the depot (if it is located at the center) and dwelling cost is negligible at the driver's home domicile.

3.4.1 Common Mathematical Models

There are two common ways to formulate a DPDFL problem. The first one uses an extended version of the assignment problem (e.g., assignment with timing constraints) to exploit the problem's characteristics. This is the most common approach in the literature (see Yang et al. 1998; Powell et al., 2000; Yang et al., 2004; Tjokroamidjojo et al., 2006). In the second one, the problem can be formulated as a variant of capacitated arc routing problems (CARP) in which each directed arc represents one load with a designated origin and destination. A recent work by Liu et al. (2010a, b) proposed an integer-programming model to formulate CARP for truckload industries and a quality lower bound. They also developed a heuristic method based on graph theory to solve the proposed model since the exact method is incapable of handling large problem instances. However, they captured neither time windows nor the fleet size of the transportation network for fulfilling demands.

Comparing the different approaches in the literature, the former is shown to be more promising to use because the dimensionality of the model grows quickly in the latter case. Among the related studies, Tjokroamidjojo et al. (2006) used an effective approach to handle DPDFL. The utilized approach consists of two parts, a preprocessing part for time-based restrictions and an assignment problem afterwards. Since time-window restrictions are explicitly handled outside the mathematical model, the approach performs well by reducing the number of constraints and decision variables. Although our approach is similar to Tjokroamidjojo et al. (2006), we must handle some of the time-based constraints inside the MIP because most of the load and truck attributes are determined after solving the model. After developing the model, with the aid of a simple example, we point out the issue of handing all time-based constraints outside the mathematical model.

3.4.2 The Model Inputs

To formulate the proposed model, notations, parameters, and decision variables are presented below.

- **Notation**

  I: set of all available trucks, indexed by *i, u*

  J: set of loads, indexed by *j*, *k, r*

  h: home domicile of trucks (i.e., the depot).

- **Parameters**

  a_j: departure location of load *j*

  b_j: destination location of load *j*

  α_j: the earliest pick-up time of load *j*

  U_j: maximum permissible delay for serving load *j*

$D(.,.)$: travel time between any two points in the service area. Traveling time between two locations can be described as function of distance.

N: maximum hours that a driver can be away from home

n_i: maximum hours left for truck *i* to be away from its home at the decision epoch

e: the revenue earned per hour while moving loads

c: the traveling cost (empty or loaded) per hour of driving

w: the penalty cost per hour for a truck being idle at any load location (dwelling cost).

l: the penalty cost per hour for late pickup

τ: time at the decision epoch

H: a very large positive number.

- **Decision Variables**

Y_{ik}^{0}: if truck *i* serves load *k* directly at the first stop, 1 otherwise 0.

Y_{ik}^{1}**:** if load *k* is served by truck *i* through its own depot at the first stop, 1 otherwise 0.

X_{jk}^{0}**:** if load *k* is served immediately after load *j,* 1 otherwise 0.

X_{jk}^{1}**:** if load *k* is served through the depot after load *j,* 1 otherwise 0.

Z_{ik}: if load *k* is served by truck *i* after another load, 1 otherwise 0.

O_k: arrival time at the pickup location of load *k*

N_{ik}: the remaining allowable time for truck *i* when arrives at the pickup location of load *k.*

The real-time location of each truck is important at each decision epoch because of the problem's dynamic nature. If the current location of truck *i* is denoted with η_i, $D(\eta_i, q)$ shows the traveling time from the current location of truck *i* to the location *q*. If the truck cannot get to the origin location of load *j* at time α_j, it can still pickup that load only if its maximum permissible delay (U_j) is not violated. However, late pickup is penalized by \$*l*/hr. Thus, it is important to note that, as it implies, lateness is computed with regards to the load's earliest availability. This is a common practice for serving more time-sensitive loads. For example, Logikor Company (a Canadian low asset based third-party logistics provider, http://www.logikor.com) uses a similar approach for delivery of commodities to manufacturing plants operating based on a Just-in-Time Philosophy).

Dwell time is the waiting time experienced by a driver/truck if the truck must wait at the pickup location (i.e., it reaches the pickup location of load j earlier than α_j). Although we consider the same dwell cost for all clients' locations in this study, the model is flexible enough to address varying dwelling costs across client locations. Still, our study does reflect that dwelling costs at the truck/driver domicile is significantly smaller than at client locations. This is because there is no extra facility usage cost for, say, a driver to dwell at his/her home or at accommodations provided by the carrier. This creates an opportunity for trucking companies in dispatching decisions since they can check the feasibility and economical impacts of trucks spending idle time at the depot rather than waiting at the clients' location. For example, if a specific load will be available the next two days and a truck is close to the depot, the truck may be sent first to the depot and then scheduled for dispatching at an appropriate time (i.e., serving the load through the depot). Even though there could be some economical reasons to schedule a load through the depot, we are required to take an important fact into account; that drivers should be returned to their home domicile (the depot) at some point. That is why decision variables have been introduced to consider the option of serving a load through the depot.

We also consider a single cost parameter for traveling empty or loaded. This is due to the number of load-independent factors that are present regardless of travelling empty or full. There are certain costs that a carrier still incurs that are not overly influenced by the amount of freight being transported. For example, factors such as driver wages, equipment depreciation, administration, compliance and insurance, act as a fixed cost that must be incurred. These costs typically compromise 70% of the total cost of driving a truck, while the remaining 30% is typically fuel related. Fuel is a unique cost of its own due to the fact that the associated costs do not vary dramatically whether driving empty or full, specifically, an empty truck requires at least three-fourths of the fuel of a fully loaded truck. This is due to cargo-independent factors such as aerodynamic drag, engine losses, and the mass of the empty truck itself (Transport Canada, 2005; American Transportation Research Institute, 2014). Interestingly, tire wear is another cost factor that actually costs more when a truck is driven empty (Trucking Information, 2015; American Trucking Associations, 2011). It is evident that the above-mentioned factors play a significant role when determining the cost of moving empty or loaded and therefore the operating costs (either empty/loaded) remain fairly stable over the course of movement (Sheffi, 2012).

Since the model is flexible enough to allow reassignment and re-sequencing of loads and diversion of empty moving trucks, the decision made at the previous decision epoch can be modified at the current decision epoch for all the loads which have not received service yet. To acknowledge this assumption, we first define TST(i) as the status of truck i at the decision epoch τ. TST(i) can take three values; 1, -1, 0 meaning truck i is moving loaded, empty (either moving or idle at any location other than the depot), or sitting idle at its own depot, respectively. If truck i is serving load j at the decision epoch τ, it will be available at the later time, $\tau + D(\eta_i, \mathrm{b}_j)$ at the destination location of load j (i.e., the diversion is not allowed if a truck is moving loaded). If a truck is idle or empty, $\mathrm{TST}(i) \leq 0$, then truck i is available for scheduling at time τ at its current location. There is also a need to keep track of load status which is denoted with LST(j). There are four possible load statuses. If load j is being served at the decision epoch, LST(j) is equal to 2. The loads which are accepted, but not yet serviced, (i.e., LST(j) =1) enter the model for possible reassigning and/or re-sequencing. In order to distinguish new loads (i.e., the loads for which acceptance is not finalized yet) from the current ones (i.e., the loads being served, LST(*j*) =2, or waiting to be served, LST(*j*) =1), their statuses will be LST(*j*) =3. Finally, the loads which have already been rejected (i.e., LST(j) =0) never enter the model. We also define ST(i,*j*) as a binary parameter to address the status of a truck and load together. If truck i is serving load *j* at the decision time, then ST(i,*j*) takes 1, otherwise 0. Another important time-dependent attribute is the maximum number of hours left for the drivers to return home. Two situations can be considered for them: sitting idle at their home domicile (i.e., $n_i = N$) or on duty away from their home ($n_i < N$). It will be explained shortly how these features are incorporated in the proposed model. Since there is no type of uncertainty considered in traveling time, it is enough to pick up loads on time to guarantee their on-time delivery.

3.4.3 Preprocessing Stage

As mentioned in section 3.4.2, we tackle the static version of problem in two stages beginning with the preprocessing stage. This stage consists of two phases. In the first phase, we show how to compute the necessary pieces of information. Then, in the second phase, it is explained how the generated information is used to solve the proposed mathematical model.

3.4.3.1 Preprocessing Stage: Phase I

At each decision epoch, trucks and loads have different attributes. Based on their current statuses, the dwelling and lateness duration can be computed. The lateness can occur in two situations: i) truck *i* serves load *j* as the first load; ii) a truck serves load *k* after load *j*.

$\mathrm{DL0}(i,j)$: the lateness duration at the load pickup location *j* if truck *i* serves load *j* first. If truck i serves load *j* directly without visiting the depot, $\mathrm{DL0}(i,j)$ is modified as $\mathrm{DL0_{WD}}(i,j)$. For $\mathrm{TST}(i) < 1$, $\mathrm{DL0_{WD}}(i,j)$=$max\left(0, \mathrm{D}(\eta_i, a_j) + \tau - \alpha_j\right)$. If the truck is moving loaded, $\mathrm{TST}(i) = 1$, toward the destination of a load (e.g., load k), $\mathrm{DL0_{WD}}(i,j) = max\big(0, \tau + \mathrm{D}(\eta_i, b_k) + \mathrm{D}(b_k, a_j) -$

$\alpha_j)$. If truck i serves load j through the depot, $\mathrm{DL0}(i,j)$ is modified as $\mathrm{DL0}_{TD}(i,j)$. For $\mathrm{TST}(i) < 1$, $\mathrm{DL0}_{TD}(i,j) = max(0, \tau + D(\eta_i, h) + D(h, a_j) - \alpha_j)$ and for a loaded truck (e.g., while serving load k) will be $\mathrm{DL0}_{TD}(i,j) = max(0, \tau + D(\eta_i, b_k) + D(b_k, h) + D(h, a_j) - \alpha_j)$.

$\mathrm{DL1}(j,k)$: the *minimum* lateness at the load pickup location *k* if the same truck serves load *k* immediately (or through its depot) after load *j*. Load *k* will experience some lateness if there is not enough time to reach the pickup location of load *k* immediately after serving load *j*. It is denoted with $\mathrm{DL1}_{WD}(j,k) = max\ \left(0, \left(\alpha_j + D(a_j, b_j) + D(b_j, a_k)\right) - \alpha_k\right)$. However, the minimum lateness of load *k* if it is served after load *j* through the depot will be $\mathrm{DL1}_{TD}(j,k) = max\left(0, \left(\alpha_j + D(a_j, b_j) + D(b_j, h) + D(h, a_k)\right) - \alpha_k\right)$.

Similar to what is explained for calculating lateness time, truck dwelling might occur in the following cases: i) truck *i* serves load *j* as the first load directly (i.e., without visiting the depot), ii) a truck serves load *k* after load *j* directly.

$\mathrm{Dw1}(i,j)$: the dwell time at the load pickup location *j* if truck *i* serves load *j* first given it was heading from the previous load location directly (in the current decision epoch). As we defined the dwell time, this happens if the truck arrives earlier at the load pick up location. For empty trucks, $\mathrm{TST}(i) = -1$, $\mathrm{Dw1}(i,j) = max\ \left(0, \alpha_j - (\tau + D(\eta_i, a_j))\right)$. If truck *i* is moving loaded toward destination *k* at the decision epoch, similar reasoning leads to dwell time being $\mathrm{Dw1}(i,j) = max\ \left(0, \alpha_j - \left(\tau + D(\eta_i, b_k) + D(b_k, a_j)\right)\right)$ in which $\tau + D(\eta_i, b_k)$ is when truck *i* is available after completing the service of load *k*.

$\mathrm{Dw2}(j,k)$: the *minimum* dwell time of a truck at the pickup location of load *k* if it comes directly after serving load *j*, $\mathrm{Dw2}(j,k) = max\ \left(0, \alpha_k - \left(\alpha_j + U_j + D(a_j, b_j) + D(b_j, a_k)\right)\right)$.

3.4.3.2 Preprocessing Stage: Phase II

In this phase, the following three tasks are performed. First, updating all dynamic attributes of trucks (e.g., hours away from home and current truck location) and loads (e.g., a load is waiting to be served or being served). Second, identifying infeasible truck-load and load-load combinations; and finally identifying feasible combinations that cannot be part of the optimal solution. Since the first part is straight forward, only the last two functions of the preprocessing stage are discussed here.

Given the current status of the trucks, we determine whether a particular truck is eligible for serving a certain load. This must be done for all available truck-load combinations. It is trivial that certain truck-load combinations are not feasible if the truck cannot be available at the pickup location of the load without violating the maximum delay. Thus, the following modifications are applied to the decision variables: $Y_{ij}^0 = 0$ if $\mathrm{DL0}_{\mathrm{WD}}(i,j) > U_{\mathrm{j}}$ and $Y_{ij}^1 = 0$ if $\mathrm{DL0}_{\mathrm{TD}}(i,j) > U_{\mathrm{j}}$.

Similar to what is done for truck-load combinations, we examine the feasibility of serving load *k* immediately (via depot of truck *i*) after load *j*. Here, the best case scenario for load-load combinations is determined. The best possible case is the time that load *j* is served on time so that no delay is carried toward serving load *k*. It is evident that load *k* cannot be served directly (or via the depot) after load *j* when there is not enough time for the truck to be at the load *k* pick-up location without violating its time window. Thus, the following adjustments are done because if a load combination is not feasible in the best-case scenario, it cannot be feasible at all (i.e., if $\mathrm{DL1}_{WD}(j,k) > U_k$ then $X_{jk}^0 = 0$ and if $\mathrm{DL1}_{\mathrm{T}D}(j,k) > U_k$ then $X_{jk}^1 = 0$). On the other hand, if the minimum lateness is smaller than or equal to the maximum allowable delay of U_{k}, the combination is not conclusively infeasible. This is extremely important because the decision at this stage is made based on the minimum lateness but not the actual lateness. Therefore, considering different possible assignment decisions, some load combinations with $\mathrm{DL1}_{WD}(j,k) \leq U_k$ or $\mathrm{DL1}_{\mathrm{TD}}(j,k) \leq U_k$ might not be feasible after solving the problem. This exactly explains why we need to have time components in the mathematical model.

We can also identify the truck-load and load-load combinations that could not be part of the optimal solution. Before identifying non-optimal truck-load and load-load assignments, it is shown that if a truck visits the depot before serving a load, the dwelling cost should be zero. This is a trivial property since for every dispatching decision from a depot with dwelling time and cost greater than zero, there exists an alternative decision with a larger profit with dwelling cost equal to zero. This can be attributed to the negligible dwelling cost assumptions at the depot. In simple words, a truck is never dispatched from the depot in such a way that it has to wait at the pick-up location of a load.

To exclude some of the load-truck combinations, it is sufficient to show that they cannot be part of the optimal solution. At optimality, truck *i* does not serve load *j* at the first stop directly (i.e., $Y_{ij}^0 = 0$) if conditions (3.1) and (3.2) hold. These conditions simply check if the saving in omitting corresponding dwelling cost (by visiting the depot) outweighs the extra travelling costs. Satisfying these conditions means that Y_{ij}^0 never shows at optimality because not only is it less cost efficient than Y_{ij}^1 but it also uses the available hours that a driver can be away from the home domicile.

$$w \times \mathrm{Dw1}(i,j) \geq \mathrm{c} \times \left(\mathrm{D}(\eta_\mathrm{i},\mathrm{h}) + \mathrm{D}(\mathrm{h},\mathrm{a_j}) - \mathrm{D}(\eta_\mathrm{i},\mathrm{a_j})\right),\quad \text{TST}(i)\text{= -1, and LST}(j)\text{=1,3} \qquad (3.1)$$

$$w \times \mathrm{Dw1}(i,j) \geq \mathrm{c} \times \left(\mathrm{D}(\mathrm{b_k},\mathrm{h}) + \mathrm{D}(\mathrm{h},\mathrm{a_j}) - \mathrm{D}(\mathrm{b_k},\mathrm{a_j})\right),\quad \text{TST}(i)\text{=1, LST}(k)\text{=2, and LST}(j)\text{=1,3} \qquad (3.2)$$

A similar reasoning is used to exclude some of the load-load combinations. Serving load *k* directly after load *j* is not part of optimal solution (i.e., $X_{jk}^0 = 0$) if condition (3.3) is satisfied.

$$w \times \mathrm{Dw2}(j,k) \geq \mathrm{c} \times \left(\mathrm{D}(\mathrm{b_j},\mathrm{h}) + \mathrm{D}(\mathrm{h},\mathrm{a_k}) - \mathrm{D}(\mathrm{b_j},\mathrm{a_k})\right),\quad \text{LST}(j)\text{=1,3 and LST}(k)\text{=1,3} \qquad (3.3)$$

3.4.4 Mathematical Model

Having defined all parameters and dynamic aspects of the model in the preprocessing stage, it is time to formulate the conceptual model. Before formulating the proposed model, it is important to check which loads enter the model and their notations. As defined earlier, J represents the set of all loads entered in the model. However, we are required to differentiate them in order to have a neat mathematical formulation. To do so, the set of new loads are denoted with $\bar{J}$ (LST(j) =3), the set of accepted loads waiting for service with $\bar{\bar{J}}$ (LST(j) =1), and the set of loads being served at the decision epoch with $\hat{J}$ (LST(j) =2). Thus, $J = \bar{J} \cup \bar{\bar{J}} \cup \hat{J}$.

Having the parameters and decision variable defined, the model will be formulated as follows. To have a better understanding of the model, we break it down into smaller components and explain them one by one. The objective function to be maximized is the profit which includes the revenue and the relevant costs. In the following model, there are some non-linear terms in the objective function and constraints that can be easily reformulated into linear terms.

- **Revenue;** the revenue depends on trip-length of the accepted loads:

$$e \sum_{i \epsilon I} \sum_{k \epsilon \bar{J} \cup \bar{\bar{J}}} D(\mathrm{a_k}, \mathrm{b_k}) \left(Y_{ik}^0 + Y_{ik}^1 + Z_{ik} \right) \tag{3.4}$$

- **Cost of moving loaded trucks**;

$$c \sum_{i \epsilon I} \sum_{k \epsilon \bar{J} \cup \bar{\bar{J}}} D(\mathrm{a_k}, \mathrm{b_k}) \left(Y_{ik}^0 + Y_{ik}^1 + Z_{ik} \right) \tag{3.5}$$

- **Cost of moving empty trucks;** empty traveling cost can be a result of moving trucks from the delivery location of one load to the pickup location of the next load:

$$c \sum_{k \epsilon \bar{J} \cup \bar{\bar{J}}} \left[\sum_{j \epsilon \bar{J} \cup \bar{\bar{J}}} D\left(\mathrm{b_j}, \mathrm{a_k}\right) X_{jk}^0 + \sum_{j \epsilon \bar{J} \cup \bar{\bar{J}}} \left[D\left(\mathrm{b_j}, h\right) + D(h, \mathrm{a_k}) \right] X_{jk}^1 \right] \tag{3.6}$$

- The empty traveling cost occurs for repositioning empty, idle or loaded trucks to the pickup location of the first load in the sequence:

$$c \sum_{i\epsilon I, TST(i)<1} \sum_{k\epsilon \bar{J} \cup \bar{\bar{J}}} D(\eta_i, \mathrm{a_k})\, Y_{ik}^0 + c \sum_{i\epsilon I, TST(i)<1} \sum_{k\epsilon \bar{J} \cup \bar{\bar{J}}} [D(\eta_i, h) + D(h, \mathrm{a_k})]\, Y_{ik}^1 +$$

$$c \sum_{i\epsilon I} \sum_{j\in \hat{J}, ST(i,j)=1} \sum_{k\epsilon \bar{J} \cup \bar{\bar{J}}} D(\mathrm{b_j}, \mathrm{a_k})\, Y_{ik}^0 + c \sum_{i\epsilon I} \sum_{j\in \hat{J}, ST(i,j)=1} \sum_{k\epsilon \bar{J} \cup \bar{\bar{J}}} [D(\mathrm{b_j}, h) + D(h, \mathrm{a_k})]\, Y_{ik}^1 \qquad (3.7)$$

- The empty traveling also exists in either of following cases. First, the truck is going back to its depot after serving all its assigned loads (in term 3.8, based on definition of deadhead coefficient policy, travel cost from the destination of the last load in the sequence to the depot is weighted with Θ). Second, a moving truck (i.e., either empty or loaded) is not assigned to any load and so it is heading back to its depot (term 3.9). If the truck is moving a load (e.g., serving load *j*), it cannot be diverted based on the predefined assumption (similar to Regan et al., 1995). This means that it continues the movement of load *j* to its delivery location (b_j). Then, the empty traveling starts from that location (b_j) to the depot (h).

$$\Theta \times c \sum_{j\epsilon \bar{J} \cup \bar{\bar{J}}} D(\mathrm{b_j}, h) \left[\sum_{i\epsilon I} \left(Y_{ij}^0 + Y_{ij}^1 \right) + \sum_{r\epsilon \bar{J} \cup \bar{\bar{J}}} X_{rj}^0 + X_{rj}^1 - \sum_{k\epsilon \bar{J} \cup \bar{\bar{J}}} X_{jk}^0 + X_{jk}^1 \right] \qquad (3.8)$$

$$c \sum_{i\epsilon I, ST(i)=-1} D(\eta_i, h) \left[1 - \sum_{k\epsilon \bar{J} \cup \bar{\bar{J}}} (Y_{ik}^0 + Y_{ik}^1) \right]$$

$$+ c \sum_{i\epsilon I} \sum_{j\in \hat{J}, ST(i,j)=1} D(b_j, h) \left[1 - \sum_{k\epsilon \bar{J} \cup \bar{\bar{J}}} (Y_{ik}^0 + Y_{ik}^1) \right] \qquad (3.9)$$

- **Dwelling cost;** this is the cost of waiting at the load pickup location which can occur when the load is either at the beginning of the sequence or after another load directly.

$$w \sum_{k\epsilon \bar{J} \cup \bar{\bar{J}}} max(0,\ \alpha_k - O_k) \qquad (3.10)$$

- **Lateness cost;** late service occurs when the truck arrives to the load's pick-up location after its availability. Lateness cost is incurred in all the following situations. A truck (e.g., moving empty, loaded or idle) is scheduled to serve a load directly from its current location, through the truck depot or after another load:

$$l \sum_{k \epsilon \bar{J} \cup \bar{\bar{J}}} max(0,\ O_k - \alpha_k) \tag{3.11}$$

Having formulated the objective function, the constraints are introduced as follows. The first and second constraint sets (3.12 and 3.13) ensure that all previous accepted loads will be served but there is no guarantee to take all new loads.

$$\sum_{i \epsilon I} (Y_{ik}^0 + Y_{ik}^1 + Z_{ik}) = 1, \qquad k \epsilon \bar{\bar{J}} \tag{3.12}$$

$$\sum_{i \epsilon I} (Y_{ik}^0 + Y_{ik}^1 + Z_{ik}) \leq 1, \qquad k \epsilon \bar{J} \tag{3.13}$$

- A truck can serve at most one load at the beginning of a sequence.

$$\sum_{k \epsilon \bar{J} \cup \bar{\bar{J}}} (Y_{ik}^0 + Y_{ik}^1) \leq 1, \qquad i \epsilon I \tag{3.14}$$

- Each accepted load can have at most one successor.

$$\sum_{k \epsilon \bar{J} \cup \bar{\bar{J}}} X_{jk}^0 + X_{jk}^1 \leq 1, \qquad j \epsilon \bar{J} \cup \bar{\bar{J}} \tag{3.15}$$

- The next set of constraints (3.16) ensures that if load *k* is served after load *j*, load *j* is either scheduled to be the first load or placed after another load *r*.

$$\sum_{k \in \bar{J} \cup \bar{\bar{J}}} X_{jk}^0 + X_{jk}^1 - \left[\sum_{i \epsilon I} Y_{ij}^0 + Y_{ij}^1 + \sum_{r \in \bar{J} \cup \bar{\bar{J}}} X_{rj}^0 + X_{rj}^1 \right] \leq 0, \qquad j \in \bar{J} \cup \bar{\bar{J}} \tag{3.16}$$

- The constraints (3.17) ensure that load *k* can be scheduled after load *j* if they are visited by the same truck.

$$Z_{ij} + Y_{ij}^0 + Y_{ij}^1 + \sum_{u\epsilon I, u\neq i} \left(Z_{uk} + Y_{uk}^0 + Y_{uk}^1\right) \leq 2 - \left(X_{jk}^0 + X_{jk}^1\right), \quad i\epsilon I,\ j,k \in \bar{J} \cup \bar{\bar{J}} \tag{3.17}$$

- The constraints (3.18) guarantee that a load is not scheduled at the beginning of a sequence if it is served after another load.

$$\sum_{j\in\bar{J}\cup\bar{\bar{J}}} X_{jk}^0 + X_{jk}^1 = \sum_{i\epsilon I} Z_{ik}\,, \qquad k \in \bar{J} \cup \bar{\bar{J}} \tag{3.18}$$

- Constraints (3.19) ensure that N_{ik} can only take a positive value if truck *i* serves load *k*. Thus, if truck *i* serves load *k*, then $\sum_{u\epsilon I} N_{uk} = N_{ik}$ which is used in constraints (20-23).

$$N_{ik} \leq \left(Z_{ik} + Y_{ik}^0 + Y_{ik}^1\right)H, \qquad k \in \bar{J} \cup \bar{\bar{J}},\ i\epsilon I \tag{3.19}$$

- Constraints (3.20) impose an upper bound for N_{ik} (a driver's allowable time while visiting the first load of the sequence). In this constraint set, M_{ik}^0 and M_{ik}^1 represent the remaining allowable time for the driver of truck *i* when serving load *k* at the beginning of the sequence either directly or through the depot. These two parameters are obtained from the preprocessing stage for all truck-load combinations.

$$\sum_{i\epsilon I} N_{ik} \leq \sum_{i\epsilon I} \left(M_{ik}^0 Y_{ik}^0 + M_{ik}^1 Y_{ik}^1\right) + \left(\sum_{j\in\bar{J}\cup\bar{\bar{J}}} \left(X_{jk}^0 + X_{jk}^1\right)\right)H, \qquad k \in \bar{J} \cup \bar{\bar{J}}, \tag{3.20}$$

- Constraints (3.21) introduce an upper bound for a driver's allowable time when serving load *k* immediately after load *j*. Constraints (3.22) perform similarly when the driver returns to the depot in-between visits.

$$\sum_{i\epsilon I} N_{ik} \leq \left[\sum_{i\epsilon I} N_{ij} - \left(\alpha_j - min\left(O_j, \alpha_j\right)\right) - D\left(\mathrm{a_j}, \mathrm{b_j}\right) - D\left(\mathrm{b_j}, \mathrm{a_k}\right)\right] + \left(1 - X_{jk}^0\right)H, .j,k \in \bar{J} \cup \bar{\bar{J}} \tag{3.21}$$

$$\sum_{i\epsilon I} N_{ik} \leq \left[N - D(\mathrm{h}, \mathrm{a_k})\right] + \left(1 - \sum_{j\in\bar{J}\cup\bar{\bar{J}}} X_{jk}^1\right) H, \qquad k \in \bar{J} \cup \bar{\bar{J}}, \tag{3.22}$$

- Constraints (3.23) guarantee that trucks have enough time to return home when they are at pick-up location of loads.

$$\sum_{i\epsilon I} N_{ik} \geq \alpha_k - min(O_k, \alpha_k) + D(\mathrm{a_k}, \mathrm{b_k}) + D(\mathrm{b_k}, \mathrm{h}) - \left(1 - \sum_{i\epsilon I}\left[Z_{ik} + Y_{ik}^0 + Y_{ik}^1\right]\right)H$$

$$k \in \bar{J} \cup \bar{\bar{J}}, \quad i\epsilon I \tag{3.23}$$

- Altogether, constraints (3.24) through (3.27) ensure that O_k does not take on an unrealistically large or small value to prevent dwelling or lateness costs. Constraints (3.24) and (3.25) apply when a truck is serving one load after another load directly while constraints (3.26) and (3.27) are for the case of a truck serving a load at the beginning of a sequence. In constraints (3.26) and (3.27), Q_{ik}^0 is the earliest time that truck *i* can be available at the pickup location of load *k*, which is computed in the preprocessing stage. If truck *i* is moving loaded (e.g., serving load *j*) at the decision epoch, $Q_{ik}^0 = \tau + D(\eta_i, \mathrm{b_j}) + D(\mathrm{b_j}, \mathrm{a_k})$; otherwise $Q_{ik}^0 = \tau + D(\eta_i, \mathrm{a_k})$.

$$O_k \geq D(\mathrm{a_j}, \mathrm{b_j}) + D(\mathrm{b_j}, \mathrm{a_k}) + max(O_j, \alpha_j) + (X_{jk}^0 - 1)H, \qquad j, k \in \bar{J} \cup \bar{\bar{J}} \tag{3.24}$$

$$O_k \leq D(\mathrm{a_j}, \mathrm{b_j}) + D(\mathrm{b_j}, \mathrm{a_k}) + max(O_j, \alpha_\mathrm{j}) + (1 - X_{jk}^0)H, \qquad j, k \in \bar{J} \cup \bar{\bar{J}} \tag{3.25}$$

$$O_k \geq \sum_{i\epsilon I} Q_{ik}^0 Y_{ik}^0 + \left(\sum_{i\epsilon I} Y_{ik}^0 - 1\right)H, \qquad k \in \bar{J} \cup \bar{\bar{J}}, \tag{3.26}$$

$$O_k \leq \sum_{i\epsilon I} Q_{ik}^0 Y_{ik}^0 + \left(1 - \sum_{i\epsilon I} Y_{ik}^0\right)H, \qquad k \in \bar{J} \cup \bar{\bar{J}}, \tag{3.27}$$

- Constraints (3.28) and (3.29) ensure that a truck arrives at the pick-up location of load *k* no sooner than after serving load *j* and traveling to load *k* through the depot if such a schedule is implemented. In constraints (3.28), Q_{ik}^1 is the earliest time that truck *i* can be available at the pickup location of load *k* if it visits the depot first. If truck *i* is moving loaded (e.g., serving load *j*) at the decision epoch, $Q_{ik}^1 = \tau + D(\eta_i, \mathrm{b_j}) + D(\mathrm{b_j}, \mathrm{h}) + D(\mathrm{h}, \mathrm{a_k})$; otherwise $Q_{ik}^1 = \tau + D(\eta_i, \mathrm{h}) + D(\mathrm{h}, \mathrm{a_k})$.

$$O_k \geq \sum_{i\epsilon I} Q_{ik}^1 Y_{ik}^1 + \left(\sum_{i\epsilon I} Y_{ik}^1 - 1\right)H, \qquad k \in \bar{J} \cup \bar{\bar{J}} \tag{3.28}$$

$$O_k \geq D(\mathrm{a_j}, \mathrm{b_j}) + D(\mathrm{b_j}, \mathrm{h}) + D(\mathrm{h}, \mathrm{a_k}) + max(O_j, \mathrm{\alpha_j}) + (X_{jk}^{1} - 1)H, \qquad j, k \in \bar{J} \cup \bar{\bar{J}} \qquad (3.29)$$

- Constraints (3.30) guarantee that accepted loads are served without violating the maximum permissible delay.

$$O_k \leq \alpha_\mathrm{k} + U_\mathrm{k}, \qquad k \in \bar{J} \cup \bar{\bar{J}} \qquad (3.30)$$

The presented model is for a general case where lateness is allowed for serving loads. If no lateness is allowed, the model can be simplified by eliminating constraints (3.24) to (3.30) and modifying the model using the generated information in the preprocessing stage (Appendix A). This approach (i.e., handling time-based constraints outside the mathematical model) is problematic if lateness is allowed. The following example illustrates the potential issue.

We assume a small 10-city example within a circle-shape area with radial travel time of 12hrs (cities are randomly selected from a 20×20 grid). For ease of exposition, one truck (with no time restriction to go back to the depot) and five loads are taken into account. Figure 3.3 depicts the transportation network along with travel times and the loads' earliest and latest departure times. For instance, the travel time to serve load B is 5.4 hrs, and it can be picked up at city 4 between times 96 and 101 for delivery to city 7. The truck is at coordinate (9, 19) at the time of decision making (time 0). As mentioned earlier, three tasks are performed at the preprocessing stage. After updating the dynamic attributes of trucks and loads, it checks for feasibility of truck-load and load-load assignments. Finally, it identifies those solutions that cannot be part of the optimal solution. The last feature can be easily explained based on the first movement of the truck (which is located in city 5 at time 0). If the first load is accepted, the truck has two options: 1) it can go directly to the pickup location of load A (travel time 8.7hrs) and wait for its availability (60-8.7 dwell time); 2) it can go to the depot (travel time 10.9hrs) and dwell there before heading to city 1 (travel time 10.3hrs) for serving the first load. The choice of the second alternative depends on the tradeoff between extra travelling cost (10.3+10.9-8.7)×60 and the saving in dwell cost (60-8.7)×25. Since the dwell cost saving outweighs the traveling cost increments, the first option is eliminated in the preprocessing stage (i.e., condition (3) is satisfied).

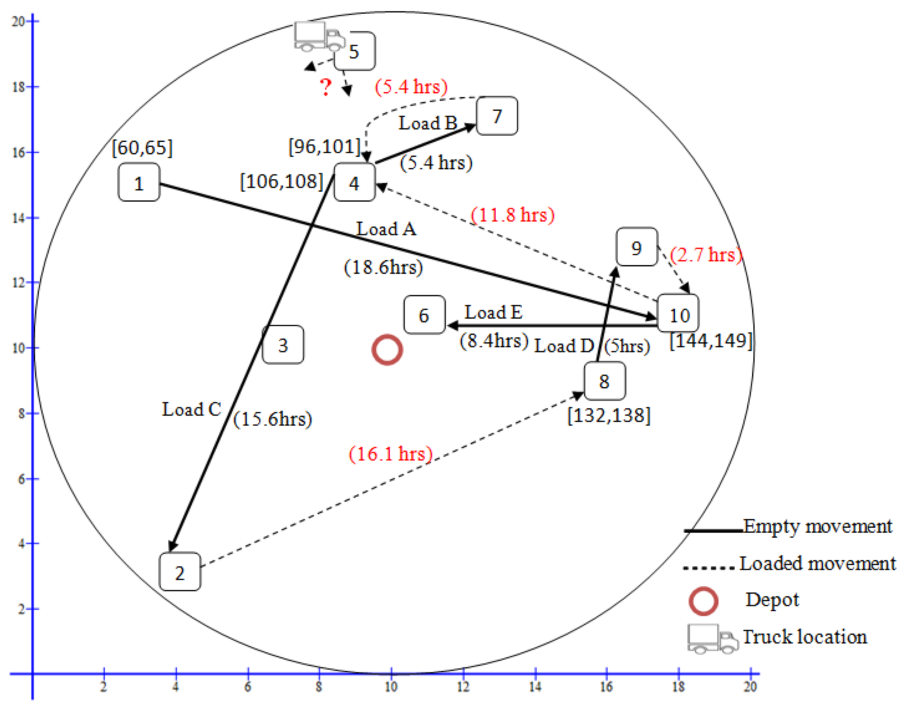


Figure 3.3. An infeasible solution when time-based constraints handled in the preprocessing stage

The problem arises if all time-based constraints are handled in the preprocessing stage (using the revised mathematical model in Appendix A). Based on the preprocessing outcome, it is feasible to serve load *k* directly after load *j* if *k*>*j*. The optimal solution of the integer programming model sends the truck to serve load A (through the depot) and the rest of the loads directly (without visiting the depot) one after another. In Figure 3.3, the broken lines show the empty movement of trucks and the solid ones represent the loaded movements. After handling load A through the depot at time 60, the truck moves toward the pickup location of load B and reaches the location too early, so it must wait until time 96. The truck cannot be at the pickup location of load C before time 106.8, but serving that load is still acceptable. This delay directly impacts the time that the truck reaches the origin of load D (time 138.5) when load D is no longer available. In this example, the preprocessing stage fails because it only compares whether two loads can be served immediately one after the other without considering the previous load(s) of that sequence. Based on the preprocessing result, serving load D after load C would be feasible if the pickup time of load 3 is 106. This example illustrates the need for explicitly including time-based constraints into the mathematical model. Since the actual dwell and lateness at each pickup location cannot be computed before solving the model, those terms should be replaced by correct terms representing dwell and lateness costs.

3.5. Experimental Design

In this section, we explain how the model's parameters are generated for use in the simulation study to provide useful insights. Our investigation of the academic literature and empirical reports suggested the potential influence of the following factors on a carrier's profitability: radius of service, trip length, load density, advance load information, and time windows. Having a quick look at the recent statistics of ATA (2013), we observe that the truckload market is highly fragmented where almost 90% of the carriers are small with six or fewer trucks. The Canadian statistics are very similar to the American ones. Moreover, as mentioned earlier, the smaller companies usually suffer more than big companies with a sophisticated network of lanes. Thus, we concentrate on small companies with six trucks.

Radius of service: defined as the furthest distance from the depot that a truckload carrier is willing to carry a load. Two levels are considered for the radius of service: a minimum of 18 hours (driving) and a maximum of 36 hours (driving).

Trip length: measured as travel time between a load's origin and destination. The test problems are generated in two categories called short and long trip-length groups. In the former, the majority of loads (80%) are shorter than the radius of service while in the latter the majority (80%) of loads are longer than the radius.

Load density: number of loads entering the system per truck per week. Load density is inversely related to the average length of loads (Powell, 1996), which usually ranges between 2 to 2.5 loads (per truck per week) for large companies with the average load length between two to four days. Since this study targets small carriers with shorter trip lengths, load density is studied at two levels, 2.5 (low load density) and 5 (high load density) loads per truck per week.

Advance load information (ALI): it is called knowledge window (KW) by Tjokroamidjojo et al. (2006) who define it as number of hours that loads' information is available in advance. Since the trucking industry is identified with excess capacity and a high level of competitiveness, last-minute call for transportation services is very common in the industry. It is also unusual for a shipper to book a load more than two or three days in advance (Frantzeskakis and Powell, 1990). Thus, acquiring load information very far in advance (e.g., a week or so) does not provide practical managerial insights. That is why we focus our attention on the three ALI/KW levels: 24, 48, and 72 hours.

Time Windows: Following Tjokroamidjojo et al. (2006), hourly dwelling and lateness costs are set to be $25 per hour. The maximum permissible delay for serving customers is examined at two levels: no lateness is allowed, and lateness is permissible in which the maximum lateness is drawn from a discrete uniform distribution with maximum of 5 hours.

The result of the abovementioned five factors at different levels becomes 48 combinations. We tested five replicates of each combination. Each replicate was a randomly generated instance of the experiments' stochastic conditions (e.g., earliest availability of loads and city locations). We use 240 test problems (5 replicates by the 48 combinations) in our simulation study. In all conducted numerical experiments, it was assumed that there are 50 cities within an area with a predefined radius of service. The number of cities in the transportation network is referred to as network size. These potential cities are representative of loads' origins and destinations. This approach is commonly used in the literature. For simulation studies, some authors consider that the shape of the service area is square (e.g., Yang et al. 2004; Özener and Ergun, 2008). To capture the radius of service as one of the transportation network settings, we assumed that the trucking company operates in a circle-shaped area. For each test problem, the parameters are generated as follows:

- To generate each load, an origin-destination pair is selected randomly from a 50-city network. The initial location of trucks is also determined by placing them randomly among the 50 cities.
- The earliest loads availability is generated from an exponential distribution in which the average inter-arrival time is determined based on load density.
- The average operating highway speed is used since the majority of cities are connected to each other via highways. The average operating speed is set to 55 mph, which is typical on US highways (refer to the recent report by the U.S. department of energy, 2011).
- In the trucking industry where drivers can easily be away from home between one to four weeks, most carriers try to have the drivers back home every fortnight (Powell, 1996). To be consistent with these statistics, this study sets the maximum number of hours that a driver can be away from home equal to 240 hours.
- Fuel cost and driver wages are the major portion of the operational cost. However, there are other miscellaneous cost components such as insurance premiums and maintenance. Given that we consider dwelling and lateness cost separately, it is fair to set the operational cost equal to $1.10 per mile (empty/loaded) and revenue to $2.25 per loaded mile. The earned revenue per mile also conforms to the TRANSCORE (provider of intelligent transportation systems) survey in 2011 from 600 small carriers. The 2:1 ratio of revenue to cost is also supported by the work of Gregory and Powell (2002).

- The overall length of the planning horizon highly depends on the average speed of the transportation mode. The slower mode of transportation usually requires a longer overall planning horizon. For example, Choong et al. (2002) considered a 15-day and 30-day planning horizon in empty container management in which a barge was one of the transportation modes. A shorter planning horizon (20-day) was considered in the truckload industry by Tjokroamidjojo et al. (2006). Similar to the latter study, we consider a three-week planning horizon.

3.6. Numerical Study

In this section, we first explain how each test problem is handled systematically in a dynamic context. Then, the simulation results are presented and analyzed. We used AIMMS modeling language and Gurobi 5.1 as a solver to run the test problems. The whole algorithm was programmed in MATLAB 2012b and run on a 2.8 GHz computer. As shown in Figure 3.4, the algorithm starts by setting the value of Θ for the deadhead coefficient policy. The clock is set equal to zero and the preprocessing engine is called every 12 hours (the time interval between two decision epochs) to update truck and load status and exclude infeasible assignments and non-optimal assignments. Then, loads with status 1, 2, and 3 are entered into the model. In other words, the loads that have been already delivered and the ones that are far in future (i.e., beyond the KW) do not enter into the model. The next step is to call the solver to solve the proposed MIP model to optimality. After the model is solved, the obtained schedule is implemented up to the next interval and checked for the termination condition (i.e., whether all the loads are considered during the overall planning horizon).

It is important to note that we need to record all the movement of trucks during the planning horizon since diversion of empty trucks is allowed. After the stopping criterion is satisfied, a simple algorithm tracks each truck's contribution to compute the system total profit for the middle two weeks of the study. The main reason for collecting data on just the middle two weeks is to control the anomalies of beginning and end-of horizon effects. For example, with KW=72hrs, the loads information of the second day must be known before the beginning of planning horizon. Thus, the problem is handled for the entire planning horizon while only the middle two weeks statistics (Day 4 to Day 17) are considered for further analysis in this chapter.

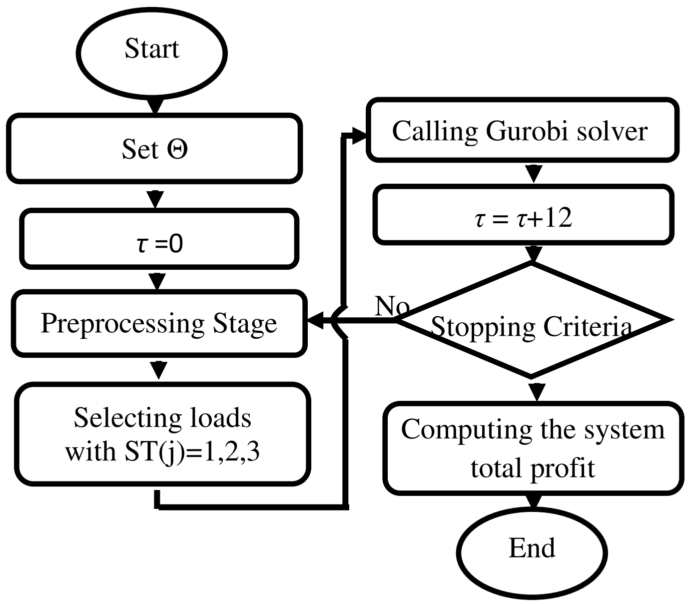


Figure 3.4. The detail of the dynamic implementation

As mentioned earlier, the Θ value can vary between 0 and 1. Six values are chosen for Θ: 0, 0.2, 0.4, 0.6, 0.8 and 1. Thus, all 240 generated test problems are solved six times, each time with one Θ value. In order to have a valid comparison, the obtained profit from the proposed policy is normalized by the optimal solution of the static version (when all loads information during the planning horizon is known in advance). Simply put, for each test problem, we divide the obtained profit of the Pure-Θ Policy by the static optimal solution. Because profit depends heavily on characteristics of the transportation network, normalization is essential in assuring fair comparison across different network settings.

Table 3.4 provides the averages of optimal profits of static versions along with CPU times (in seconds) and load rejection rates for low load density. Combinations are coded with three letters. The combination code represents radius of service (High/Low), trip length (High/Low), and load density (High/Low), respectively. Some intuitive results can be observed from Table 3.4. For example, having more high-revenue loads (the majority of loads are long) reduces the rejection rate while increasing the CPU time. Moreover, if lateness is allowed, the profit improvement is higher when trucks operate in a smaller service area (i.e., radius is low). Table 3.5 is similar to Table 3.4 but for the high load density combinations. Compared to Table 3.4, the higher load density does not necessarily increase the rejection rate because higher load density improves economy of scope (defined earlier) by lowering empty repositioning, thereby making low revenue loads more profitable. That is why increasing load density lowers the rejection rate when most loads are short. Another interesting observation is that the CPU time depends on the network settings. For example, it takes slightly more than 10 minutes on average to solve LLH coding with no lateness while HHH coding with lateness takes more than 50 hours to be solved to optimality.

Table 3.4. The averages of static optimal solutions (low load density)

CODE	Averages with no lateness			Averages when lateness allowed			Improvement
	Optimum Profit	CPU time (Sec)	Rejection Rate	Optimum Profit	CPU time (Sec)	Rejection Rate	
LLL	4034.8	1.8	59%	4576.2	16.0	57%	13.42%
LHL	15113.3	32.2	26%	17414.5	86.2	22%	15.22%
HLL	10727.6	3.4	63%	11733.2	8.6	62%	9.37%
HHL	28703.5	1045.2	37%	29388.8	2472.8	36%	2.39%

Table 3.5. The averages of static optimal solutions (high load density)

CODE	Averages with no Time Window			Averages with Time Window			Improvement
	Optimum Profit	Solution time (Sec)	Rejection Rate	Optimum Profit	Solution time (Sec)	Rejection Rate	
LLH	14787.4	701.7	56%	16453.8	1221.1	48%	11.27%
LHH	34833.1	2679.3	41%	39104.3	9805.7	38%	12.26%
HLH	31341.0	25729.7	61%	33482.1	43541.0	56%	6.83%
HHH	48347.8	120960.9	60%	50199.2	181440.3	59%	3.83%

3.6.1 The Impact of the Deadhead Coefficient Policy

The simulation result for the impact of Θ on the normalized profit during the middle two weeks period is shown in Figure 3.5 for the case of low load density. Since the profit is normalized by the static optimal solution (which is usually an unrealistic benchmark), small ratios do not necessarily indicate a low performance of the policy (this will be discussed in the next section). Since the normalized profit is insensitive does not show a remarkably different behavior to the choice of Θ with-and-without lateness, we first present the obtained results for the test problems with lateness allowed and then point out the differences if delay is not permissible. As a general observation, the policy produces a lower normalized profit if lateness is not allowed. Figure 3.5 comprises 4 charts that can be interpreted as follows.

- **Fig. 3.5, Chart (a):** Low service radius-Low trip length-Low load density (LLL)

  Seen from chart (a), a properly tuned deadhead coefficient policy can obtain 80% of static optimal profit with 72 hrs KW. We also observe that the profit is sensitive around the best value of Θ but the sensitivity declines by extending the KW. For example, shifting from the best Θ of 0.8 to a value of 0.6 when lateness is allowed decreases the normalized profit by 9 percentage points (from 68% to 59%) for the smallest KW while it only drops by 8 and 5 percentage points for larger KWs.

- **Fig. 3.5, charts (b):** Low service radius-High trip length-Low load density (LHL)

  A properly tuned deadhead coefficient policy of Θ=0.6, yields the highest profit independent of the KW and the lateness option. The obtained normalized profit is the best across all combinations with low load density (almost 90% with KW=72hrs). The profit is less sensitive around the best value of Θ when the knowledge window is longer than 24hrs. Thus, under this setting, the choice of Θ becomes more important when KW is limited to one day. For the no lateness option, it is more crucial to obtain loads information beyond the next day (under Θ=0.6, the improvement is 16 percentage points, from 73% to 89%).

- **Fig. 3.5, charts (c):** High service radius-Low trip length-Low load density (HLL)

  Unlike the previous combinations, the best Θ value (resulting in the highest profit) depends on KW and is identical for the no lateness option. The best Θ values for 24, 48, and 72hrs advance load information are 0.4, 0.6, and 0.8, respectively. This intuition can be explained by looking at the problem from the dispatcher's point of view. Given that the carrier is operating in a large service area, the dispatching policy typically needs to be less conservative to become more profitable when the KW is shorter. The lower level of conservatism helps the carrier to improve load acceptance by taking more risk. Although a smaller Θ is recommended for a shorter KW, the profit will drop remarkably if the selected value is too small (e.g., Θ=0.2). The main reason is that most loads are short (low revenue) so taking more risk does not necessarily lead to a higher profit.

- **Fig. 3.5, charts (d):** High service radius-High trip length-Low load density (HHL)

  Similar to the HLL combination, the Θ value that produces the highest normalized profit depends on the KW. Using the same approach, the dispatching policy needs to be more optimistic in order to gain more profit when KW becomes smaller. However, unlike the HLL combination, a low value of Θ (e.g., 0.2 or 0) still produces a high profit for short KW (e.g., 24hrs) because the dispatching policy tends to be less conservative (smaller Θ) compared to the HLL setting in which a majority of loads are low revenue.

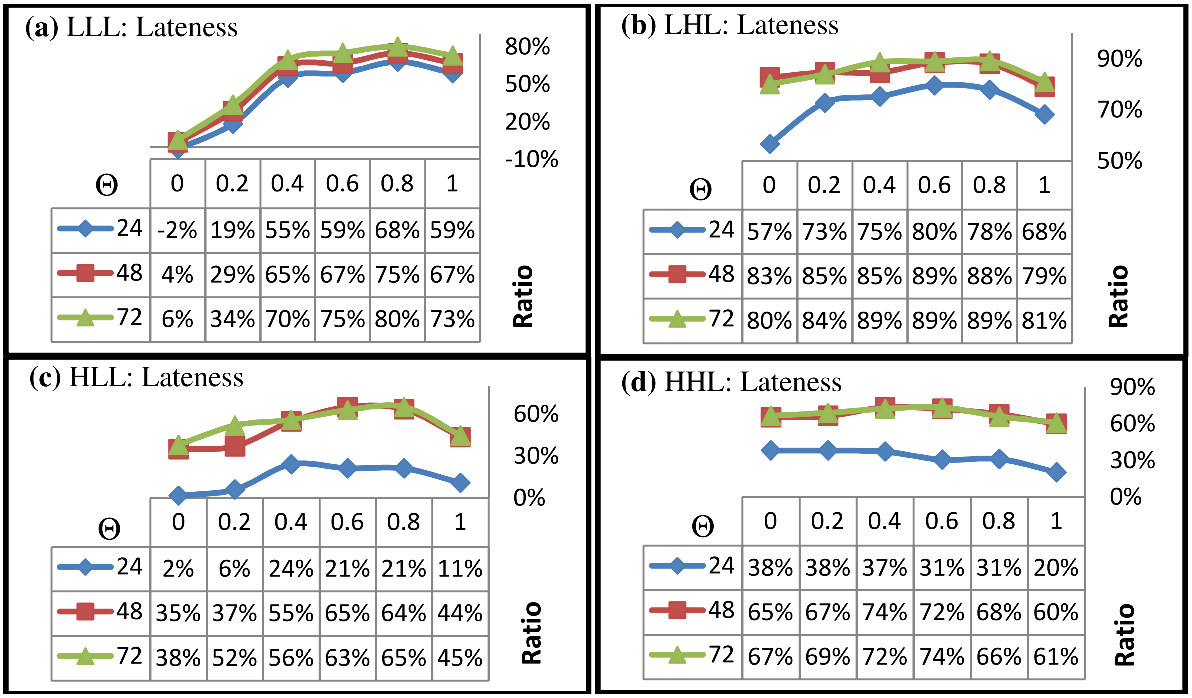


Figure 3.5. Simulation results for combinations with low load density

Figure 3.6 depicts the simulation results on the normalized profit for the case of high load density. The charts of this figure are interpreted as follows.

- **Fig. 3.6, chart (a):** Low service radius-Low trip length-How load density (LLH)

  If chart (a) of Figure 3.5 is compared to chart (a) of Figure 3.6, it is easy to explain why a smaller Θ (Θ=0.4) yields a higher normalized profit. A less conservative dispatching policy makes more profitable choices because the number of incoming loads is higher in the current combination. In the case of no lateness option, the dispatching policy tends to be more conservative because the loads cannot be served even if they are one minute late.

- **Fig. 3.6, chart (b):** Low service radius-High trip length-How load density (LHH)

  The normalized profit is insensitive to Θ in the range of 0.4-0.8 regardless of the size of the knowledge window and lateness option. This can be explained from the fact that even if the dispatching policy's level of conservatism leads to different loads selection, there are enough high revenue loads to make almost the same amount of profit. This argument is also valid when the service radius is high (Fig. 3.5, chart (d)).

- **Fig. 3.6, chart (c):** High service radius-Low trip length-High load density (HLH)

  The best Θ value (resulting in the highest profit) depends on KW. The best Θ values for 24, 48, and 72hrs advance load information are 0.4, 0.6, and 0.8, respectively. Under this setting, the decision maker with the shorter knowledge window should select a less conservative policy to make more profitable decisions.

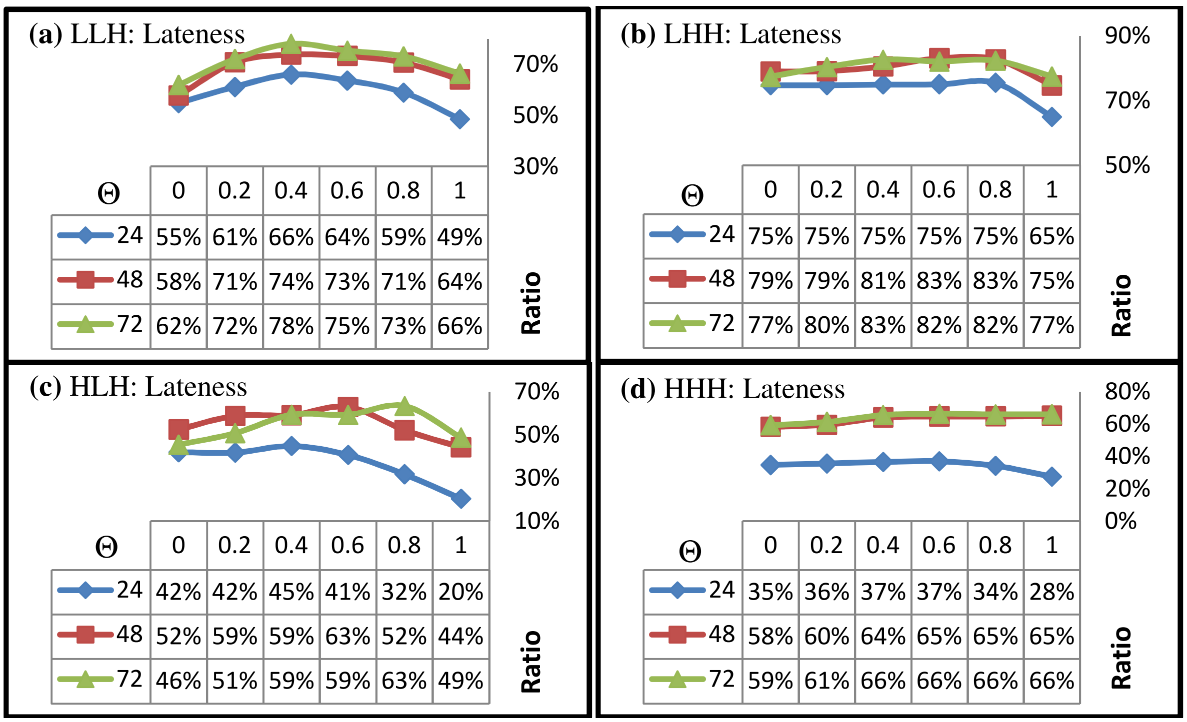


Figure 3.6. Simulation results for combinations with high load density

We have examined the performance of the algorithm under a variety of network settings. It is interesting to consider which of the settings often seen in practice. TRANSCORE conducted a carrier benchmark survey on more than 600 for-hire trucking companies in 2011 (www.transcore.com). The majority of surveyed companies (66%) were small companies with fewer than 6 trucks. The average trip length reported was about 900 miles with an average of slightly less than 3 loads per truck per week. The reported characteristic is very close to the LHL network setting examined in our simulation study. Thus, it is fair to state that the Pure-Θ Policy can produce quality solutions (almost 90% of static optimal solution) for a practical transportation network setting.

3.6.2 The Benefit of Advance Load Information

The benefit of advance load information is briefly discussed under the best choice of Θ. Figure 3.7a illustrates the normalized profit for all combinations with low load density. As explained earlier, the coding represents radius of service, trip length, and load density, respectively. For example LLL-24 is where all factors have low values and only first-day load information is available. Some interesting insights can be drawn from this figure. First, the Pure-Θ Policy often performs better when lateness is allowed. Second, the majority of benefit is gained by acquiring the second-day load information. Although access to the third-day load information yields much smaller marginal benefits, it is still worthwhile when the majority of loads are short because it helps the carrier to select more profitable sequence(s) of short loads. However, this small benefit disappears when most loads are high revenue. Finally, access to the second-day load information becomes crucial when the carrier operates in a larger service area (i.e., larger radius of service). This is because only one-day advance load information is not enough for the dispatcher to position the trucks in a vast area and thus many profitable loads might be lost.

The normalized profit for all combinations with high load density is shown in Figure 3.7b. Similar results are obtained from these test problems. At first glance, we can observe that the policy is often more effective when lateness is allowed. There is a benefit from getting advance load information, however, the margin of benefit decreases as KW increases to the third-day load information. Moreover, similar to the low-density case, the second-day load information is very important for carriers that operate in a larger service area. Finally, for carriers that operate in a smaller service area, the benefit of ALI is larger when the majority of loads are short. One possible explanation is that advance load information in such a setting provides a higher level of flexibility for the dispatcher to select and determine a sequence(s) of loads.

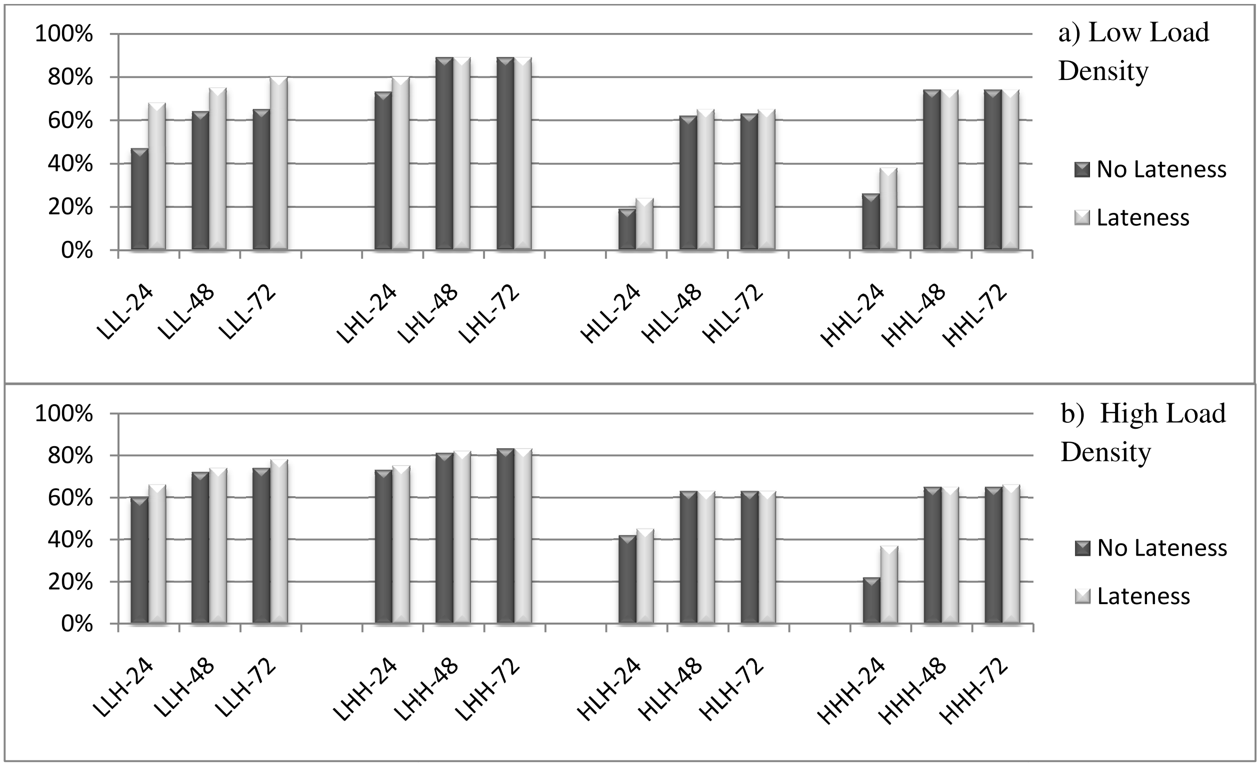


Figure 3.7. Normalized profit with the best choice of Θ

3.7. Policy Comparison

Based on the evaluation of the Pure-Θ Policy, we observed its very good performance for one of the most practical transportation settings. For some less practical settings, as mentioned earlier, low optimality ratios cannot be considered with certainty as a low performance of the Pure-Θ Policy. Thus, we take the following two steps for additional analysis. First, the enhanced version of the Pure-Θ Policy is developed based on a widely used approach in the DVRPs literature. Second, the Pure-Θ Policy and its enhanced version are numerically compared with two other dispatching methods to provide a better understanding of their efficiency.

3.7.1 Enhanced Deadhead Coefficient Policy

The literature approach we used to enhance the Pure-Θ Policy is the Multiple Scenario Approach, hence we refer to this enhanced version as the MSA-Θ Policy. The multiple scenario approach (also referred to as sampling method) is a widely used approach for incorporating common features of solutions to create a good plan (Pillac et al., 2011).

Bent and Van Hentenryck (2004) proposed the multiple scenario approach (MSA) for partially dynamic vehicle routing problems with stochastic demands. Experimental results show a dramatic improvement compared with the approaches not using stochastic demand information. Ichoua et al. (2006) also proposed a solution method for a dynamic stochastic vehicle routing problem with time windows (VRPTW). Their proposed method, which extended the parallel Tabu search by Gendreau et al. (1999), benefits from stochastic knowledge of future demand.

Hvattum et al. (2006) studied dynamic stochastic vehicle routing problems with time windows. Minimizations of the total traveling distance and the number of used vehicles were considered as the objective functions. They developed a deterministic model for the VRPTW and then extended it to a two-stage stochastic one. Since computing the expected recourse function is extremely hard, a sampling approach called hedging heuristic was proposed. A recent research in the dynamic stochastic context is the study by Schilde et al. (2011). They analyzed a daily problem arising in the Austrian Red Cross. The problem of serving patients between their home and hospital was modeled as dial-a-ride with the expected return transport. To solve the problem, they proposed four variants of variable neighborhood search. Only two of the proposed meta-heuristics take the advantage of stochastic demand information.

Since all the multiple scenario approaches in the literature were used for less than truckload problems, the method had to be modified for a proper implementation of our MSA-Θ Policy in truckload situations. Among the available approaches in the literature, the hedging heuristic by Hvattum et al. (2006) deals with multiple period problems, which are closer to the proposed problem in this study. The MSA can improve the deadhead coefficient policy from two aspects: 1) it can suggest a dwelling strategy before returning drivers to the home depot 2) it can virtually extend the knowledge window of the dispatcher by generating multiple scenarios.

At each decision epoch, based on the predefined knowledge window, new load(s) may enter into the system. Since the dispatcher does not have any information about the loads' arrivals after the KW, a number of scenarios are generated for an interval (referred to as the scenario interval) beyond the KW of the dispatcher. These loads are generated based on the transportation network characteristics. Since the MSA operates with a number of scenarios, it first needs a consensus function to develop a final plan (called the distinguished plan) to be implemented at each decision epoch. Second, it should decide whether to keep drivers at the delivery location of the last assigned load or return them to the depot. Table 3.6 presents the outer loop of the MSA-Θ Policy while Table 3.7 explains how a distinguished plan is formed at each decision epoch.

Table 3.6. The outer loop

A. At the start of each decision epoch
    a. Take the distinguished plan of the previous interval and freeze the plan up to the current time. Then, update the attributes of all loads and trucks.
    b. Add the new loads to the system (i.e., the realized loads during the last interval)
    c. Find the distinguished plan for the current interval by using the MSA sub-procedure
B. Evaluate the solution during the middle two weeks

Based on the abovementioned points, the MSA has different parameters, which should be set carefully for achieving a quality solution. These parameters are 1) the number of generated scenarios (ω) at each decision epoch, 2) the backhaul (deadhead) coefficient (Θ), 3) scenario interval (δ), 4) acceptance threshold (Ψ), and 5) waiting threshold (Φ). The first parameter is self explanatory and the second one is the deadhead coefficient discussed as introducing the Pure-Θ Policy. The acceptance threshold (Ψ) is a measure to determine frequently accepted loads in the ω scenarios. A load is considered frequently accepted if the number of scenarios where the load is accepted, among all scenarios, is greater than or equal to Ψ. The scenario interval (δ) is the time interval after the KW for which the algorithm generates future loads (called stochastic loads). Finally, waiting threshold (Φ) is used to decide whether a truck should wait or return to the depot after serving the last load of the sequence. In a single scenario, scheduling a truck to serve a stochastic load (as the last load) is a direct indication that waiting for future loads is justifiable. Thus in all ω scenarios, one can count the number of scenarios that waiting is recommended for a truck. If this frequency is larger than the waiting threshold (Φ), the recommendation for that truck is to wait. The waiting decision can be changed at the next decision epoch.

Table 3.7. The MSA-Θ sub-procedure

1. Generate ω sample scenarios beyond the KW for the next δ intervals. Each sample scenario includes known loads which are within KW and the stochastic loads that come after the KW.
2. Develop a preliminary plan by repeating the following for all generated scenarios.
   2.1. Call the Gorubi solver to optimize the mathematical model for the sample scenarios.
   2.2. For *i* :=1 to number of available trucks, repeat the following:
      2.2.1. Find a load that appears the most as the first load of truck *i* while disregarding the ones that have been already placed. If the relative frequency of the selected load is greater than or equal to the acceptance threshold (Ψ), the truck *i* sequence is formed by scheduling that load at the first stop of that truck.
      2.2.2. The process of constructing the truck *i* sequence continues by finding the subsequent loads using a similar approach. While forming the truck *i* sequence, a load is served through the depot only if the majority of solutions in the pool support such a decision. It is also checked that the formed truck sequence does not violate the allowable number of hours that a truck driver can be away from the depot.
3. Generate the distinguished plan from the preliminary one. For each truck, count the number of scenarios where waiting is reasonable for a truck. If this number is larger than the waiting threshold (Φ), the recommendation is to wait.
   3.1. If stochastic loads are scheduled after a known load, this suggests waiting at the delivery location of the known load.
   3.2. If stochastic loads are scheduled at the beginning of a truck sequence, this means that nothing is assigned to the truck and so it waits at the delivery location of the load currently being served (if it is moving loaded) or at its current location.
   3.3. Stochastic loads are removed from the truck's sequence to form the distinguished plan

The preliminary numerical experiments reveal that increasing the number of scenarios slightly above 25 (e.g., 30 or 40) is not statistically helpful to improve the accuracy of the algorithm. Based on this observation, we set the number of scenarios equal to 25 and tune the other parameters as follows. The values 0.2, 0.5, and 0.8 are considered for each parameter Φ, Ψ, and Θ. These parameters were chosen (from the range of 0 to 1) because the preliminary experiments showed that the algorithm is insensitive within ±0.1 of the selected values. The scenario interval (δ) is set to one, two, and three days. Although the number of parameters was reduced to only four factors with a few levels, the resulting 81 combinations is still large. In order to overcome this issue, we use the idea of orthogonal array in Taguchi method (Taguchi and Yokoyama, 1994). This method is a statistical tool which helps us not only to identify the importance of each parameter but also suggests which combination of parameters result in a higher performance without checking all possibilities. Using Taguchi method, the parameters of MSA are set by checking only nine setting combinations (L_9 design). The parameters' values (Φ=0.5, Ψ=0.5, Θ=0.8) consistently result in higher performance of the algorithm across different combinations. However, the best choice of scenario interval (δ) depends on how much load information is available at the time of decision making. The δ value decreases when the KW increases, so the values are one, two, and three days for 72hrs, 48hrs, and 24hrs KWs, respectively.

3.7.2 Comparison with Other Policies

The proposed policies (Pure-**Θ** and MSA-**Θ**) are compared with two other dispatching methods. The first one is rooted in practice and used by small trucking companies (called a practical policy). The second one is only based on the multiple scenario approach and statistically independent of **Θ** value (called Pure MSA). Given the distinct features of our problem when compared to other problems reported in the literature (e.g., returning trucks back home on a regular basis), there are limited appropriate policies available.

Practical Policy (PP): The steps of this policy are developed according to the Operations and Traffic Department of Logikor Company (introduced earlier). The details of this algorithm are depicted in Table 3.8. This process is often a manual task in small trucking companies. The dispatcher starts with a truck with the lowest remaining service hours. Then, the loads are visualized on a map and a sequence with the largest positive contribution is assigned to the truck. Although this is a manual process, we improve it by checking all feasible sequences (this is not a computational drawback because of the small number of new loads at each decision time). Finally, the process continues for the remaining trucks. The contribution of each truck is computed based on the earned revenue, moving cost (either empty or loaded), lateness cost, and dwelling cost. Moreover, in order to reduce the attractiveness of isolated locations, a portion of the average empty movement cost from the delivery location of the last load to other pickup locations is deducted from the truck contribution. This portion is almost half of the overall average cost to avoid overestimating the potential cost of the next empty movement. This is to hedge against possible empty movement cost as a result of choosing a particular load.

Table 3.8. The Steps of the Practical Policy

A. At the start of each decision epoch
- a. Sort the trucks in ascending order of their remaining service hours
- b. Sort the available new loads in ascending order of their pickup times and put them in a set called $\aleph$
- c. Repeat the following for each truck
  - c.1. Create all feasible load sequences
  - c.2. Calculate the truck contribution for feasible sequences. If there is no sequence with positive contribution, move to the next truck; otherwise go to step c.3
  - c.3. Schedule the truck to serve a sequence with the largest contribution
  - c.4. Update the set $\aleph$ by removing the currently assigned loads

B. Evaluate the solution during the middle two weeks

Pure MSA: This policy is developed according to the main concept of the Pure-**Θ** Policy (in Section 2) where the **Θ** coefficient only applies to the cost of the last movement. As we have observed in the previous section, the impact is remarkable in case of practical advance load information (e.g., 3 days). However, it is evident that sufficiently large advance load information lowers the **Θ** impact with regard to cost of other movements. After conducting statistical tests, it was observed that the **Θ** impact becomes insignificant if the scenario interval (δ) is set to at least 10 days regardless of the ALI choice. Despite the computational deficiency (because of a long scenario interval), the solution quality of the Pure MSA can be viewed as a proper benchmark. The other parameters of this policy are tuned with aid of the same previous approach (Taguchi method): ω=50, Φ=0.5, and Ψ=0.5.

As mentioned earlier, the Pure-Θ Policy performs slightly better when lateness is allowed. Thus, we considered combinations without the lateness option to provide sterner test of our proposed methods. Table 3.9 provides information on CPU time and normalized profit (ratio of optimality) for all four dispatching policies under different network settings. Some interesting and intuitive results can be obtained by comparing the Pure-**Θ** Policy with the MSA-**Θ** Policy. First, the benefit of scenario generation becomes more important when the dispatcher knowledge window is limited. Thus, with only one-day load information, the MSA-Θ Policy yields a higher normalized profit than the Pure-Θ Policy in almost all factor combinations. Next, while it is true that CPU time increases dramatically compared to the Pure-Θ Policy, the solution times at each decision epoch are still less than one minute. Finally, the impact of scenario generation disappears when load density is high, most loads are long, and more than one day ALI is available. It is easy to see that in such a good market, the Pure-Θ Policy brings the same benefit but much faster. Table 3.9 also shows that, in comparison to the PP, our Pure-Θ Policy yields consistently superior solutions; beyond 24 hours of ALI, this marginal of superiority is particularly substantial. The Pure MSA Policy is just marginally superior to the MSA-Θ Policy: an average of only 1.2 percentage points with ALI=24. However, as the table also shows, the small improvement in solution quality comes at a significant computational price. This reinforces the earlier observation that our proposed policies represent a better tradeoff between solution quality and run time. Thus, whether compared to what occurs in practice (the PP) or to a more sophisticated model grounded in the scientific literature (i.e., the Pure MSA Policy), it is clear that the policies we have proposed are very competitive alternatives.

Table 3.9. The proposed policies versus the Practical Policy (PP) and the Pure MSA

KW: 24 hrs	Ratio of Optimality				CPU Time (Sec) at Each Decision Epoch			
CODE	Pure-**Θ**	MSA-**Θ**	PP	Pure MSA	Pure-**Θ**	MSA-**Θ**	PP	Pure MSA
LLL	47.9%	52.4%	33.2%	52.6%	0.01	6.0	<0.01	120.4
HLL	18.7%	26.3%	16.4%	28.6%	0.01	6.2	<0.01	119.0
HHL	25.3%	31.1%	25.0%	33.9%	0.02	6.5	<0.01	150.4
LHL	73.3%	81.2%	64.1%	82.4%	0.02	6.4	<0.01	124.4
LLH	60.8%	66.1%	43.0%	66.4%	0.04	23.1	<0.01	396.1
HHH	22.3%	32.9%	16.7%	33.4%	0.08	24.7	<0.01	482.1
HLH	42.2%	45.2%	28.5%	46.3%	0.05	24.2	<0.01	486.2
LHH	73.1%	76.3%	64.7%	77.3%	0.05	23.4	<0.01	397.5
Average	**45.4%**	**51.4%**	**36.4%**	**52.6%**	**0.04**	**15.1**	**<0.01**	**284.5**
KW: 48 hrs	**Ratio of Optimality**				**CPU Time (Sec) at Each Decision Epoch**			
LLL	64.4%	69.6%	36.5%	69.7%	0.02	5.9	<0.01	132.5
HLL	62.8%	69.8%	41.9%	71.1%	0.03	6.0	<0.01	126.1
HHL	74.2%	79.5%	58.9%	79.8%	0.03	6.3	<0.01	157.2
LHL	89.0%	91.8%	69.5%	92.3%	0.03	6.1	<0.01	137.3
LLH	72.4%	74.6%	48.6%	75.1%	0.06	20.1	<0.01	397.3
HHH	64.9%	64.9%	50.8%	65.2%	0.11	21.6	<0.01	498.6
HLH	63.3%	64.8%	53.5%	64.8%	0.08	21.2	<0.01	488.4
LHH	81.2%	81.2%	69.4%	81.5%	0.06	21.1	<0.01	402.9
Average	**71.5%**	**74.5%**	**53.7%**	**74.9%**	**0.05**	**13.5**	**<0.01**	**292.5**
KW: 72 hrs	**Ratio of Optimality**				**CPU Time (Sec) at Each Decision Epoch**			
LLL	65.4%	71.2%	40.6%	71.5%	0.03	5.7	<0.01	143.9
HLL	63.5%	70.4%	42.2%	70.7%	0.05	5.7	<0.01	137.6
HHL	74.3%	79.7%	59.1%	80.0%	0.06	5.9	<0.01	160.9
LHL	89.4%	91.8%	69.9%	92.3%	0.05	5.8	<0.01	144.0
LLH	74.8%	77.1%	48.9%	77.2%	0.13	17.5	<0.01	404.4
HHH	66.0%	66.0%	54.2%	66.2%	0.29	19.9	<0.01	510.4
HLH	64.0%	64.9%	53.9%	65.3%	0.21	19.3	<0.01	501.1
LHH	82.5%	82.5%	71.6%	82.9%	0.17	18.2	<0.01	414.0
Average	**72.5%**	**75.5%**	**55.1%**	**75.8%**	**0.12**	**12.3**	**<0.01**	**302.0**

3.8. Conclusion and Future Research Directions

There are many research studies on long haul transportation dispatching rules that did not address the requirement of drivers and trucks to regularly return to their domiciles. This overestimates the capacity of the transportation network. Moreover, the majority of them assume that all loads information is available in advance. Thus, many of these models are not suitable to be implemented in a dynamic context. One contribution of this study is that it develops a comprehensive two-index MIP model that is flexible enough to include many operational details and can be implemented in a dynamic environment by using a rolling horizon approach. The two-index MIP is more efficient compared to existing general three-index models in the literature. Using the characteristics of the problem at the preprocessing stage along with the two-index MIP enables us to find the optimal solution of the static problem for small trucking companies.

Another contribution of this research is to develop a policy that can help carriers improve their razor-thin profit. To achieve this goal, a simple policy (deadhead coefficient policy/Pure-Θ Policy) was proposed and its performance evaluated under a wide variety of network settings through the simulation study. Although the static optimal solution is not a realistic bound, it is used as a benchmark to normalize the obtained profit of the Pure-Θ Policy. The policy performs the best (almost 90% of static optimal solution) in one of the practical transportation network settings when the second-day load information is available, regardless of whether the lateness option is in effect.

Finally, we incorporated the idea of a multiple scenario approach in hedging heuristic by Hvattum et al. (2006) to improve the Pure-Θ Policy. The MSA-Θ Policy has a more noticeably higher solution quality when the knowledge window is limited. The average ratio of optimality improves from 45.4% to 51.4% when only the next day load information is available. The margin of benefit will decrease as the dispatcher knowledge window increases. Moreover, the possible benefit can disappear if a carrier operates in a good market with more than one day ALI. Also, the Pure-Θ Policy and MSA-Θ Policy were compared against two other policies (namely a Practical Policy and the Pure MSA). The numerical experiments show that our proposed policies are competitive dispatching alternatives in terms of solution quality and computational efficiency.

In practice, the home depot is usually close to the area with more demands (loads). Since there is one home depot in our simulation study and loads are uniformly generated within the area of service, it is a reasonable assumption to consider the location of the depot at the center. Finding the optimum location of depot is beyond the scope of this work and can be viewed as a future research direction. Another possible research direction is to address information uncertainty since loads information (e.g., pick-up time or cancellation) may change even after it is received by the carrier. Issues such as truck breakdowns and accidents can also be considered by relaxing the assumption of constant traveling time. Eventually, designing efficient dispatching policies to handle large trucking companies with a few hundred trucks and drivers can be viewed as other interesting extensions.

CHAPTER 4

OPERATIONAL FLEXIBILITY IN THE TRUCKLOAD TRUCKING INDUSTRY

4.1. Introduction

In the trucking industry, many freight transportation service providers (also known as carriers) face highly variable demands from clients as well as other challenging issues reported by the American Trucking Associations (ATA, 2014). The tight US regulations on hours of services followed by driver shortage and high driver turnover are among the most important concerns. These challenges along with the rise in operational cost (e.g., driver wage) and market fluctuations have forced many small trucking companies to file for bankruptcy, e.g., in the first quarter of 2014, 390 carriers with 10,650 tractors went out of business. The Ontario Trucking Association also reports that Canadians face similar concerns (http://ontruck.org/).

To survive in this environment, carriers are continuously investigating various strategies to improve their operational efficiency. One of the major and indisputable operational issues is empty repositioning of the assets (Crainic, 2000; Wieberneit, 2008; Özener et al., 2011). The statistics on empty repositioning signal sub-optimal operational efficiency in the trucking industry. For example, empty mile as a percent of total miles are 22% for reefer fleets, 27.5% for private fleet flatbeds, and 21% for bulk operations in the US (http://www.logisticsmgmt.com). A similar issue reported by Barla et al. (2010) is that one in every three heavy trucks on major Canadian highways travels empty. Given the size of North America's trucking industry, empty repositioning costs carriers over a hundred billion dollars annually (Ergun et al., 2007a).

It is important to keep in mind that eliminating empty repositioning is rarely *a fruitful quest* since there are several empty repositioning determinants that are not under the full control of managers (e.g., geographic imbalance, market conditions, hours of service rules, trip length of loads). There are also some other potential factors (e.g., fleet size) over which managerial influence is limited, at least in the short run (Repoussis and Tarantilis, 2010). In such circumstances, collaboration and information sharing among logistic participants is viewed as an attractive alternative. Sharing advance load information (ALI) is considered because one of the least costly methods when freight transportation service clients and carriers collaborate with each other is to communicate timely load information (from clients to carriers) and pickup and delivery plans (from carriers to clients). The potential benefits of ALI were examined in a few research studies (e.g., Tjokroamidjojo et al., 2006; Zolfagharinia and Haughton, 2014) in the truckload context.

Apart from ALI, there are some other strategies that can help carriers improve operational efficiency. One of these strategies is diversion capability, which was first defined by Regan et al. (1995) as the dispatcher capability to divert an empty moving vehicle to serve a newly arising request for shipment delivery. Several studies indicate the potential benefit of diversion capability in the context of vehicle routing problems, VRPs (Ichoua et al., 2006; Branchini et al., 2009; Klundert et al., 2010; Respen et al., 2014, Ferrucci and Bock, 2015). However, with the exception of Regan et al. (1998), we are not aware of any other study that investigates this strategy in the context of the truckload trucking industry.

Another key strategy is to set an appropriate decision interval; i.e., the duration between time points at which the dispatcher makes the core operational decisions aimed at serving shipment requests (e.g., deciding which vehicle will serve a given request). A longer interval (i.e., lower decision frequency) means delaying the decisions to account for additional information on loads to be delivered. However, this benefit of more informed decisions comes at the detriment of newly arrived loads waiting longer to be taken into account. Most studies in the VRPs literature consider a continuous decision interval, triggered by a new load arrival, (Ichoua et al., 2006; Jaillet and Wanger, 2006; Branchini et al., 2009; Respen et al., 2014). An exception is Klundert et al. (2010) who discussed the possible benefit of extending the decision interval to one minute instead of every 30 seconds. Still, to the best of our knowledge, no study investigates the potential benefit of choosing an appropriate decision interval on the performance measure of a carrier.

Our present study to address this gap and several other gaps in the research literature is inspired by a low asset-based third party logistics provider (3PL) located in Ontario, Canada. They have a few drivers and tractors that operate in a relatively small geographic area. The next-day load information is often collected until late evening and the dispatching decision is made daily. The decision is whether to handle new loads using their own trucks or outsource them to other carriers. This problem is a pick-up and delivery with full truckload (DPDFL) in which load requests are realized as time progresses (i.e., dynamic nature). The primary goal of this study is to simultaneously investigate how a carrier's performance is affected by the above-mentioned three strategy factors: ALI, diversion capability and decision interval. A contribution of jointly studying multiple strategies over which a carrier has some control is to extend the scope of analysis beyond solely ALI. To achieve that goal, we focus on three key points. We first develop a mixed integer programming model that is flexible enough to properly handle the problem's dynamic aspects. Second, an efficient algorithm based on time-window discretization is developed and its convergence to optimality is proven. This algorithm is helpful for solving the problem's much larger static version: the version in which the carrier has advance information on all loads in the planning horizon of interest (e.g., a one-month horizon). Finally, we examine the impact of potential factors including the aforementioned strategies. For the purpose of this chapter, we define the term policy to mean any combination of the three strategy factors of interest here.

The rest of this article is organized as follows. In section 4.2, we review the related research works to position the current study in relation to the existing literature and to present its novelty.

Section 4.3 is devoted to defining the problem, stating its underlying assumptions, and formulating the proposed problem. In section 4.4, we explain how a special case of the problem formulation is used to develop an efficient algorithm for the static version. The major focus of section 4.5 is on dynamic implementation of the model, designing the experiments, solving the test problems, and conducting the statistical analysis. In section 4.6, we first evaluate the performance of the proposed algorithm (to obtain the benchmark solution) for the static version under different network settings. Then, multiple policies are compared against each other based on their deviation from the benchmark solution. This comparison helps us to draw valuable managerial insights. Finally, in section 4.7, we conclude our work and propose interesting future research directions.

4.2. Literature Review

In this section, the related studies are reviewed and classified based on the important features. Although the proposed problem is a truckload case, we also consider less-than-truckload studies that consider or test at least one of the following factors: advance load information, diversion capability, or decision interval.

4.2.1 Advance Load Information (ALI)/Knowledge Window (KW)

The dispatcher's KW is defined as how much advance notice the dispatcher has about relevant particulars on clients' loads (shipment requests); e.g., earliest and latest pick-up time. That is, the KW increases when the client of transportation services communicates load information further in advance of when loads are available for pick-up. A few studies have investigated the importance of ALI. The study by Powell (1996) proposed a stochastic dynamic load assignment problem formulation. He showed that when some stochastic information about future demand is available, the proposed model outperforms the deterministic one, which is updated as new information arrives. The model was evaluated under three conditions: fleet size, demand uncertainty and ALI. Not surprisingly, the stochastic model is superior with more fleet density, higher uncertainty but not with more ALI.

Mitrović-Minić et al. (2004) developed a double-horizon heuristic algorithm for the same-day dynamic pick-up delivery problems with time windows. The heuristic solved the problem with short-term (minimizing total distance) and long-term goals (efficiently serving future requests). The benefit of ALI was found to be positive but smaller for larger instances. Also, Jaillet and Wanger (2006) addressed the benefit of advance information for two variations of the traveling salesman problem. By defining the notion of disclosure dates for incoming requests, they analytically showed how ALI helps to improve competitive ratios.

Recently, the studies by Tjokroamidjojo et al. (2006) and Zolfagharinia and Haughton (2014) evaluated the benefit of ALI. Zolfagharinia and Haughton (2014) extended the earlier study by accounting for the requirement that drivers must regularly return to their home base. They found that the majority of profit improvement is attainable from acquiring the second-day load information.

4.2.2 Diversion Capability

As mentioned earlier, truck diversion is a model capability of changing the immediate destination of an empty truck (not a loaded one) to serve a new request. Regan et al. (1998) developed a dynamic framework to simulate the operations of small trucking companies. They evaluated the performance of relatively easy-to-implement and fast heuristics in truckload operations. The combinations of three load acceptance rules, eight assignment rules, and two modification strategies (including diversion capability) were taken into account. The authors found significant profit improvement from diversion capability. However, incorporating simple heuristic rules might not take full advantage of available information and result in myopic decisions. In their model, Zolfagharinia and Haughton (2014) incorporated the diversion capability but did not investigate its benefit. Even though we are not aware of any other work considering diversion capability in truckload transportation, there are multiple studies in the literature on vehicle routing problems, which are briefly reviewed below.

Ichoua et al. (2006) addressed a pick-up (or delivery) problem with courier service applications. The customers should be served within their time windows where service delay is also acceptable but penalized. The objective was to minimize the distance and lateness costs. By incorporating a diversion option as a dynamic rule, Ichoua et al. (2006) enhanced the Tabu search algorithm that Gendreau et al. (1999) developed. They found, through numerical experiments, that diversion capability can reduce the cost by 4.3%.

Branchini et al. (2009) consider a real-life Brazilian transportation problem with a relatively low level of dynamism where 60% to 80% of requests are known before the day starts. The objective was to maximize profit (revenue – cost of traveling – lateness cost). They developed three heuristics (Nearest Neighbor, Best Insertion, and Granular Local Search) and included diversion capability along with two other strategies. They found that the diversion reduces travel distances and customer rejection. However, the improvement in profit is not very significant.

The impact of diversion capability on the traveling salesman problem (TSP) was investigated by Klundert et al. (2010). Their work was inspired by the largest service organizations in the Netherlands. Their problem was different from the traditional TSP in the sense that requests materialize as time progresses and customers are mobile (e.g., leased cars). Their analyses reveal that the diversion of salespersons can improve the system responsiveness by an average of 20% (measured as the customer's average waiting time for receiving service). In a more recent study, Respen et al. (2014) evaluated the impact of vehicle tracking devices on the performance of VRPs with soft time windows and dynamic traveling times. The obtained results show significant reductions in total travel time and lateness.

As previously noted, there could be some advantages in operational flexibility through diversion. However, the benefit depends on the measure of performance and the structure of the problem. The margin of benefit is quite important because of the potential diversion drawbacks (Ferrucci and Bock, 2015). For one thing, technical devices must be installed for communications between drivers and the dispatcher(s). Moreover, because drivers may need to react quickly (e.g., sudden lane switch on a highway) and must deal with the visual distraction of using devices such as the truck's navigating system, there is an ever-present risk of driver errors that can cause road accidents (Young and Salmon, 2012; Stavrinos et al, 2013).

4.2.3 Decision Interval

For conformity with the literature, the decision interval (defined in section 4.1) is taken here to be synonymous with the re-optimization interval since the objective at each decision time is optimize dispatch operations in light of new information since the previous decision time. In some studies, dispatching decisions (e.g., load-vehicle assignments) are triggered when a new request enters the system. Because the times of those entries are dynamic, the decision times are not known in advance (this is called the continuous decision case). Other studies assume that the decision times are predefined (the discrete case). The discrete case is often seen in the truckload context (Powell , 1987; Powell et al., 1988; Powell, 1996; Powell et al., 2000; Godfery and Powell , 2002; Tjokroamidjojo et al., 2006; Zolfagharinia and Haughton, 2014; Zolfagharinia and Haughton, forthcoming). This helps carriers to receive more inputs before taking action about the requests. In the discrete case, the observed decision interval ranges between four hours (Godfery and Powell, 2002) to daily decisions (e.g., Tjokroamidjojo et al., 2006). However, the existing works in the truckload literature neither explain their choice of discrete decision interval nor investigate its impact on carriers' efficiency. The present work fills that gap in the extant literature.

4.2.4 Research Contributions

To highlight the other novel aspects of this work, we carefully point out the limitations of relevant works in the truckload literature (Powell, 1987; Powell et al., 1988; Powell, 1996; Regan et al., 1998; Yang et al., 1998; Powell et al., 2000; Godfery and Powell, 2002; Yang et al., 2004; Tjokroamidjojo et al., 2006; Zolfagharinia and Haughton, 2014). Table 4.1 highlights the key features within this body of works in order to cast a clear light on the features that represent sources of novelty for the present work. A concise and informative picture of this body of work now follows as a preamble to clarifying the novelty of our contributions.

As Table 4.1 shows, a very persistent feature for works on truckload (TL) problems is the use of tour capability; i.e., designing continuous truckload routes. We incorporated this in our choice of modeling approach because it is well known to be particularly important for the type of TL carrier operation we study: small trucking operations within relatively small geographic regions (i.e., unlike large nationwide or international carrier operations for which the average time to serve a load is very long: two to four days). Prominent in the stream of the literature on the large carrier context are works by Powell and colleagues who simplified the problem to different versions of assignment problems. Relevant works in the literature stream on small TL carrier operations include Regan et al. (1998) who developed heuristic rules for the continuous decision case defined earlier. Other studies in this latter stream used mixed integer programming to formulate the problem and a rolling horizon approach for implementation (Yang et al, 1998; Yang et al, 2004; Tjokroamidjojo et al., 2006; Zolfagharinia and Haughton, 2014).

As noted in subsection 4.2.2 on diversion capability, the work by Regan et al. (1998) relied only on heuristic rules. The major concern is that simple heuristic rules do not take full advantage of available information and may lead to myopic decisions. The defined problem was the same in the studies by Yang et al. (1998, 2004). The objective was to minimize the total cost (including delay, empty and loaded movement, and load rejection costs). However, the very important cost of vehicle dwelling (waiting at the pick-up location for a load) was not part of their model. Another limitation of their model is that it does not consider any subcontracting option, which is a very common practice in reality.

Tjokroamidjojo et al. (2006) addressed a full truckload pickup and delivery problem in which the carrier's total cost was taken into account. They also investigated how much a trucking company can reduce cost by obtaining additional information further in advance. However, their proposed mathematical model was subject to some limitations. For example, their method was unsuitable if delay is permissible. Moreover, diversion was not part of the model and the decision interval was set as daily (without testing its potential impact). Compared to the work by Zolfagharinia and Haughton (2014), the problem in this study is simpler because its focus on local truckload operations means that it need not consider the issue of regularly returning drivers to their home depot. Nonetheless, we develop an efficient algorithm that can be extended to solve more general cases like Zolfagharinia and Haughton (2014). Summarizing the limitation of related studies, we can put the contributions of this chapter in three broad categories:

- We developed a flexible mixed integer programming formulation to model the dynamic pickup and delivery problem faced by a real-world logistics provider. The special case of the model (no lateness allowed) is reformulated using integer programming.
- To provide a quality benchmark solution, we developed an efficient algorithm using the idea of time window discretization (introduced by Wang and Regan, 2002). We proved that the algorithm converges to the optimal total cost and test its computational efficiency.
- We uncovered managerial insights through a comprehensive simulation study. To the best of our knowledge, this is the first study to examine the impact of diversion capability and re-optimization interval in the presence of different levels of ALI. Moreover, this study assesses how different policies deviate from the benchmark solution.

Table 4.1. Summarizing the most related studies to the current study

Author(s) and publication year	Problem Type	Tour Capability	Demand Prob. Information	Objective Function	Modeling Approach	Diversion		Decision Interval	The Benefit of ALI	Sub-contract
						Included	Tested			
Powell (1987)	TL	X	√	Max. Profit	SF	X	X	Discrete, 24hrs	X	X
Powell et al. (1988)	TL	X	X	Max. Profit	SF	X	X	Discrete, 6hrs	X	X
Powell (1996)	TL	X	√	Max. Profit	SF	X	X	Discrete, 24hrs	√	X
Regan et al. (1998)	TL	√	X	Max. Profit	Heuristic Rules	√	√	Cont.	X	X
Yang et al. (1998)	TL	√	X	Min. Cost	MIP	√	X	Cont.	X	X
Godfery and Powell (2002)	TL	X	X	Max. Profit	ADP	X	X	Discrete, 4hrs	X	X
Mitrović-Minić et al. (2004)	VRP	√	X	Min. Total Distance	Tabu Search	X	X	Discrete, 15min	√	X
Yang et al. (2004)	TL	√	√	Min. Cost	MIP	√	X	Cont.	X	X
Jaillet and Wanger (2006)	TSP	√	X	Min. Cost	Heuristic Rules	X	X	Cont.	√	X
Ichoua et al. (2006)	DVRPTW	√	X	Min. Cost	Tabu Search	√	√	Cont.	X	X
Tjokroamidjojo et al. (2006)	TL	√	X	Min. Cost	IP	X	X	Discrete, 24hrs	√	√
Branchini et al. (2009)	DVRPTW	√	√	Max. Profit	Heuristic Rules	√	√	Cont.	X	X
Klundert et al. (2010)	TSP	√	X	Max. Responsiveness	IP	√	√	Discrete, 30 & 60Sec.	X	X
Respen et al. (2014)	DVRPTW	√	X	Min. Travel Time & Lateness	Heuristic Rules	√	√	Cont.	X	X
Zolfagharinia and Haughton (2014)	TL	√	X	Max. Profit	MIP	√	X	Discrete, 12hrs	√	X
Ferrucci and Bock (2015)	DVRP	√	X	Max. Responsiveness	Tabu Search	√	√	Discrete, 20Sec.	X	X
The present study	TL	√	X	Min. Cost	IP and MIP	√	√	Discrete, 12 & 24hrs	√	√

Note: TL=Truckload; TSP=Traveling Salesman Problem; DVRPTW=Dynamic Vehicle Routing Problem with Time Windows
ADP=Adoptive Dynamic Programming
IP: Integer Programming; MIP=Mixed Integer Programming
SF: Stochastic Formulation

4.3. Problem Definition

The dynamic pickup and delivery with full truckload (DPDFL) problem under study is defined by the following assumptions:

- The carrier has a fixed fleet of trucks.
- Each truck can handle one load at a time (i.e., full truckload transportation).
- The carrier knows of customers' demands (loads) gradually as time elapses.
- It adopts the literature's standard assumption that each trip is executed without a break.
- Each truck's attributes are current location and status.
- Each load has static attributes (the earliest and latest pickup time, the maximum permissible delay time, the pickup location, and the delivery location) and dynamic attributes (e.g., load has been previously accepted and waiting to be served by one of the company-owned trucks).
- The shipment cost is a linear function of travel time which itself is a linear function of distance.
- There is a hard time-window to serve a load. Thus, a load will be subcontracted if it cannot be served within the predefined time window.

Taking all the aforementioned assumptions into account, the optimal DPDFL solution specifies the carrier's cost minimization decisions concerning (i) whether to serve new loads using available trucks or subcontract them, and (ii) the sequence of accepted loads that should be served by each truck.

4.3.1 Common Mathematical Models

There are two common ways to formulate a DPDFL problem. The first approach uses an extended version of the assignment problem (e.g., assignment with timing constraints) to exploit the problem's characteristics. This is the most common approach in the literature (see Yang et al. 1998; Powell et al., 2000; Yang et al., 2004; Tjokroamidjojo et al., 2006; Zolfagharinia and

Haughton, 2014). In the second approach, the problem can be formulated as a variant of capacitated arc routing problems (CARP) in which each directed arc represents one load with designated origin and destination. A recent work by Liu et al. (2010a, b) proposed an integer-programming model to formulate CARP for truckload industries and a quality lower bound. They also developed a heuristic method based on graph theory to solve the proposed model since the exact method is incapable of handling large problem instances. However, they captured neither time windows nor the fleet size of the transportation network for fulfilling demands.

Comparing both approaches in the literature, the former is shown to be more promising to use because the dimensionality of the model grows quickly in the latter case. Among the related studies, the one by Tjokroamidjojo et al. (2006) used an effective approach to handle DPDFL. The utilized approach consists of two parts, a preprocessing part for time-based restrictions and an assignment problem afterwards. Since time-window restrictions are explicitly handled outside the mathematical model, the approach performs well by reducing the number of constraints and decision variables. Although our approach is similar to Tjokroamidjojo et al. (2006), we must handle some of the time-based constraints inside the MIP because most of the loads and trucks attributes are determined after solving the model (Zolfagharinia and Haughton, 2014).

4.3.2 The Model Inputs

The notations, parameters, and decision variables used in formulating the proposed model are presented below.

Notations:

I set of trucks, indexed by i, and u=1,…, $|I|$

J set of loads, indexed by j, k, and r=1,…, $|J|$

L set of dummy loads, indexed by j, k, and r= $|J|+1$,…, $|J|+|I|$

Parameters:

η_i : the location of truck i at the time of decision making

U_j: maximum permissible delay for serving load j

a_j: departure city of load j (pick up location)

b_j: destination city of load j (delivery location)

$D(.,.)$: travel time between any two points in the service area. It can be described as function of distance.

α_j: the earliest availability of load j

s_j: cost of subcontracting load j

c: the empty traveling cost per hour of driving

w: the penalty cost per hour for a truck being idle at any load location (dwelling cost)

l: the penalty cost per hour for late pickup

H: a very large positive number.

τ: time at the decision epoch

Decision Variable:

Y_{ik}: if truck i serves load k directly at the first stop, 1 otherwise 0.

X_{jk}**:** if load k is served immediately after load j, 1 otherwise 0.

Z_{ik}: if load k is served by truck i after another load, 1 otherwise 0.

O_k: arrival time at the pickup location of load k

The real-time location of each truck is important at each decision epoch because of the problem's dynamic nature. If the current location of truck i is denoted with η_i, $D(\eta_i, q)$ shows the traveling time from the current location of truck i to the location q. Dwell time is the waiting time experienced by a driver/truck if the truck must wait at any city location to pick up the next load. Although we consider the same dwell cost for all locations in this study, the model is flexible enough to address varying dwelling costs across locations.

To acknowledge dynamic features of the problem, we first define TST(i) as the status of truck i at the decision epoch τ. TST(i) can take two values 1 and -1 meaning truck i is moving loaded or empty (either moving empty or idle at any city location) respectively. If truck i is moving loaded (i.e., TST(i)=1) at the time of decision making (e.g., serving load j), it will be available at the later time, $\tau + D(\eta_i, \mathrm{b}_j)$ at the destination location of load j (i.e., the diversion is not allowed if a truck is moving loaded).

If a truck is empty, TST(i)=-1, then truck i is available for scheduling at time τ at its current location. There is also a need to keep track of load status which is denoted with LST(j). There are four possible load statuses. If load j is being served at the decision epoch, LST(j) is equal to 2. The other loads which were already served by company-owned trucks (i.e., LST(j) =0). The loads which are accepted but have not received service yet (i.e., LST(j) =1) enter the model. In order to distinguish new loads from the current ones (i.e., the loads being served, LST(j) =2, or waiting to be served, LST(j) =1), their statuses will be LST(j) =3. We also define ST(i,j) as a binary parameter to address the status of the truck and load together. If truck i is serving load j at the decision time, then ST(i,j) takes 1, otherwise 0.

To distinguish between pre-planning and diversion strategies, a parameter $v(j)$ is defined at the preprocessing stage. If an empty truck (e.g., truck i) is moving toward the pickup location of a load (e.g., load j), the pre-planning strategy freezes the assignment of load j to truck i (i.e., if the assignment of load j is fixed to truck i, $v(j)=i$ otherwise v(j)=0). This feature will be incorporated in the proposed mathematical model.

4.3.3 Preprocessing Stage

As mentioned in section 4.3.1, we tackle the static version of the problem in two stages. The first stage is preprocessing. This stage consists of two phases. In the first phase, we show how to compute the necessary pieces of information. Then, in the second phase, we explain how the generated information is used to solve the proposed mathematical model.

4.3.3.1 Preprocessing Stage: Phase I

At each decision epoch, trucks and loads have different attributes. Based on their current statuses, the dwelling and lateness duration can be computed. The lateness can occur in two situations: i) truck *i* serves load *j* as the first load; ii) a truck serves load *k* after load *j*.

$\mathrm{DL0}(i,j)$: the lateness duration at the load pickup location *j* if truck *i* serves load *j* first. For, $\mathrm{TST}(i) = -1, \mathrm{DL0}(i,j) = max\left(0, \mathrm{D}(\eta_i, a_j) + \tau - \alpha_j\right)$. If the truck is moving loaded, $\mathrm{TST}(i) = 1$, toward the destination of a load (e.g., load *k*), $\mathrm{DL0}(i,j) = max\left(0, \tau + \mathrm{D}(\eta_i, b_k) + \mathrm{D}(b_k, a_j) - \alpha_j\right)$.

$\mathrm{DL1}(j,k)$: the *minimum* lateness at the load pickup location *k* if the same truck serves load *k* immediately after load *j*. Load *k* will experience some lateness if there is not enough time to reach the pickup location of load *k* immediately after serving load *j*. This is denoted with $\mathrm{DL1}(j,k) = max\left(0, \left(\alpha_j + \mathrm{D}(a_j, b_j) + \mathrm{D}(b_j, a_k)\right) - \alpha_k\right)$.

To some extent similar to what is explained for calculating lateness time, truck dwelling might occur in following cases: i) truck *i* serves load *j* as the first load, ii) a truck serves load *k* after load *j*, iii) truck dwell time after delivery of the last assigned load, and iv) dwell time if the truck is not assigned to any load.

$\mathrm{DW0}(i,j)$: the dwell time at the load pickup location *j* if truck *i* serves load *j* first in the current decision epoch. For empty trucks, $\mathrm{TST}(i) = -1$, $\mathrm{DW0}(i,j) = max\left(0, \alpha_j - (\tau + \mathrm{D}(\eta_i, a_j))\right)$. If truck *i* is moving loaded toward the destination of load *k* at the current decision epoch, similar reasoning leads to dwell time being $\mathrm{DW0}(i,j) = max\left(0, \alpha_j - \left(\tau + \mathrm{D}(\eta_i, b_k) + \mathrm{D}(b_k, a_j)\right)\right)$ in which $\tau + \mathrm{D}(\eta_i, b_k)$ is when truck *i* is available after completing the service of load *k*.

$\mathrm{DW1}(j,k)$: the *minimum* dwell time of a truck at the pickup location of load *k* if it comes directly after serving load *j*, $\mathrm{DW1}(j,k) = max\left(0, \alpha_k - \left(\alpha_j + U_j + \mathrm{D}(a_j, b_j) + \mathrm{D}(b_j, a_k)\right)\right)$.

It is important to note that all the calculated minimum values (e.g., $\mathrm{DW1}(j,k)$;) turn to actual values if no lateness is allowed. This means that the exact lateness and dwell times can be computed in the pre-processing stage under this assumption.

4.3.3.2 Preprocessing Stage: Phase II

In this phase, the following two tasks are performed: 1) updating all dynamic attributes of trucks (e.g., current truck location) and loads (e.g., a load is waiting to be served or being served), and 2) identifying infeasible truck-load and load-load combinations. Since the first task is straightforward, only the last function of the preprocessing stage is discussed here.

Given the current status of the trucks, we determine whether a particular truck can serve a certain load. This must be done for all available truck-load combinations. It is obvious that certain truck-load combinations are not feasible if the truck cannot be available at the pickup location of the load without violating the maximum delay. Thus, the following modifications are applied to the decision variables: $Y_{ij} = 0 \text{ if } \mathrm{DL0}(i,j) > U_{\mathrm{j}}$.

Similar to what is done for truck-load combinations; we examine the feasibility of serving load k immediately after load *j*. Here, the best-case scenario for load-load combinations is determined. The best possible case is if load j is served on time so that no delay is carried toward serving load k. It is evident that load k cannot be served after load j when there is not enough time for the truck to be at the load k pick-up location without violating its time window. Thus, the following adjustments are done because if a load-load combination is not feasible in the best-case scenario, it cannot be feasible at all (i.e., if $\mathrm{DL1}(j,k) > U_k$ then $X_{jk} = 0$). On the other hand, if the minimum lateness is smaller than or equal to the maximum allowable delay of U_k, the combination is not conclusively infeasible. This is extremely important because the decision at this stage is made based on the minimum lateness, but not the actual lateness. Therefore, considering different possible assignment decisions, some load combinations with $\mathrm{DL1}(j,k) \leq$

U_k might be infeasible after solving the problem. This illustrates why we need to have time components in the mathematical model.

4.3.4 Mathematical Model

Having defined all parameters and dynamic aspects of the model in the preprocessing stage, it is time to formulate the conceptual model. Before formulating the proposed model, the two following points should be addressed.

First, it is important to check which loads enter the model and their notations. As defined earlier, J represents the set of all loads entered in the model. However, we are required to differentiate them in order to have a neat mathematical formulation. To do so, the set of new loads is denoted with $\bar{J}$ (LST(j) =3), the set of accepted jobs waiting for service, $\bar{\bar{J}}$ (LST(j) =1), and the set of jobs being served at the decision epoch, $\hat{J}$ (LST(j) =2). Thus, $J = \bar{J} \cup \bar{\bar{J}} \cup \hat{J}$.

Second, we introduce a simple concept to calculate the dwelling cost that a truck experiences. Since the mathematical model is re-optimized during the knowledge window, a truck may incur dwell costs before picking up loads, after serving all the assigned loads, or when it is not assigned to any load. The approach is to introduce one dummy load for each available truck. These loads have zero trip length (i.e., $D(\mathrm{a}_j, \mathrm{b}_j) = 0,\ j \in L$). It is also assumed that the driving distance from the delivery location of all loads to the pickup location of these dummy loads is zero (i.e., $D(\mathrm{b}_j, \mathrm{a}_k) = 0,\ j \in J$ and $k \in L$). Moreover, the earliest pickup time of these loads are the same and equal to the predefined knowledge window (i.e., $\alpha_k = \tau + KW,\ k \in L$). Finally, in the second phase of the pre-processing stage, the possibility of serving dummy loads before any other load will be excluded (i.e., they must be served at the end of each truck sequence). Introducing dummy loads aids in calculating dwelling cost by only including the time that a truck has to wait for a load to become available (either an actual load or a dummy one). Having the parameters and decision variables defined, the model is then formulated as follows.

$$\mathbf{M1}: Min\ \ l \sum_{k\in(\bar{J}\cup\bar{\bar{J}})} max\,(0, O_k - \alpha_k) + w \sum_{k\in(\bar{J}\cup\bar{\bar{J}}\cup L)} max\,(0, \alpha_k - O_k) + \sum_{k\in\bar{J}} s_k \left[1 - \sum_{i\epsilon I}(Y_{ik} + Z_{ik})\right]$$

$$+ c \sum_{i\epsilon I, TST(i)=-1} \sum_{k\epsilon\bar{J}\cup\bar{\bar{J}}} D(\eta_i, \mathrm{a_k})\, Y_{ik} + c \sum_{i\epsilon I} \sum_{j\in\hat{J}, ST(i,j)=1} \sum_{k\epsilon\bar{J}\cup\bar{\bar{J}}} D(b_{\mathrm{j}}, \mathrm{a_k})\, Y_{ik}$$

$$+ c \sum_{k\epsilon\bar{J}\cup\bar{\bar{J}}} \sum_{j\epsilon\bar{J}\cup\bar{\bar{J}}} D(\mathrm{b_j}, \mathrm{a_k})\, X_{jk} \tag{4.1}$$

Subject to:

$$\sum_{i\epsilon I}(Y_{ik} + Z_{ik}) \le 1, \qquad k\epsilon\bar{J} \tag{4.2}$$

$$\sum_{i\epsilon I}(Y_{ik} + Z_{ik}) = 1, \qquad k\epsilon\bar{\bar{J}} \cup L \tag{4.3}$$

$$Y_{ik} = 1, \qquad i \in I, k \in \bar{\bar{J}}, v(k) = i \tag{4.4}$$

$$\sum_{k\epsilon\bar{J}\cup\bar{\bar{J}}\cup L} Y_{ik} = 1, \qquad i \in I \tag{4.5}$$

$$\sum_{k\in\bar{J}\cup\bar{\bar{J}}\cup L} X_{jk} - \left[\sum_{i\in I} Y_{ij} + \sum_{r\in\bar{J}\cup\bar{\bar{J}}} X_{rj}\right] = 0, \qquad .j \in \bar{J} \cup \bar{\bar{J}} \tag{4.6}$$

$$Z_{ij} + Y_{ij} + \sum_{u\epsilon I, u\neq i} (Z_{uk} + Y_{uk}) \le 2 - X_{jk}, \qquad i \in I, .j, k \in \bar{J} \cup \bar{\bar{J}} \cup L \tag{4.7}$$

$$\sum_{j\epsilon\bar{J}\cup\bar{\bar{J}}} X_{jk} - \sum_{i\epsilon I} Z_{ik} = 0\,, \qquad k \in \bar{J} \cup \bar{\bar{J}} \cup L \tag{4.8}$$

$$O_k - D(\mathrm{a}_j, \mathrm{b}_j) - D(\mathrm{b}_j, \mathrm{a}_k) - max(O_j, \alpha_j) \ge (X_{jk} - 1)H, \qquad j, k \in \bar{J} \cup \bar{\bar{J}} \cup L \tag{4.9}$$

$$O_k - D(\mathrm{a}_j, \mathrm{b}_j) - D(\mathrm{b}_j, \mathrm{a}_k) - max(O_j, \alpha_j) \le (1 - X_{jk})H, \qquad j, k \in \bar{J} \cup \bar{\bar{J}} \cup L \tag{4.10}$$

$$O_k - \sum_{i\epsilon I} \sum_{j\epsilon\hat{J}, ST(i,j)=1} [\tau + D(\eta_i, b_j) + D(b_j, a_k)]Y_{ik} - \sum_{i\epsilon I, TST(i)=-1} [\tau + D(\eta_i, a_k)]\, Y_{ik}$$

$$\ge \left(\sum_{i\epsilon I} Y_{ik} - 1\right)H, \qquad k \in \bar{J} \cup \bar{\bar{J}} \cup L \tag{4.11}$$

$$O_k - \sum_{i\epsilon I} \sum_{j\epsilon\hat{J}, ST(i,j)=1} [\tau + D(\eta_i, b_j) + D(b_j, a_k)]Y_{ik} - \sum_{i\epsilon I, TST(i)=-1} [\tau + D(\eta_i, a_k)]\, Y_{ik}$$

$$\leq \left(1 - \sum_{i \in I} Y_{ik}\right) H, \quad k \in \bar{J} \cup \bar{\bar{J}} \cup L \qquad (4.12)$$

$$O_k - \alpha_k - U_k \leq 0, \qquad k \in \bar{J} \cup \bar{\bar{J}} \qquad (4.13)$$

The objective function is composed of six terms. The first term computes the lateness cost during the interval over which the problem is optimized. Late service occurs when the truck arrives to the load's pick-up location after its availability. This penalty only applies to actual loads and not dummy loads. The second term captures total dwelling costs that all trucks experience. This is the cost of waiting at the pickup locations of loads (including dummy ones), which can occur when a load is either at the beginning of the sequence or after another load. The rest of the objective function calculates, respectively, the subcontracting costs and the empty repositioning costs (to serve loads with the carrier's own trucks).

The mathematical model has twelve constraint sets. The first two sets (4.2 and 4.3) ensure that all previous accepted loads (including dummy loads) will be covered, but there is no guarantee to take all new loads by company-owned trucks. Constraints (4.4) are preplanning restrictions to prevent trucks diversion. Constraint set (4.5) forces trucks to serve exactly one load at the beginning of a sequence. If the first load is dummy load for a truck, the decision for that truck is to wait. The next set of constraints (4.6) ensures that if load *k* is served after load *j*, load *j* is either scheduled to be the first load or placed after another load *r*. The constraints (4.7) ensure that load *k* can be scheduled after load *j* if they are visited by the same truck. The constraints (4.8) guarantee that a load is not scheduled at the beginning of a sequence if it is served after another load. Altogether, constraints (4.9)–(4.12) ensure that O_k does not take on an unrealistically large or small value to prevent dwelling or lateness costs. Constraints (4.9) and (4.10) apply when a truck is serving one load after another load directly while constraints (4.11) and (4.12) are for the case of a truck serving a load at the beginning of a sequence. Finally, constraints (4.13) guarantee that all accepted loads are served within U_k hours from their earliest availabilities.

4.3.5 Special Case: No Lateness is allowed

If no lateness is allowed, the mathematical model can be simplified by taking time-based constraints into the pre-processing stage. Keskinocak and Tayur (1998) used the same method for an Aircraft scheduling problem when the exact departure times should be met. This method was also utilized by Tjokroamidjojo et al. (2006) in the truckload trucking industry.

In this special case, constraints (4.9) to (4.13) are checked in the pre-processing stage and therefore eliminated from the mathematical model. Moreover, there should be a small modification in the objective function. This modification includes omitting the second term of the objective function (lateness cost) and replacing the total dwell costs by the following terms:

$$w \sum_{i \in I} \sum_{j \epsilon \bar{J} \cup \bar{\bar{J}} \cup L} \mathrm{DW0}(i,k) Y_{ik} + w \sum_{j \epsilon \bar{J} \cup \bar{\bar{J}}} \sum_{k \epsilon \bar{J} \cup \bar{\bar{J}} \cup L} \mathrm{DW1}(j,k)\, X_{jk} \qquad (4.14)$$

4.4. Developing a Benchmark

In this section, we propose an efficient method that can handle medium sized problems. This method helps us to gauge the efficiency of different policies compared to the lowest attainable cost where all information is available to the decision maker. This method is based on time windows partitioning. We adopt the idea of time window discretization, which was proposed by Wang and Regan (2002). They broke down time windows into several parts and treated each part as a sub-load. Our method is different in the sense that it breaks the time windows into several time points where each load can be only handled at one of those time points. Our approach is more efficient because it allows us to use the formulation of the special case (i.e., no lateness option).

4.4.1 Discretization of Time Windows

If truck i serves a load (e.g., load k), one of the following cases will occur. In the first case, the truck reaches load k earlier than α_k so it must wait. It is evident that the load is handled as soon as it becomes available. In the other case, the truck arrives at the pickup location of load k between α_k and $\alpha_k + U_k$. In this case, if the time window is modified to two single points (i.e., α_k and $\alpha_k + U_k$), the truck can handle the load at those times only.

LEMMA 1. If $l \geq w$, the adjustment of pickup time to $\alpha_k + U_k$ will never reduce the minimum total cost.

PROOF. It is enough to show that for any arbitrary feasible solution of the original problem, the time adjustment to $\alpha_k + U_k$ never reduces the total cost. Thus, for any arbitrary feasible solution of the original problem, we are required to check the impact of this adjustment on the current load (e.g., load k) and all the subsequent load(s). Regardless of the pickup location and time windows of the next load, this adjustment simply increases the lateness cost of the current load by $l \times [\alpha_k + u_k - \mathrm{O_k}]$. Without partitioning the time windows, if there is another load scheduled after load k (e.g., load r), there are only two possible scenarios.

1- The truck arrives at load r before the load availability: $O_\mathrm{r} = O_\mathrm{k} + \mathrm{D(a_k, b_k)} + \mathrm{D(b_k, a_r)} < \alpha_r$

Without any adjustment, the truck must wait for $\alpha_r - O_\mathrm{r}$. If the truck picks up load k at a later time (i.e., $\alpha_k + u_k$), it will spend less time waiting to pick up load r. The maximum saving on the dwell cost is equal to $w \times [\alpha_k + u_k - O_k]$. Since lateness cost is greater than or equal to dwell cost, the maximum possible saving is never greater than the cost increment.

2- The truck arrives at load r after its earliest availability, but without violating its maximum permissible delay: $\mathrm{\alpha_r} \leq \mathrm{O_r} = O_\mathrm{k} + \mathrm{D(a_k, b_k)} + \mathrm{D(b_k, a_r)} \leq \mathrm{\alpha_r} + U_r$

Since $\mathrm{\alpha}_k + u_k \geq \mathrm{O_k}$, applying the adjustment will result in one of following situations:

2.1. It may make scheduling of load r after load k infeasible (i.e., $\mathrm{O_r} > \alpha_\mathrm{r} + U_r$), thus excluding some of the feasible solutions will never lower the total cost.

2.2. Serving load r after load k may be feasible, but it inflates the lateness cost by $l \times [\alpha_k + u_k - O_\mathrm{k}]$.

Using the same approach, it is evident that this adjustment will never reduce the total cost for subsequent loads that can be scheduled after load *r*. Thus, the adjustment of pickup time to $\alpha_k + U_k$ will never reduce the minimum total cost.

COROLLARY 1. If $l<w$, replacing the lateness cost by the dwell cost and adjusting the pickup time to $\alpha_k + u_k$ will never reduce the minimum total cost.

PROOF. Follows from Lemma 1 and a simple property that replacement of the lateness cost by a larger value will never reduce the minimum total cost. The property is very trivial because for every dispatching decision with the lateness cost adjustment (**>** *l*), there exists a dispatching decision with a lower or equal total cost with the original l value.

Since the adjustment of pickup times to $\alpha_k + u_k$ makes the problem over-constrained (because it might eliminate some of the feasible load combinations), we denote the minimum total cost with Z^{over}. If $l \geq w$, Z^{over} is obtained after the pickup time adjustment to $\alpha_k + u_k$; otherwise we first replace l with w and then make the pickup time adjustment. Thus, based on Lemma 1 and Corollary 1, $Z^{over} \geq Z^{opt}$.

LEMMA 2. If $l \geq w$, the pickup time is set to α_k will never increase the minimum total cost.

PROOF. It is sufficient to show that for any arbitrary feasible solution of the original problem, the time adjustment to α_k never increases the total cost. Thus, for any arbitrary feasible solution of the original problem, we have to check the impact of the adjustment on the current load (e.g., load *k*) and the all the subsequent load(s). Regardless of the pickup location and time windows of the next load, this adjustment simply reduces the lateness cost of the current load by $l \times [O_k - \alpha_k]$. Without partitioning of time windows, if there is another load scheduled after load *k* (e.g., load *r*), only one of the following two scenarios will happen.

1- The truck arrives at load *r* before the load availability: $O_r = O_k + D(a_k, b_k) + D(b_k, a_r) < \alpha_r$ Without any adjustment, the truck must wait for $\alpha_r - O_r$. If the truck picks up load *k* earlier as the result of adjustment (i.e., α_k), it will spend more time waiting to handle load *r*. The maximum cost increment on the dwell cost is equal to $w \times [O_k - \alpha_k]$. Since *l*≥*w*, the achieved saving is greater than the maximum cost increment.

2- The truck arrives at load *r* after its earliest availability, but without violating the maximum permissible delay: $\alpha_r \leq O_r = O_k + D(a_k, b_k) + D(b_k, a_r) \leq \alpha_r + U_r$

Given the amount of lateness at loads *r* and *k* before the adjustment, there are only two possible scenarios.

2.1. The lateness at load *r* is greater than or equal to the lateness at load *k*: $[O_k - \alpha_k] \leq [\alpha_r - O_r]$. This means that the adjustment of reducing the lateness at load *r* will drop by $[O_k - \alpha_k]$.

2.2. The lateness at load *r* is less than the lateness at load *k*: $[O_k - \alpha_k] > [\alpha_r - O_r]$. In this circumstance, the entire lateness at load *r* is eliminated. However, it increases the dwell time by the difference of the lateness amounts. Based on the assumption of *l*≥*w*, this cost increment never exceeds the already obtained saving ($l \times [O_k - \alpha_k]$).

Following the same approach, it is evident that this adjustment does not increase the cost of serving subsequent loads that are scheduled after load *r*. Therefore, under no circumstance will this adjustment increase the minimum total cost.

COROLLARY 2. If *l*<*w*, replacing the dwell cost by the lateness cost and then adjustment of pickup time to α_k will never increase the minimum total cost.

PROOF. The proof is similar to Corollary 1 proof. It follows from Lemma 2 and a trivial property that replacement of the dwell cost by a smaller value will never increase the minimum total cost.

Since the adjustment of pickup times to α_k makes the problem under-constrained (because it might include some of the infeasible load combinations), we denote the minimum total cost with Z^{under}. If $l \geq w$, Z^{under} is obtained after the pickup time adjustment to α_k; otherwise we first replace w with l and then apply the pickup time adjustment. Hence, $Z^{under} \leq Z^{opt}$ according to Lemma 2 and Corollary 2.

4.4.2 Discretization Scheme of Time Windows

Generally, the larger time window results in a larger gap between Z^{under} and Z^{over}. Thus, one can use an iterative algorithm to narrow the gap by increasing time points. In the context of our proposed problem, Figure 4.1 illustrates how increasing the time points does not guarantee reducing the gap between Z^{under} and Z^{over} (and can even worsen the gap) when time windows are partitioned equally. In Figure 4.1 (a) and (b), time windows are partitioned into two and three equal intervals (three and four time points), respectively. Assume that because of a high subcontracting cost, the lowest total system cost can only be achieved by serving both loads using the available truck at city 2. If the time window of load *k* is partitioned to three points (Figure 4.1 (a)), it is still feasible to serve load *r* in the over-constrained problem. However, as seen from Figure 4.1 (b), increasing the time points (i.e., shorter but equal intervals) not only increases the lateness of load *k,* but also makes serving load *r* infeasible. Thus, increasing the number of equally spaced time points does not guarantee reducing (or maintaining) the (Z^{over} - Z^{under}) gap.

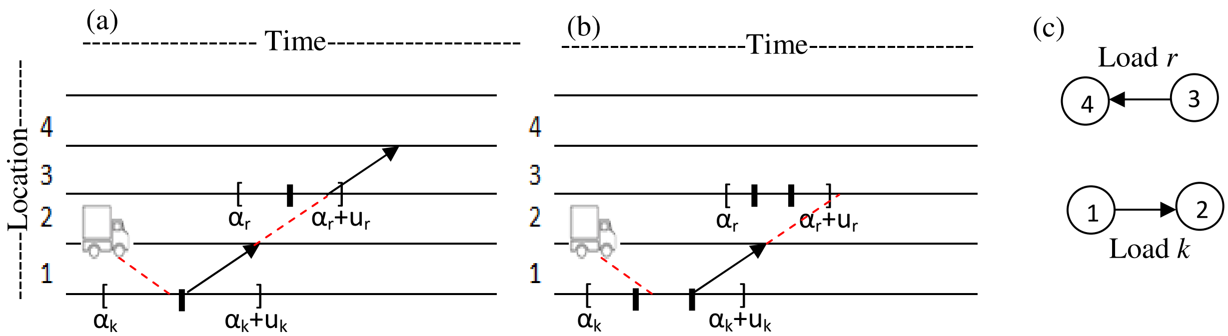


Figure 4.1. The inefficiency of partitioning time windows into equal intervals

In order to make sure that increasing the number of time points will never increase the time adjustments in under- and over- constrained problems and so reduce (or maintain) the (Z^{over} - Z^{under}) gap, the time points from previous integrations are maintained as the iterative algorithm runs. The following is our proposed iterative algorithm where $Z^{under, \Delta}$ and $Z^{over, \Delta}$ are minimum total costs for under- and over-constrained problems at iteration Δ.

Step 1. For the first iteration, Δ=1, the time window of each load is modified into two points by using the beginning and end points of the original time window (set of all time points are denoted with Ω).

Step 2. Optimize under- and over-constrained problems and compute $Z^{under,\Delta}/Z^{over,\Delta}$. If the obtained ratio is less than or equal to a specified threshold, the algorithm stops; otherwise go to step 3. Since adding time points increases the size of problem, the algorithm can be stopped if a predefined solution time is exceeded.

Step 3. Add one time point to the existing time points of each load from the last iteration and update Ω. A new time point is added in a way that the largest available time interval breaks into two parts. Return to step 2 after updating Ω.

LEMMA 3. In the proposed iterative algorithm, the ratio of $Z^{under,\Delta}/Z^{over,\Delta}$ converges to one as the number of iterations grows.

PROOF. We prove this lemma with the aid of mathematical induction.

Δ=1: if a truck reaches the pickup location of a load (e.g., load k) within its time window, the time adjustment will be $\alpha_k + u_k$ and α_k in over- and under-constrained problems, respectively. As previously noted, the time adjustment in an over-constrained problem is $[\alpha_k + u_k - O_\mathrm{k}]$ and in an under-constrained problem is $[O_\mathrm{k} - \alpha_k]$.

Δ=2: The third point, call it p3, is between α_k and $\alpha_k + u_k$. Thus, O_k is either between α_k and p3 or between p3 and $\alpha_k + u_k$. In an over-constrained problem, if $\alpha_k \leq O_\mathrm{k} \leq$p3, the time adjustment is $[\mathrm{p3} - \mathrm{O_k}]$ which is smaller than the time adjustment with Δ=1 (i.e., $[\alpha_k + u_k - O_\mathrm{k}]$). Otherwise, the time adjustments are the same in both Δ=1 and 2 (i.e., $[\alpha_k + u_k - O_\mathrm{k}]$). Hence, the time adjustments do not increase and so $Z^{\text{over},2} \leq Z^{\text{over},1}$. In an under-constrained problem, if $\alpha_k \leq O_\mathrm{k} \leq$p3, the time adjustment is $[\mathrm{O_k} - \alpha_k]$ which is the same as the time adjustment with Δ=1. Otherwise, the time adjustment is $[\mathrm{O_k} - \mathrm{p3}]$ which is smaller than the time adjustment with Δ=1 (i.e., $[O_\mathrm{k} - \alpha_k]$). Thus, the time adjustments do not increase and so $Z^{\text{under},1} \leq Z^{\text{under},2}$).

Δ=n, n+1:Using the same approach, it is simple to show that $Z^{\text{under,n}} \leq Z^{\text{under, n+1}}$ and $Z^{\text{over,n+1}} \leq Z^{\text{over ,n}}$; Thus, $\lim_{\Delta\to\infty} \frac{Z^{\text{under},\Delta}}{Z^{\text{over},\Delta}} = 1$.

In order to implement the proposed algorithm, the original mathematical model (**M1**) should be modified in an appropriate way (**M2**) for solving under- and over-constrained problems. Thus, Y_{ik} and X_{jk} are replaced by Y_{ik}^{ψ} and $X_{jk}^{\theta\psi}$, respectively. Y_{ik}^{ψ} takes the value of one if truck i serves load k at time point ψ at the first stop. Similarly, $X_{jk}^{\theta\psi}$ becomes one if load k is served at time point ψ after load j, which was served at time point θ by the same truck. The modified model is presented below.

$$\textbf{M2}: Min\ l \sum_{i\in I} \sum_{k\epsilon(\bar{J}\,\cup\,\bar{\bar{J}})} \sum_{\psi\epsilon\Omega} \mathrm{DL0}(i,k,\psi)\, Y_{ik}^{\psi} + l \sum_{j\epsilon(\bar{J}\,\cup\,\bar{\bar{J}})} \sum_{k\epsilon(\bar{J}\,\cup\,\bar{\bar{J}})} \sum_{\psi\epsilon\Omega} \sum_{\theta\epsilon\Omega} \mathrm{DL1}(j,k,\theta,\psi)\, X_{jk}^{\theta\psi}$$

$$+w\sum_{i\in I}\sum_{k\epsilon(\bar{J}\cup\bar{\bar{J}}\cup L)}\sum_{\psi\epsilon\Omega}\mathrm{DW0}(i,k,\psi)\,Y_{ik}^{\psi}+w\sum_{j\epsilon(\bar{J}\cup\bar{\bar{J}})}\sum_{k\epsilon(\bar{J}\cup\bar{\bar{J}}\cup L)}\sum_{\psi\epsilon\Omega}\sum_{\theta\epsilon\Omega}\mathrm{DW1}(j,k,\theta,\psi)\,X_{jk}^{\theta\psi}$$

$$+\sum_{k\in\bar{J}}s_k\left[1-\sum_{i\epsilon I}Z_{ik}-\sum_{i\epsilon I}\sum_{\psi\epsilon\Omega}Y_{ik}^{\psi}\right]+c\sum_{i\epsilon I,TST(i)=-1}\sum_{k\epsilon(\bar{J}\cup\bar{\bar{J}})}D(\eta_i,\mathrm{a_k})\sum_{\psi\epsilon\Omega}Y_{ik}^{\psi}$$

$$+c\sum_{i\epsilon I}\sum_{j\in\hat{J},ST(i,j)=1}\sum_{k\epsilon(\bar{J}\cup\bar{\bar{J}})}D(b_j,a_k)\sum_{\psi\epsilon\Omega}Y_{ik}^{\psi}+c\sum_{j\epsilon(\bar{J}\cup\bar{\bar{J}})}\sum_{k\epsilon(\bar{J}\cup\bar{\bar{J}})}D(\mathrm{b_j},\mathrm{a_k})\sum_{\psi\epsilon\Omega}\sum_{\theta\epsilon\Omega}X_{jk}^{\theta\psi} \tag{4.15}$$

Subject to:

$$\sum_{i\epsilon I}\sum_{\psi\epsilon\Omega}Y_{ik}^{\psi}+\sum_{i\epsilon I}Z_{ik}\leq 1,\qquad k\epsilon\bar{J} \tag{4.16}$$

$$\sum_{i\epsilon I}\sum_{\psi\epsilon\Omega}Y_{ik}^{\psi}+\sum_{i\epsilon I}Z_{ik}=1,\qquad k\epsilon\left(L\cup\bar{\bar{J}}\right) \tag{4.17}$$

$$\sum_{\psi\epsilon\Omega}Y_{ik}^{\psi}=1,\qquad i\in I,k\in\bar{\bar{J}},v(k)=i \tag{4.18}$$

$$\sum_{k\epsilon(\bar{J}\cup\bar{\bar{J}}\cup L)}\sum_{\psi\epsilon\Omega}Y_{ik}^{\psi}=1,\qquad i\in I \tag{4.19}$$

$$\sum_{k\epsilon(\bar{J}\cup\bar{\bar{J}}\cup L)}\sum_{\psi\epsilon\Omega}X_{jk}^{\theta\psi}-\left[\sum_{i\epsilon I}Y_{ij}^{\theta}+\sum_{r\epsilon(\bar{J}\cup\bar{\bar{J}})}\sum_{\psi\epsilon\Omega}X_{rj}^{\psi\theta}\right]=0,\qquad j\in\bar{J}\cup\bar{\bar{J}},\theta\epsilon\Omega \tag{4.20}$$

$$\sum_{j\in(\bar{J}\cup\bar{\bar{J}})}\sum_{\theta\epsilon\Omega}\sum_{\psi\epsilon\Omega}X_{jk}^{\theta\psi}-\sum_{i\epsilon I}Z_{ik}=0,\qquad k\in\bar{J}\cup\bar{\bar{J}}\cup L \tag{4.21}$$

$$Z_{ij}+\sum_{\psi\epsilon\Omega}Y_{ij}^{\psi}+\sum_{u\epsilon I,u\neq i}\left(Z_{uk}+\sum_{\psi\epsilon\Omega}Y_{uk}^{\psi}\right)\leq 2-\sum_{\theta\epsilon\Omega}\sum_{\psi\epsilon\Omega}X_{jk}^{\theta\psi},\qquad i\epsilon I,\ j,k\in\bar{J}\cup\bar{\bar{J}}\cup L \tag{4.22}$$

Similar to the original model (**M1**), the objective function is to minimize the total cost including lateness cost, dwell cost, subcontracting cost, and empty repositioning cost. Since time windows are converted into time points, the loads can only be handled at some specific times. Thus, the model becomes similar to the special case where no lateness is allowed. This means that all dwell and lateness times (and associated costs) can be calculated outside the mathematical model. Moreover, all time-based constraints are considered at the pre-processing stage. The constraints (4.16-4.22) are similarly defined as constraints (4.2-4.8) of the original model (**M1**). Although discretization of time windows increases the size of problem (e.g., more decision variables), the under- and over-constrained problems are integer programming problems that can be solved more efficiently by applying the pre-processing stage (numerical illustrations are presented in section 4.6).

4.5. Experimental Design, Implementation, and Analysis

4.5.1 Factor Selection and Levels

In this section, we explain how the model's parameters are generated for use in the numerical study to provide useful insights. We observe in recent statistics from the ATA (2013) that the truckload market is highly fragmented where almost 90% of the carriers are small with six or fewer trucks. The Canadian statistics are very similar to the American ones. Moreover, the smaller companies usually suffer more than big companies with sophisticated lane networks. Therefore, we concentrate on small companies with six trucks. Trucking companies may operate within different service areas; however, it is more likely for smaller trucking companies to operate locally. Therefore, our numerical experiments use a radius of 18 driving hours to fittingly portray the operating area of local operators (e.g., the low asset-based 3PL company which inspired this study).

Our investigation of the academic literature and empirical reports suggested the potential influence of the following factors on a carrier's operational costs: trip length, load density, and subcontracting cost. These factors are usually not easily controlled by trucking companies and are often dictated by the market conditions. In such circumstances, trucking companies should focus on strategies that are usually under managerial influence. This presents us with the choice of three other factors; namely advance load information, diversion capability and re-optimization interval.

Trip length: measured as travel time between a load's origin and destination. The test problems are generated in two categories called short and long trip-length groups. In the former, the majority of loads (80%) are shorter than the radius of service while in the latter the majority (80%) of loads are longer than the radius.

Load density: number of loads entering the system per truck per week. Load density is inversely related to the average length of loads (Powell, 1996), which usually ranges between 2 to 2.5 loads (per truck per week) for large companies with the average load length between two to four days. Since this study targets small carriers with shorter trip lengths, load density is studied at two levels, 2.5 (low load density) and 5 (high load density) loads per truck per week.

Subcontracting Costs: subcontracting happens when it is uneconomical or impossible to serve a load. The associated cost may have different sources. If the load is subcontracted to another carrier, it is often more costly than serving it using company-owned trucks. If the load is rejected, lost revenue and loss of goodwill are translated into cost terms (Powell 1996, Yang et al., 2004). Similar to the previous studies, we assume the subcontracting cost is explained as a linear function of a load's duration (trip length). Thus, subcontracting longer loads are more costly for the company. Two values are selected for subcontracting costs, low of $20/hr and high of $80/hr. These values were chosen after conducting preliminary experiments and observing the behavior of the model.

Advance load information (ALI): We consider the low and high ALI/KW values as, respectively 24 and 72 hours of advance notice to the carrier about loads. The low value of 24 hours is chosen because the trucking industry's excess capacity and intense competition make last-minute calls for transportation services very common. The choice of 72 hours as the high value is based on the rarity of shippers booking loads more than two or three days in advance

(Frantzeskakis and Powell, 1990). Thus, experiments on acquiring load information very far in advance (e.g., a week or so) will not provide practical managerial insights.

Diversion capability: Consistent with earlier discussion about the connection between preplanning and diversion strategies, we treat diversion as a binary factor. When the capability is on (vehicle diversion is allowed), the dispatcher can change truck-load assignments in subsequent decision epochs; when the capability is off, the pre-planning strategy is in effect; i.e., all decisions are fixed once they are made.

Re-optimization Interval (R_{opt}): this interval determines at what frequency the mathematical model should be optimized. As we have noted, this paper adopts the literature's convention of using re-optimization interval and decision interval interchangeably. Though, as demonstrated in the literature review, no published study has provided analytical or statistical justifications for its choice of a re-optimization interval, the literature still gives some insight on a suitable range of intervals to be tested. Specifically, based on Zolfagharinia and Haughton (2014) who used half a day and Tjokroamidjojo et al. (2006) who used one day (but also suggested every other day as a possible alternative in future research), we tested two intervals: every 12 hours and every 48 hours.

4.5.2 Test Problems and Dynamic Implementation

Here we explain the particulars of the test problems, the relevant experimental context, and deployment of the DPDLF model to generate the output data to be analyzed. Starting with the test problems, we covered 320 of them in our numerical experiments. The 320 resulted from testing 5 replicates for each of the resulting 2^6 factor combinations (two levels for each of the abovementioned six factors). Each replicate was a randomly generated instance of the experiments' stochastic conditions (e.g., earliest availability of loads and city locations).

In all conducted numerical experiments, it was assumed that there are 50 cities within the service area. The number of cities in a transportation network is referred to as network size. These potential cities are representative of loads' origins and destinations. This approach is commonly used in the literature. Some authors of past studies consider that the shape of the service area is square (e.g., Yang et al. 2004; Özener and Ergun, 2008) while others (e.g., Zolfagharinia and Haughton, 2014) assume a circle-shaped area. Similar to the latter studies, we consider a circle-shaped area with following parameters:

- To generate each load, an origin-destination pair is selected randomly from a 50-city network. The initial location of trucks is also determined by placing them randomly among the 50 cities.
- Following the common assumption in the literature, hourly dwelling and lateness cost are set to be \$25 per hour. The maximum lateness is drawn from a discrete uniform distribution with maximum of 5 hours.
- The earliest availability of each load is generated from an exponential distribution in which the average inter-arrival time is determined based on load density.
- The average operating highway speed is used since the majority of cities are connected to each other via highways. The average operating speed is set to 55 mph, which is typical on US highways (refer to the recent report by the U.S. department of energy, 2011).
- Fuel cost and driver wages are the major portion of the operational cost. However, there are other miscellaneous cost components such as insurance premiums and maintenance. Given that we consider dwelling and lateness cost separately, it is reasonable to set the operational cost equal to \$1.10 per mile. This operational cost is also supported by theoretical studies (e.g., Gregory and Powell, 2002) and empirical reports (e.g.,

TRANSCORE, provider of intelligent transportation systems, survey in 2011 from 600 small carriers).

- Consistent with what the research literature has established as an appropriate horizon length for truckload operations (e.g., Tjokroamidjojo et al., 2006 and Zolfagharinia and Haughton, 2014), we assume a three-week planning horizon.

We used AIMMS modeling language and Gurobi 5.5 as a solver to run the 320 test problems. The whole algorithm was programmed in MATLAB 2012b. As shown in Figure 4.2, the algorithm starts with setting ALI/KW and R_{opt}.

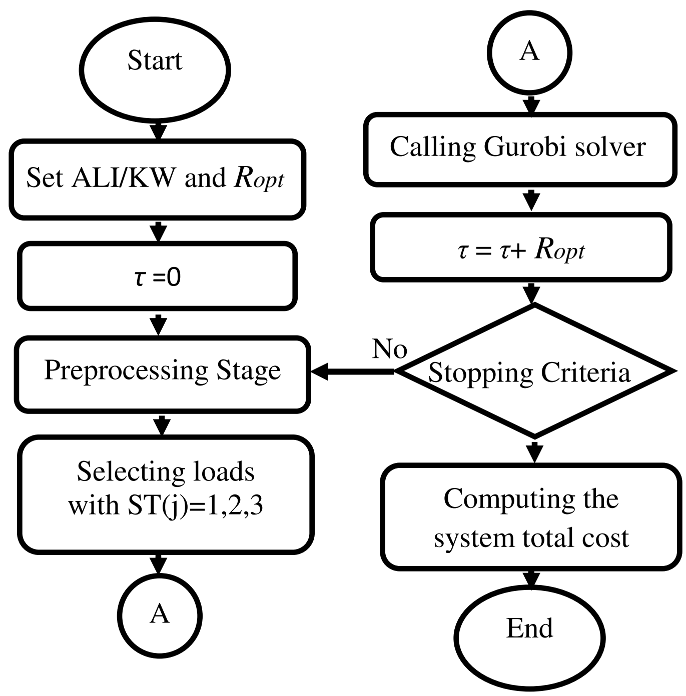


Figure 4.2. The detail of the dynamic implementation

The clock is set equal to zero and the preprocessing engine is called to update the truck and load status and exclude infeasible schedules. Then, the loads with status 1, 2, and 3 are entered into the model. In other words, the loads that have already been delivered and the ones that are far in future (i.e., beyond the knowledge window) are not included in the model. The next step is to call the solver to handle the proposed MIP model to optimality. After the model is solved, the obtained schedule is implemented up to the next interval ($\tau = \tau + R_{opt}$) and checked for the termination condition (i.e., whether all the loads are considered during the overall planning horizon). It is important to note that we have to record all the movement of trucks during the planning horizon if diversion of empty trucks is allowed. After the stopping criterion is satisfied, a simple algorithm tracks each truck's cost to compute the system total costs for the middle two weeks of the study.

4.5.3 Statistical Analysis

After solving all the test problems, we used a linear regression model (comprising the six selected factors) to statistically test how the carrier's cost is impacted by the subset of factors we already specified as being under managerial influence: ALI, diversion capability, and the re-optimization interval.

The dependent variable is the total cost during the middle two weeks of the planning horizon. The effect coding is used for categorical variables to make them appropriate for inclusion in the regression model. As the main focus of our study is to examine the impact of ALI, diversion capability and re-optimization interval, we control the impact of other factors (i.e. trip length, load density, and subcontracting cost). Thus, the control factors are first entered in the model (refer to model 1 of Table 4.2). Then, at the next step, ALI, diversion and re-optimization intervals are entered (model 2 of Table 4.2). The obtained results illustrate that slightly less than half of the variation in the total cost is explained by the control factors. Furthermore, it indicates that strategy factors (regression model 2) can explain the variations in total cost by almost 5% over and above all the control factors. Finally, the statistical output reveals the existence of interaction effects (regression model 3).

Table 4.2. The summary of the regression model

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.684[a]	.468	.463	7121.99
2	.720[b]	.519	.510	6806.90
3	.741[c]	.549	.523	6711.80

a. Predictors: (Constant), Subcontracting cost, Density, Length

b. Predictors: (Constant), All main factors including ALI, Diversion, Re-optimization

c. Predictors: (Constant), All main factors and two-way interactions of ALI, Diversion and Re-optimization with other factors

The details of statistical tests for the third model are depicted in Table 4.3. It shows that all the main factors (including control and strategy ones) except diversion are significant at the 5% level. The obtained statistical results are intuitive and easily explained for trip length, load density, subcontracting cost, and ALI. Not surprisingly, the trip length, load density, and subcontracting costs are positively correlated with the total cost. Neither diversion capability nor its two-way interactions with other control factors significantly impact the total cost. This observation is not consistent with what was found by Ichoua et al. (2006), who showed that diversion can improve system performance for vehicle routing problems. One possible explanation is the quality of advance load information in our model. Similar to previous works in the truckload literature, we assumed that the quality of information is perfect during the knowledge window of the dispatcher. For example, if the knowledge window of a dispatcher is three days, all load information during the next three days remains constant (i.e., no new loads will be realized and none of current loads will be cancelled). This assumption significantly reduces the need for truck diversion because no changes occur during the knowledge window. Although both ALI and re-optimization interval significantly impact the total cost (more ALI and shorter re-optimization intervals reduce total cost), it is more appropriate to interpret their significant interactions.

Table 4.3. The detailed statistical results of the regression model

Model 3	Standardized Coefficients Beta	t	Sig.
(Constant)		200.744	.000
Length	.219	5.673	.000
Density	.480	12.423	.000
Subcontracting Cost	.435	11.258	.000
ALI	-.210	-5.420	.000
Re-optimization (R_{OPT})	.082	2.109	**.036**
Diversion (Di)	.008	.208	.835
ALI* Length	-.017	-.450	.653
ALI* Density	-.030	-.788	.431
ALI* Subcontracting Cost	-.085	-2.203	**.028**
R_{OPT} * Length	-.002	-.058	.954
R_{OPT} *Density	-.023	-.606	.545
R_{OPT} *Subcontracting Cost	.074	1.924	**.055**
R_{OPT} *ALI	-.122	-3.156	**.002**
Di * Length	-.015	-.380	.704
Di * Density	.005	.122	.903
Di * Subcontracting Cost	-.003	-.086	.931
Di *ALI	.009	.239	.811

Starting with ALI, the extent to which more ALI reduces total cost (i.e., helps the carrier to improve its performance) depends on subcontracting costs. As Figure 4.3(a) depicts, a carrier reaps savings from advance load information as subcontracting cost rises. By the same token, ALI becomes less attractive as subcontracting cost falls. The interactions of re-optimization intervals with subcontracting cost and ALI have remarkable impacts on the carrier's cost. Figure 4.3(b) and (c) helps to explain these impacts. Seen from Figure 4.3(b), the re-optimization interval does not make a considerable difference when subcontracting cost is low. However, the impact becomes significant when subcontracting cost is large. This result can be explained as follows. More frequent re-optimization means greater operational responsiveness by delivering more shipments with company-owned trucks instead of incurring the penalty cost of subcontracting those shipments to other carriers. Thus, the higher that penalty, the larger will be the carrier's costs reductions from being responsive.

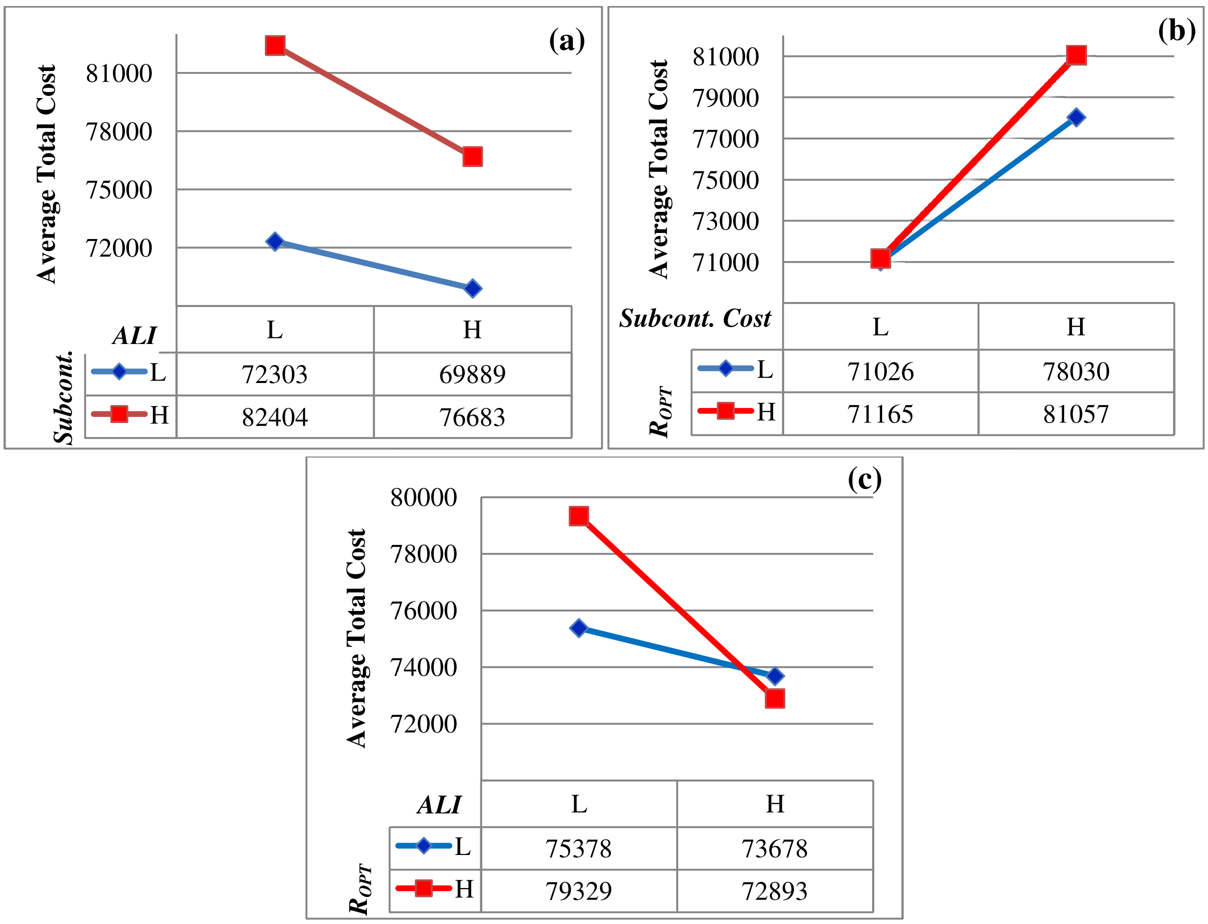


Figure 4.3. Significant two-way interaction effects

Figure 4.3(c) depicts that the impact of the re-optimization interval is highly dependent on how much advance load information is available. The longer re-optimization interval is equivalent to a greater level of postponement in decision making. Based on the assumptions, once a decision is made at a decision epoch, no modification is possible regarding whether to serve load(s) by using the company's trucks or subcontracting. Therefore, on the one hand, postponement can provide the dispatcher with more information before decision making. On the other hand, it reduces the company's responsiveness in using its own trucks. When limited advance load information is available (e.g. one day), the company's responsiveness in using its own trucks becomes essential.

Thus, the benefit (in terms of cost reduction) from responsiveness outweighs the possible benefit of longer intervals (i.e. having additional information). In short, the suggestion is to shorten the intervals when access to ALI is limited. The situation alters when more advance load information is available (e.g., three days). In such circumstances, the importance of company responsiveness in using its own truck reduces since greater gains will come from accessing additional ALI. In this case, a longer re-optimization interval leads to lower cost.

4.6. Comparison with the Benchmark Solutions

In this section, we pursue three important goals. First, the efficiency of the proposed algorithm (in terms of solution quality and run-time) is examined under different network settings. Second, with different levels of advance load information, the various carrier policies are compared against each other according to their deviations from the proposed benchmark. Although the benchmark solution is obtained when all load information is available in advance (and so results in unrealistically low total cost), it is still a fair illustration of how much further the obtained results might be improved. Finally, to further illustrate the performance of the proposed benchmark, we numerically investigate a case where lateness cost is not equal to dwell cost.

Table 4.4 illustrates the efficiency of the proposed algorithm in two iterations. The first column represents each individual combination of controlled factors (i.e. trip length, load density, and subcontracting cost). As mentioned earlier, each combination was replicated five times and the key information about $Z^{under,\Delta}$, run-time (in seconds), and $Z^{under,\Delta}/Z^{over,\Delta}$ (ratio) are collected for each iteration. As seen from this table, the lowest ratio is 0.9866 and the optimal solutions were found for more than half of the replicates in the first iteration. Moreover, the second iteration improves the lowest ratio to 0.9967. The level of load density has direct impact on the problem size and the run time. It takes on average 4 seconds to compute $Z^{under,1}$ where load density is low (i.e., the average number of loads is 51 per test problem) while it takes almost

11 seconds in the case of high load density (i.e., the average number of loads is 96). It is evident that the proposed algorithm produces a very high quality ratio in a matter of seconds.

Table 4.4. The performance of the proposed algorithm

Code	Iteration Δ=1			Iteration Δ=2		
	$Z^{under,1}$	CPU (Sec)	$[Z^{under,1}/Z^{over,1}]$	$Z^{under,2}$	CPU (Sec)	$[Z^{under,2}/Z^{over,2}]$
LLL-1	69301	3.15	1	----	----	----
LLL-2	69466	4.51	1	----	----	----
LLL-3	70369	3.4	0.9995	70369	5.21	0.9998
LLL-4	68231	3.71	1	----	----	----
LLL-5	71584	7.24	1	----	----	----
LLH-1	70918	4.48	1	----	----	----
LLH-2	70437	3.25	1	----	----	----
LLH-3	71293	3.95	1	----	----	----
LLH-4	68931	3.87	1	----	----	----
LLH-5	74303	4.12	1	----	----	----
LHL-1	65001	7.46	0.9987	65019	12.21	0.9999
LHL-2	65900	4.57	0.9997	65917	8.7	1
LHL-3	63816	12.31	1	----	----	----
LHL-4	65959	8.46	0.9989	65972	16.21	0.9998
LHL-5	68068	12.12	0.9979	68118	13.12	0.9996
LHH-1	73298	17.82	0.9982	73317	18.01	0.9994
LHH-2	73555	11.68	0.9952	73707	13.73	0.9999
LHH-3	67898	20.22	0.9995	67898	24.26	1
LHH-4	72861	9.14	0.9986	72892	12.43	0.9997
LHH-5	75862	11.45	0.9966	75949	13.56	0.9997
HLL-1	63459	3.1	1	----	----	----
HLL-2	64318	4.84	1	----	----	----
HLL-3	62681	4.37	1	----	----	----
HLL-4	65480	3.89	1	----	----	----
HLL-5	65360	3.51	1	----	----	----
HLH-1	65444	3.62	1	----	----	----
HLH-2	67207	5.1	1	----	----	----
HLH-3	67301	3.29	1	----	----	----
HLH-4	66869	3.32	1	----	----	----
HLH-5	68784	3.56	1	----	----	----
HHL-1	65245	8.09	0.9986	65284	11.87	0.9996
HHL-2	62241	14.12	1	----	----	----
HHL-3	63176	11.5	0.9993	63213	15.32	0.9999
HHL-4	67890	9.38	0.9937	68200	13.59	0.9993
HHL-5	63915	11.03	0.9976	63997	13.42	0.9996
HHH-1	84328	8.86	0.9981	84391	10.31	0.9990
HHH-2	85638	14.65	0.9996	85641	15.43	0.9998
HHH-3	87634	8.61	0.9973	87789	13.96	0.9994
HHH-4	94932	9.41	0.9866	95894	10.81	0.9967
HHH-5	89239	11.15	0.9867	89385	18.98	0.9994

Note: Code=Trip Length-Load Density-Subcontracting Cost

To have a clearer understanding of the efficiency of the proposed algorithm, after the preprocessing stage, Gurobi 5.5 was used to solve the mathematical model (**M1)** where all information is available in advance. Since it takes a relatively short time (less than 10 minutes) to find a value for a LP[4] lower bound solution and a feasible solution across the board, we run the solver for a much longer period (e.g., three hours) to get a higher quality solution. The best LP lower bound solution and the best feasible solution are analogous to under- and over-constrained solutions in that they both provide a range that includes the optimal total cost. It is evident that a narrower range indicates a higher quality solution. This helps us to provide stronger evidence on the efficiency of the proposed algorithm compared to solving the original mathematical model (**M1**).

Figure 4.4 illustrates two average ratios (named performance ratios) in a radar chart: 1) the average ratio of under-constrained to over-constrained solutions in two iterations; 2) the average ratio of the best LP lower bound solution to the best feasible solution in a three-hour run time for the M1 model. The performance ratio (ranges between 0 and 1) indicates the solution quality of the solution method. As also seen from Table 4.4, within a few seconds, the proposed algorithm is able to find the optimal solution for more than half of test replicates and a very competitive ratio for the rest. However, not surprisingly, the obtained results from solving the **M1** model indicate a poor performance. Although the average percentage is higher under the HHH setting (generally converges faster when the factors take higher values), it is still not acceptable because the best LP lower bound solution deviates remarkably from the best feasible solution (e.g., the performance ratio can be as low as 0.4). Moreover, it is not efficient at all from the computational aspect and so solving model **M1** is not considered as a quality benchmark.

---

[4] Linear Programming

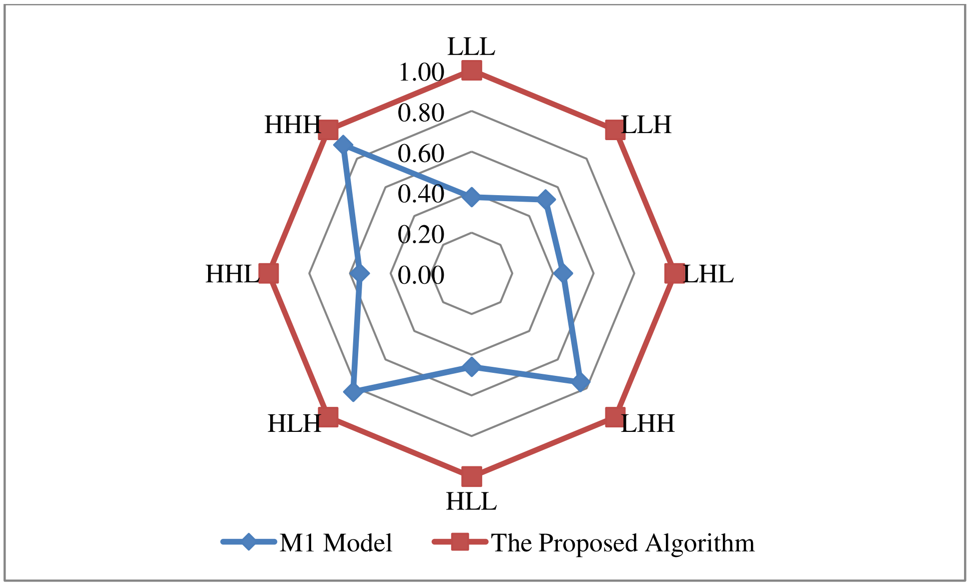


Figure 4.4. The efficiency of the proposed algorithm based on performance ratios

Now, we can compare different policies against each other based on their deviation from the quality proposed lower bound. These policies are identified with the aid of the statistical analysis in Section 5. Both ALI and re-optimization interval have significant impact on the total cost while diversion capability does not. Thus, one can define four policies considering different levels of advance load information and the re-optimization interval.

To make the policies comparable under different settings, we focus on their deviation from the proposed lower bound. If the optimal solution is found at the first iteration (i.e., $Z^{under,1}/Z^{over,1}=1$), this solution will be used as a benchmark; otherwise $Z^{under,2}$ will be chosen (the fifth column of Table 4.4). This approach does not impact our general conclusion since the ratio ($Z^{under,\Delta}/Z^{over,\Delta}$) is very competitive (i.e., close to one). Table 4.5 illustrates the deviation of four policies from the proposed benchmark. Applying ANOVA and Student Newman Keuls (SNK) tests reveal a significant difference between the average deviations of policies at 5% level. When ALI is limited to one day, the choice of an appropriate re-optimization interval (R_{opt}=12hrs) significantly impacts the total cost. This impact is more remarkable when subcontracting is high. The deviation drops by 5.74% on average when the model is re-optimized more frequently.

Table 4.5. The average deviation of each policy (in percentage) from the benchmark solution

Code	% of deviation from the benchmark solution			
	ALI=24hrs, R_{opt}=24hrs	ALI=24hrs, R_{opt}=12hrs	ALI=72hrs, R_{opt}=24hrs	ALI=72hrs, R_{opt}=12hrs
LLL	7.22%	6.03%	4.03%	5.49%
LLH	16.45%	3.68%	2.97%	2.86%
LHL	19.34%	16.93%	10.59%	13.95%
LHH	18.36%	10.50%	6.75%	7.45%
HLL	15.21%	10.79%	9.96%	11.41%
HLH	20.08%	11.07%	5.46%	6.34%
HHL	22.22%	20.11%	14.51%	17.60%
HHH	20.66%	14.45%	9.68%	9.80%
Average	**17.44%**	**11.70%**	**7.99%**	**9.36%**

Although the choice of the appropriate re-optimization interval is helpful to reduce total cost, additional advance load information can result in more savings. With three-day ALI, there is a reduction in the average deviation from the benchmark solution regardless of the re-optimization interval. It is interesting to point out that having access to additional load information and selecting an appropriate re-optimization interval can reduce the deviation from the benchmark solution to less than 10% in most combinations.

Although there is an average improvement by acquiring additional information regardless of the re-optimization interval, it is not true across all individual combinations. In the HLL setting, not only does additional load information not improve the deviation, but it also worsens it if the appropriate re-optimization interval is not chosen. This reinforces the importance of the re-optimization interval selection in relation to advance load information. The obtained results also provide an important managerial insight for carriers with limited ALI. By using an appropriate re-optimization interval, such carriers limit their loss resulting from insufficient ALI to an average of no more than 4 percentage points (the deviation increases to 11.70% (with ALI=24hrs, R_{opt}=12hrs) from 7.99% (with ALI=72hrs, R_{opt}=24hrs).

To further illustrate the performance of the suggested algorithm, we conducted the numerical experiments for a case that dwell cost is greater than lateness cost ($w>l$). Even though a very large value can be considered for the theoretical ratio of dwelling cost to lateness cost, a more practical value was selected (e.g., dwell cost is double lateness cost: *w*=$50/hr, *l*=$25/hr).

The performance of the algorithm with three iterations is illustrated in Table 4.6.

Table 4.6. The performance of the proposed algorithm when dwell cost is double lateness cost

Code	Iteration Δ=1		Iteration Δ=2		Iteration Δ=3	
	$Z^{under,1}$	$[Z^{under,1}/Z^{over,1}]$	$Z^{under,2}$	$[Z^{under,2}/Z^{over,2}]$	$Z^{under,3}$	$[Z^{under,3}/Z^{over,3}]$
LLL-1	66301	**1**	-	-	-	-
LLL-2	66665	0.9967	66759	0.9991	66787	0.9996
LLL-3	67521	0.9939	67664	0.9981	67664	**1**
LLL-4	65209	0.9989	65243	0.9997	65243	0.9999
LLL-5	68683	**1**	-	-	-	-
LLH-1	67897	0.9994	67916	0.9997	67916	0.9999
LLH-2	67613	0.9956	67719	0.9989	67754	0.9995
LLH-3	68821	0.9908	68965	0.9975	69056	**1**
LLH-4	65879	0.9977	65935	0.9994	65935	0.9997
LLH-5	71337	0.9983	71399	0.9996	71399	0.9998
LHL-1	62313	0.9877	62725	0.9968	62741	0.9984
LHL-2	63713	0.9833	64110	0.9909	64124	0.9993
LHL-3	61787	0.9976	61893	0.9994	61902	0.9997
LHL-4	63808	0.9895	63965	0.9967	64031	0.9986
LHL-5	65787	0.9892	66011	0.9976	66076	0.9984
LHH-1	70475	0.9837	71021	0.9956	71074	0.9973
LHH-2	71617	0.9840	72113	0.9974	72147	0.9987
LHH-3	65882	0.9939	66028	0.9985	66069	0.9992
LHH-4	70780	0.9862	71034	0.9970	71090	0.9985
LHH-5	73543	0.9874	73866	0.9971	73919	0.9986
HLL-1	60518	0.9996	60525	0.9999	60528	**1**
HLL-2	63487	**1**	-	-	-	-
HLL-3	60466	0.9949	60673	0.9991	60701	0.9995
HLL-4	62426	0.9981	62463	0.9993	62470	0.9996
HLL-5	62395	0.9988	62426	0.9995	62435	0.9997
HLH-1	62624	**1**	-	-	-	-
HLH-2	66376	**1**	-	-	-	-
HLH-3	65012	0.9936	65256	0.9983	65296	0.9991
HLH-4	63853	0.9978	63853	0.9992	63860	0.9995
HLH-5	65682	0.9947	65826	0.9986	65851	0.9993
HHL-1	63718	0.9923	64005	0.9981	64020	0.9984
HHL-2	59794	0.9917	59944	0.9977	59975	0.9982
HHL-3	62344	0.9974	62488	0.9999	62488	**1**
HHL-4	65203	0.9832	65812	0.9961	65872	0.9982
HHL-5	61897	0.9851	62281	0.9961	62357	0.9982
HHH-1	83252	0.9966	83252	0.9980	83290	0.9982
HHH-2	83389	0.9974	83389	0.9989	83399	0.9991
HHH-3	86789	0.9922	87185	0.9986	87226	0.9993
HHH-4	92645	0.9850	93491	0.9969	93566	0.9985
HHH-5	87958	0.9863	87958	0.9966	88028	0.9978

Note: Code=Trip Length-Load Density-Subcontracting Cost

Compared to the results with original parameters (Table 4.4), the lowest ratio (0.9832) is marginally lower than 0.9866, the lowest ratio found earlier at the first iteration. However, additional iterations (i.e. third iteration) can improve the lowest ratio to 0.9973. This provides strong evidence that the algorithm is highly efficient in producing a quality benchmark even if dwell cost is larger than lateness cost.

4.7. Conclusion and Future Research Directions

In the trucking industry, many carriers face highly variable demands from clients as well as other challenging issues reported by the American Trucking Associations (ATA, 2014). Despite several attempts in the literature, the need for operational improvements by incorporating simple and effective policies is still felt.

Inspired by a real life case, we model a small trucking company located in Ontario, Canada. The main approach is to design a mathematical model for the static version of the problem and apply it in the dynamic context using a rolling horizon approach. The computational efficiency of the proposed mathematical model is improved by adding a feature named the pre-processing stage. This feature serves the following roles. First, it is possible to reduce the dimensionality of the problem by eliminating some infeasible solutions. Second, it helps us develop an efficient formulation for a special case of the problem (where no lateness is allowed) by handling all the time-based constraints outside the mathematical model.

One of the major contributions of this study is that it develops an algorithm based on discretization of time windows. This method allows us to convert any problem with the lateness option to the special case of the problem. That is why the algorithm is computationally efficient and can easily handle medium sized problems with almost 100 loads in a matter of seconds. Moreover, we proved that this algorithm converges to the optimal total cost. The numerical analysis shows that the proposed algorithm converges very quickly under various network settings and even different parameters.

Another contribution of this study is to provide valuable insights for carriers on how they can improve their operational efficiency. Through comprehensive numerical experiments and statistical analysis, we found that ALI and the re-optimization interval significantly influences the total cost. However, diversion capability and its interactions with other factors are not statistically significant. The findings also show that the impact of the re-optimization interval depends on the subcontracting cost and level of advance load information.

Finally, given the values we considered for the ALI and the re-optimization interval, four policies are investigated and compared against each other according to their deviations from the benchmark solution. The obtained results emphasize the importance of the re-optimization interval when ALI is limited to one day. Moreover, by choosing the appropriate re-optimization interval, carriers do not lose more than an average of 4 percentage points in deviation from the benchmark compared to accessing three-day load information. Finally, it was observed that three-day load information and the appropriate re-optimization interval can reduce the deviation from the benchmark to less than 10% in most combination settings.

This research study can be extended in various directions. There are several real-life situations in which the quality of information is uncertain during the knowledge window of the dispatcher (e.g., the possibility of load cancelation). There are also some circumstances that some clients are not willing to communicate their load information in advance, e.g., military clients because of security issues (existence of partial load information). Thus, it is interesting to model and investigate some of the flexibility features (e.g., diversion capability) under the new assumption.

Another interesting research direction is to investigate the possible benefit of diversion capability, re-optimization interval, and ALI where historical information is available and reliable. This would provide clarification regarding how historical information might impact the significance of those factors. We also consider a fixed travel time as a function of distance. However, there are various factors such as road congestion, weather conditions, or accidents that might impact travel times. Therefore, it is insightful to see the impact of those complexities in deriving managerial insights. We targeted small trucking companies that constitute the majority of carriers in North America. Designing an efficient solution algorithm to handle large trucking companies with a few hundred trucks and drivers is an important step to seek answers for similar research questions, but in another context.

CHAPTER 5

CONCLUSION: INSIGHTS AND LOOKING AHEAD

Past research studies on long haul transportation dispatching rules do not incorporate the concept of a home domicile. This is quite important because of human related considerations and maintaining the trucks on a regular basis. Lack of this feature will result in overestimating the capacity of a transportation network therefore reducing the model accuracy. In addition, most studies assume that all loads information is available in advance. Thus, many of these models are not suitable to be implemented in a dynamic context. A major contribution of the second chapter is that it develops a comprehensive MIP model that is flexible enough to include many operational details. The use of a rolling horizon approach allows it to be implemented in a dynamic environment. The other major contribution of the second chapter is gauging the benefit of advanced load information in the truckload industry. In this regard, a comprehensive set of numerical experiments covering five factors is designed. The results of the study illustrate that access to the second day loads information can improve profit by an average of 22%. Obtaining more information can further increase the benefit, however the margin decreases to 6%. Moreover, other transportation network settings have the ability to affect the overall impact of ALI. For example, the impact of ALI on a carrier's profit is greater when the majority of carrier's loads are long or the carrier is operating within a large service area. The rejection rate can be also reduced by accessing loads information further in advance by obtaining the second-day loads information. The reduction in rejection rate becomes trivial by moving beyond the second day load information. It is also important to note that the improvement in the rejection rate depends on the radius of service and load density. The benefit (in terms of lowering the rejection rate) becomes larger if the radius of service grows. The rejection rate also improves to a greater extent when the load density is lower.

The third chapter extends the previous study by addressing the uncertainty after the knowledge window. In this work, the focus is on developing novel policies to help trucking companies to improve their razor-thin profit. The contributions of this chapter are threefold. First, the dynamic pickup and delivery problem is reformulated as a two-index mixed integer program. This formulation is more efficient compared to three-index formulations in literature (e.g., Keskinocak and Tayur,1998; Tjokroamidjojo et al., 2006; Zolfagharinia and Haughton, 2014). The main reason for reformulation is to solve the static version of the problem to optimality during the entire planning horizon. Second, we design a novel dispatching policy that generates quality solutions under one of the most practical transportation network settings. With only two day ALI, the proposed policy produces almost 90% of the attainable profit during the planning horizon. Finally, the proposed policy is enhanced by incorporating the scenario generation approach. The finding shows that scenario generation can significantly improve the performance of the policy when the ALI is limited to one day. The scenario generation benefits decline when additional load information becomes available to the dispatcher. We also compared our developed polices with two other dispatching methods. The first method is rooted in practice and designed by consulting our industry partner. The second method is purely based on the scenario generation approach. The result shows the performance of the proposed policies both in terms of solution quality and computational efficiency.

Similar to the previous works, the last study is within the context of truckload transportation. However, it differs in two aspects: 1) it targets local operators; 2) all load requests are being handled through either the company owned trucks or other carriers (subcontractors). The main inspiration of this study is a small third party logistics provider (Logikor Inc.) located in Ontario, Canada. The major goal of this study is to identify effective strategies that reduce the total operational costs. To achieve this goal, the current study extends the existing literature in the following ways. First, it develops a mathematical model that can capture all important cost components during the knowledge window of the dispatcher by introducing dummy loads.

Second, it proposes an efficient method based on time window discretization to solve the static version of problem during the planning horizon. Third, it investigates the impact of ALI, re-optimization interval, and diversion capability.

The statistical analyses reveal that ALI and re-optimization interval have significant impact on the total cost, but that diversion capability does not. The result on the impact of diversion capability is not consistent with some studies in the context of vehicle routing problems (e.g., Ichoua et al., 2006). One possible explanation is the quality of ALI in the problem under investigation. Similar to previous works in the truckload literature, we assumed that the quality of information is perfect during the knowledge window of the dispatcher (i.e., no new loads will be realized and none of current loads will be cancelled during the KW). Our industry partner also experiences a negligible load cancelation rate. This assumption significantly reduces the need for truck diversion because no changes occur during the knowledge window. Finally, we introduce different policies based on combinations of significant strategies (i.e., ALI and re-optimization interval). The obtained results illustrate that selecting an appropriate re-optimization interval is essential when ALI is limited to one day.

The studies in this thesis can be expanded in various directions. Some of these interesting research areas are listed below:

- In chapters 2 and 3, it is assumed that when a truck returns to the home domicile, it is immediately ready for the next trip. However, this is true only if a backup driver is available to take the responsibility of the incoming truck and the truck does not require major maintenance. Thus, the impact of relaxing that assumption is worth investigating.
- In chapters 2 and 3, it was assumed that the home base of the carrier is located in the center of the service area. This cannot always be the case. Thus, it is interesting to

know how an appropriate selection of the depot within the service area can improve the operational efficiency of carriers. We anticipate observing stronger impact when high geographical imbalance exists.

- One of the strong assumptions in all previous chapters is the quality of load information during the knowledge window. Although we assume the quality of load information is perfect, many real world cases experience a high level of uncertainty. Thus, another possible research direction is to address information uncertainty as loads information (e.g., pick-up time or cancellation) may change even after it is received by the carrier.
- The travel time was assumed to be a linear function of distance. In reality, the constant travel time can be viewed a restrictive assumption. Since various factors (e.g., weather condition, road accidents, and truck breakdowns) can influence travel time, it is interesting to test the robustness of different policies where travel times are not constant.
- Using a mathematical model is an appropriate choice where small trucking companies are targeted. However, designing an efficient algorithm to handle large trucking companies with a few hundred trucks and drivers is an important step to seek answers for similar research questions.
- Another fruitful research agenda is to evaluate the benefits of collaboration through information sharing in intermodal transportation. This would represent an intriguing transition from the unimodal focus of this dissertation.

APPENDIX A.

When no lateness is allowed, there would be no non-linear terms to be linearized. Moreover, all the time constraints (3.24-3.30) can be handled outside the mathematical model. Thus, we only need to make four changes in the original model. First, omitting the lateness term from the objective function since no lateness is allowed. Second, in the objective function, the dwell cost term in (3.10) is replaced by term (A.1). As explained in section 3.4.3.2, the dwell cost will not apply for the loads that are scheduled right after the depot. The final changes are conducted in the body of constraints by replacing constraints (3.21) and (3.23) with (A.2) and (A.3), respectively.

$$w \sum_{j \epsilon \bar{J} \cup \bar{\bar{J}}} \left[\sum_{i \in I} \mathrm{DW1}(i,j) Y_{ij}^{0} + \sum_{r \epsilon \bar{J} \cup \bar{\bar{J}}} \mathrm{DW2}(r,j)\, X_{rj}^{0} \right] \tag{A.1}$$

$$\sum_{i \epsilon I} N_{ik} \leq \left[\sum_{i \epsilon I} N_{ij} - \sum_{i \epsilon I} \mathrm{DW1}(i,j) Y_{ij}^{0} - \sum_{r \in \bar{J} \cup \bar{\bar{J}}} \mathrm{DW2}(r,j) X_{rj}^{0} - D(a_j, b_j) - D(b_j, a_k) \right] + \left(1 - X_{jk}^{0}\right) H, \quad j,k \in \bar{J} \cup \bar{\bar{J}}, \tag{A.2}$$

$$\sum_{i \epsilon I} N_{ik} \geq \sum_{i \epsilon I} \mathrm{DW1}(i,j) Y_{ij}^{0} + \sum_{j \in \bar{J} \cup \bar{\bar{J}}} \mathrm{DW2}(j,k) X_{jk}^{0} + D(a_k, b_k) + D(b_k, h_0) - \left(1 - \sum_{i \epsilon I} \left[Z_{ik} + Y_{ik}^{0} + Y_{ik}^{1} \right] \right) H, \quad k \in \bar{J} \cup \bar{\bar{J}}, \tag{A.3}$$

REFERENCES

American Trucking Associations. (2011). Radial Tire Conditions Analysis Guide. Fourth Edition.

American Trucking Association (2013) Retrieved from: http://www.logisticsmgmt.com/article/ata_releases_american_trucking_trends_2013

American Trucking Association (2014) Retrieved from: www.intrucking.org/.../atri_2014_top_industry_issues_report_final.pdf

American Transportation Research Institute. (2014). Retrieved from: http://atri-online.org/wp-content/uploads/2014/09/ATRI-Operational-Costs-of-Trucking-2014-FINAL.pdf

Angelelli, E., Bianchessi, N., Mansini, R., & Speranza, M.G. (2009). Short Term Strategies for a Dynamic Multi-Period Routing Problem. Transportation Research Part C: Emerging Technologies, 17(2), 106-119.

Anupindi, R., Morton, T.E., & Pentico, D. (1996). The Non-stationary Stochastic Lead-Time Inventory Problem: Near-Myopic Bounds, Heuristics, and Testing. Management Science, 42(1), 124-129.

Barla, P., Bolduc, D., Boucher, N., & Watters, J. (2010). Information technology and efficiency in trucking. Canadian Journal of Economics. 43 (1) 254-279.

Bent, R.W., & Van Hentenryck, P. (2004). Scenario-Based Planning for Partially Dynamic Vehicle Routing with Stochastic Customers. Operations Research, 52(6): 977-987.

Berbeglia, G., Cordeau, J-F., & Laporte, G. (2010). Dynamic Pickup and Delivery Problems. European Journal of Operational Research, 202(1), 8-15.

Berger, S., & Bierwirth, C. (2010). Solutions to the request reassignment problem in collaborative carrier networks. Transportation Research Part E: Logistics and Transportation Review, 46(5): 627-38.

Bookbinder, J.H. H'ng, B.T. (1986). Production Lot Sizing for Deterministic Rolling Schedules. Journal of Operations Management, 6(3-4), 349-362.

Bourland, K.E., Powell, S.G., Pyke, D.F., 1996. Exploiting Timely Demand Information to Reduce Inventories. European Journal of Operational Research, 92(2), 239-253.

Branchini M. R., Amaral Armentano V., & Løkketangen A. 2009. Adaptive Granular Local Search Heuristic for a Dynamic Vehicle Routing Problem. Computers & Operations Research 36(11): 2955-2968.

Caplice, C., & Sheffi. Y., (2003). Optimization Based Procurement for Transportation Services.

Journal of Business Logistics, 24(3): 109–128.

Caplice, C. (2007). Electronic Markets for Truckload Transportation. Production and Operations Management, 16(4), 423-436.

Chand, S., Hsu, V.N., & Sethi, S. (2002). Forecast, Solution, and Rolling Horizons in Operations Management Problems: A Classified Bibliography. Manufacturing & Service Operations Management, 4(1), 25-43.

Chang, T.S. (2009). Decision Support for Truckload Carriers in One-Shot Combinatorial Auctions. Transportation Research Part B: Methodological, 43(5): 522-541.

Cheevaprawatdomrong, T., & Smith, R.L. (2004). Infinite Horizon Production Scheduling in Time-Varying Systems under Stochastic Demand. Operations Research, 52(1), 105-115.

Choong, S.T., Cole, Michael H., & Kutanoglu, E. (2002). Empty Container Management for Intermodal Transportation Networks. Transportation Research Part E: Logistics and Transportation Review, 38(6): 423-438.

Council of Supply Chain Management Professionals, (2006). Posted on http://cscmp.org/.

Crainic, T.G. (2000). Service Network Design in Freight Transportation. European Journal of Operational Research, 122(2): 272-288.

Ergun, Ö., Kuyzu, G., & Savelsbergh, M. (2007a). Shipper Collaboration. Computers and Operations Research, 34(6): 1551-1560.

Ergun, Ö., Kuyzu, G., & Savelsbergh, M. (2007b). Reducing Truckload Transportation Costs Through Collaboration. Transportation Science, 41(2): 206-221.

Ferrucci, F. & Bock, S. (2015). A General Approach for Controlling Vehicle en-route Diversions in Dynamic Vehicle Routing Problems. Transportation Research Part B: Methodological 77(0): 76-87.

Frantzeskakis, L., & Powell, W.B. (1990). A Successive Linear Approximation Procedure for Stochastic, Dynamic Vehicle Allocation Problems. Transportation Science, 24(1): 40-57.

Frohlich, M., & Westbrook, R. (2001). Arcs of Integration: An International Study of Supply Chain Strategies. Journal of Operations Management, 19(2), 185-200.

Gavirneni, S., Kapuscinski, R., & Tayur, S. (1999). Value of Information in Capacitated Supply Chains. Management Science, 45(1), 16-24.

Gendreau, M., Guertin, F., Potvin, J.Y., & Taillard, E. (1999). Parallel Tabu Search for Real-Time Vehicle Routing and Dispatching. Transportation Science, 33(4): 381-390.

Godfrey, G.A., & Powell, W.B. (2002). An Adaptive Dynamic Programming Algorithm for Dynamic Fleet Management, I: Single period travel times. Transportation Science, 36(1), 21-39.

Gronalt, M., Hartl, R.F., & Reimann, M. (2003). New Savings Based Algorithms for Time Constrained Pickup and Delivery of Full Truckloads. European Journal of Operational Research, 151(3):520-35.

Helper, C. M., Davis, L. B., & Wei, W., (2010). Impact of Demand Correlation and Information Sharing in a Capacity Constrained Supply Chain with Multiple-Retailers. Computers and Industrial Engineering, 59(4), 552-560.

Hvattum, L. M., Løkketangen, A., & Laporte G. (2006). Solving a Dynamic and Stochastic Vehicle Routing Problem with a Sample Scenario Hedging Heuristic. Transportation Science, 40(4):421-38.

Ichoua, S., Gendreau, M., & Potvin, J-Y, (2006). Exploiting Knowledge about Future Demands for Real-Time Vehicle Dispatching. Transportation Science, 40(2): 211-225.

Jaillet, P., & Wagner, M.R., (2006). Online Routing Problems: Value of Advanced Information as Improved Competitive Ratios. Transportation Science, 40(2), 200-210.

Keller, S.B. (2002). Driver Relationships with Customers and Driver Turnover: Key Mediate Variables Affecting Driver Performance in the Field. Journal of Business Logistics, 23(1): 39-64.

Kern, G.M., & Wei, J.C. (1996). Master Production Rescheduling Policy in Capacity-Constrained Just-In-Time Make-To-Stock Environments. Decision Sciences, 27(2), 365-387.

Keskinocak, P., & Tayur, S. (1998). Scheduling of Time-shared Jet Aircraft. Transportation Science, 32(3): 277-294.

Klundert, J.V.D. & Wormer L. (2010). ASAP: The After-Salesman Problem. Manufacturing & Service Operations Management. 12(4): 627-641.

Lewis, I., & Talalayevsky, A. (1997). Logistics and Information Technology: a Coordination Perspective. Journal of Business Logistics, 18(1), 141-157.

Liu, R., Jiang, Z., Liu, X., & Chen, F., (2010a). Task Selection and Routing Problems in Collaborative Truckload Transportation. Transportation Research Part E: Logistics and Transportation Review, 46(6): 1071-1085.

Liu, R., Jiang, Z., Fung, R.Y.K., Chen, F., & Liu, X., (2010b). Two-phase Heuristic Algorithms for Full Truckloads Multi-Depot Capacitated Vehicle Routing Problem in Carrier Collaboration. Computers and Operations Research, 37(5): 950-959.

Lynch, K. (2001). Collaborative Logistics Networks-Breaking Traditional Performance Barriers for Shippers and Carriers. White paper, Nistevo. http://www.nistevo.com/v1/downloads/index.html.

Mitrović-Minić, S., Krishnamurti, R., & Laporte, G. (2004). Double-Horizon Based Heuristics for the Dynamic Pickup and Delivery Problem with Time Windows. Transportation Research Part B: Methodological, 38(8), 669-685.

Özener, Ö.Ö., & Ergun, Ö. (2008). Allocating Costs in a Collaborative Transportation Procurement Network. Transportation Science 42(2), 146-165.

Özener, Ö.Ö., Ergun, Ö., & Savelsbergh, M. (2011). Lane-exchange Mechanisms for Truckload Carrier Collaboration. Transportation Science, 45(1): 1-17.

Patterson, K.A., Grimm, C.M., & Corsi, T.M. (2003). Adopting New Technologies for Supply Chain Management. Transportation Research Part E-Logistics and Transportation Review, 39(2), 95-121.

Pillac, V., Gendreau, M., Guéret, C., & Medaglia, A.L. (2013). A Review of Dynamic Vehicle Routing Problems. European Journal of Operational Research, 225(1), 1-11.

Powell W.B. (1986). A Stochastic Formulation of the Dynamic Vehicle Allocation Problem, Transportation science, 20(2): 117-129.

Powell, W.B., (1987). An Operational Planning Model for the Dynamic Vehicle Allocation Problem with Uncertain Demands. Transportation Research Part B: Methodological, 21(3): 217-232.

Powell, W.B. (1996). A Stochastic Formulation of the Dynamic Assignment Problem, with an Application to Truckload Motor Carriers. Transportation science, 30(3): 195-219.

Powell, W.B., Bouzaïene-Ayari, B., & Simão, H.P., (2007). Chapter 5, Dynamic Models for Freight Transportation. Barnhart, C., Laporte, G. (Eds.), Handbooks in Operations Research and Management Science, Elsevier. 14: 285-365.

Powell, W.B., Sheffi, Y., Nickerson, K.S., Butterbaugh, K., & Atherton, S. (1988). Maximizing Profits for North American Van Lines' Truckload Division: A New Framework for Pricing and Operations. Interfaces, 18(1): 21-41.

Powell, W.B., Towns, M.T., & Marar, A. (2000). On the Value of Optimal Myopic Solutions for Dynamic Routing and Scheduling Problems in the Presence of User Noncompliance. Transportation Science, 34(1): 67-85.

Regan, A.C., Mahmassani, H.S., & Jaillet, P. (1995). Improving the Efficiency of Commercial Vehicle Operations Using Real-Time Information: Potential Uses and Assignment Strategies. Transportation Research Record, (1493): 188-198.

Regan, A.C., Mahmassani, H.S., & Jaillet, P. (1998). Evaluation of Dynamic Fleet Management Systems: Simulation Framework. Transportation Research Record, 1645, 176–184.

Repoussis, P.P. & Tarantilis, C.D. (2010). Solving the Fleet Size and Mix Vehicle Routing Problem with Time Windows via Adaptive Memory Programming. Transportation Research Part C: Emerging Technologies 18(5): 695-712.

Respen J., Zufferey. N., Potvin, J.-Y. (2014) (working paper). Impact of Online Tracking on a Vehicle Routing Problem with Dynamic Travel Times. Retrieved from: https://www.cirrelt.ca/DocumentsTravail/CIRRELT-2014-05.pdf

Rodriguez, J.M., & Griffin, G.C. (1990). The Determinants of Job Satisfaction of Professional Drivers. Transportation Research Forum, 30(2): 453-64.

Rodriguez, J., Kosir, M., Lantz, B., Griffen, G. & Glatt, J. (2000). The Costs of Truckload Driver Turnover. Upper Great Plains Transportation Institute, NDS University, Fargo, ND.

Schilde, M., Doerner, K.F., & Hartl, R.F. (2011). Metaheuristics for the Dynamic Stochastic Dial-a-Ride Problem with Expected Return Transports. Computers & Operations Research, 38(12): 1719-30.

Schmid, V. (2012). Solving the Dynamic Ambulance Relocation and Dispatching Problem Using Approximate Dynamic Programming. European Journal of Operational Research, 219(3): 611-21.

Shaw, J.D., Delery, J.E., Jenkins, G.D. Jr & Gupta, N. (1998). An Organization-Level Analysis of Voluntary and Involuntary Turnover. Academy of Management, 41(5): 511-25.

Sheffi, Y. (2012). Logistics clusters, delivering value and driving growth. MIT Press, Cambridge, MA.

Simao, H.P., Day, J., George, A.P., Gifford, T., Nienow, J., & Powell, W.B. (2009). An Approximate Dynamic Programming Algorithm for Large-Scale Fleet Management: A Case Application. Transportation Science, 43(2): 178-197.

Stavrinos, D., Jones, J.L., Garner, A.A., Griffin, R., Franklin, C.A., Ball, D., Welburn, S.C., Ball, K.K., Sisiopiku V.P., & Fine P.R. (2013). Impact of Distracted Driving on Safety and Traffic Flow. Accident Analysis & Prevention 61(0): 63-70.

Suzuki, Y., M.R. Crum & Pautsch, G.R. (2009). Predicting Truck Driver Turnover. Transportation Research Part E: Logistics and Transportation Review 45(4): 538-550.

TRANSCORE (2011) Retrieved from: http://www.werc.org/assets/1/Publications/935%20CarrierBenchmarkSurvey2011.pdf

Taguchi, G. and Yokoyama, Y., (1994). Taguchi Methods: Design of Experiments, American Supplier Institute, Dearborn MI, in conjunction with the JS Association, Tokyo, Japan.

Thomas, B.W. 2007. Waiting Strategies for Anticipating Service Requests from Known Customer Locations. Transportation Science, 41(3): 319-331.

Thomas, B.W. & White ,C.C. (2004). Anticipatory Route Selection. Transportation Science,38(4):473-87.

Tjokroamidjojo, D., Kutanoglu, E., & Taylor, G.D., (2006). Quantifying the Value of Advance Load Information in Truckload Trucking. Transportation Research Part E-Logistics and Transportation Review, 42(4): 340-357.

Trucking Information (2015) Retrieved from: http://www.truckinginfo.com/article/story/2009/12/mistaken-identity-do-light-loads-really-cause-excessive-tire-wear.aspx

TRANSCORE (2011) Retrieved from: http://www.werc.org/assets/1/Publications/935%20CarrierBenchmarkSurvey2011.pdf

Transport Canada (2005) Retrieved from: http://publications.gc.ca/collections/collection_2011/tc/T46-14-2005-eng.pdf

US Department of Energy (2011) Retrieved from: http://www1.eere.energy.gov/vehiclesandfuels/facts/m/2011_fotw671.html

White, W. (1972). Dynamic Transshipment Networks: An Algorithm and its Application to the Distribution of Empty Containers. Networks, 3(2), 211-236.

White, W., Bomberault, A. (1969). A Network Algorithm for Empty Freight Car Allocation. IBM Systems Journal, 2(8), 147-171.

Wieberneit, N. (2008). Service Network Design for Freight Transportation: a Review. OR Spectrum, 30 (1), 77-112.

Wang, X., & Regan, A.C. (2002). Local Truckload Pickup and Delivery with Hard Time Window Constraints. Transportation Research, Part B, 36, 97–112.

Yang, J., Jaillet, P., & Mahmassani, H.S. (1998). On-line Algorithms for Truck Fleet Assignment and Scheduling Under Real-Time Information. Transportation Research Record, (1667): 107-113.

Yang, J., Jaillet, P., & Mahmassani, H.S., (2004). Real-Time Multivehicle Truckload Pickup and Delivery Problems. Transportation Science, 38(2): 135-148.

Young, K.L. & Salmon, P.M. (2012). Examining the Relationship Between Driver Distraction and Driving Errors: A discussion of theory, studies and methods." Safety Science 50(2): 165-174.

Zolfagharinia, H., & Haughton, M.A. (2012). The Benefit of Information Sharing in a Logistics Outsourcing Context. Int. J. of Logistics Systems and Management 13 (2), 187–208.

Zolfagharinia, H. & Haughton M. (2014). The Benefit of Advance Load Information for Truckload Carriers. Transportation Research Part E: Logistics and Transportation Review 70(0): 34-54.

Zolfagharinia, H., & Haughton, M. (forthcoming). Effective Truckload Dispatch Decision Method with Incomplete Advance Load Information. European Journal of Operational Research .