1-1-2012

An integrated framework for freight forwarders:exploitation of dynamic information for multimodal transportation

Farshid Azadian
Wayne State University,

AN INTEGRATED FRAMEWORK FOR FREIGHT FORWARDERS: EXPLOITATION OF DYNAMIC INFORMATION FOR MULTIMODAL TRANSPORTATION

by

FARSHID AZADIAN

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2012

MAJOR: INDUSTRIAL ENGINEERING

Approved by:

_____

Advisor Date

_____

_____

_____

_____

DEDICATION

To the memory of my grandfather,

Houssein Dehmand, who taught me

the value of knowledge at an early age

and to my parents Morteza and Rabeeh

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my appreciation to the many people who have contributed their time and energy to the furtherance of this research. This dissertation would not have been possible without the kind help and support of my professors, colleagues, friends, and family.

First and foremost, I would like to express my utmost gratitude to my mentors and advisors Dr. Alper Murat and Dr. Ratna Babu Chinnam for their unselfish and unfailing support and inspiration during this research and throughout my graduate education; I am forever grateful to both of them. I would also like to thank my dissertation committee members Dr. Leslie Monplaisir, Dr. Darin Ellis, and Dr. John Taylor for their time & effort, helpful suggestions, and constructive criticism.

I would like to especially thank Dr. Monplaisir, ISE department chair, and Dr. Kenneth Chelst, former ISE department chair, for their support and encouragement during my PhD program.

I am grateful to the University Transportation Center at University of Toledo for sponsoring this research. This work was supported by funds from the US Department of Transportation through the University of Toledo University Transportation Center.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter One

Introduction

1.1 Motivations

In recent decades, access to cheap labor and raw materials, better financing opportunities, larger product markets, arbitrage opportunities, and additional inducements offered by host governments to attract foreign capital encouraged companies to extend their supply chains over the globe (Manuj and Mentzer 2008). However, the success of these longer supply chains highly relies on the performance of the firms' logistics (Green, Whitten, and Inman 2008). While logistics includes a wide range of activities, one of its fundamental elements is transportation (Bookbinder and Matuk 2009).

The nature of supply chains requires efficient short and long-haul shipping of raw materials, components, and products between manufacturers, retailers and customers. In recent years, many companies have adopted new manufacturing and inventory management strategies (e.g., make-to-order and just-in-time) that aim to reduce costs while improving responsiveness to market demands. However, these approaches demand for fast, and more importantly reliable, transportation. However, since transportation infrastructure has not kept pace with business growth, excess demand over the transportation network capacity has led to

growing congestion and uncertainty in transportation lead-times. According to a 2011 Urban Mobility Report, the US experienced a steady increase in travel time index[1] since 1982 from 1.09 to a national average of 1.2 in 2011 (Schrank and Lomax 2011). They also reportthe cost of congestion to be about $101 billion for delay and fuel waste in 2010 alone. Average line-haul speed on rail freight is about 22 mph (US DoT 2007). Congestion is also an issue in other modes of transportation (US DoT 2007, 2009). A survey conducted by Golob and Regan (2000) shows that 82% of the interviewed companies recognize congestion as a problem (somewhat to critically serious) for their business and over 27% of them often or very often miss their schedules due to congestion. The aforementioned problem can be recognized as a major factor in shift of shippers' demand toward more expensive modes of transportation that provide faster and more reliable services (US DoT 2006).

Using more expensive modes of transportation translates to increased shipping service level expectation that demands for more sophisticated decision making that in turn requires better system-wide information. The advent of the Intelligent Transportations Systems (ITS) provides opportunity for improvement in transportation performance and quality. The core of ITS consists of obtaining, processing, and distributing information for better use of the transportation system, infrastructure and services (Crainic, Gendreau, and Potvin 2009). This includes Geographical Positioning System (GPS), Automatic Vehicle Location System (AVL), Fleet

[1] The Travel Time Index (TTI) is the ratio of peak period travel time to free flow travel time. The TTI expresses the average amount of extra time it takes to travel in the peak relative to free-flow travel. A TTI of 1.3, for example, indicates a 20-minute free-flow trip will take 26 minutes during the peak travel time periods, a 6-minute (30 percent) travel time penalty.

Telematics System (FTS), Wireless Communication, Electronic Data Interchange (EDI), along with internet and other real-time data sharing systems that inform the decision maker about the location of the vehicles and freight and provide better understanding of the network status, especially, under congestion.

The importance of logistics and its complexities are motivating companies to outsource their logistic operations in different levels to third parties known as “freight forwarders” to reduce cost and increase efficiency (Razzaque and Sheng 1998). Freight forwards generally act as an intermediary between shippers and carriers and are responsible for transporting goods in supply chains. Indeed, freight forwarding industry, as part of the broader supply chain management industry, is undergoing a profound transition with the rise of multinational freight forwarders based in Europe, the United States, and Japan that perform integrated logistics services in addition to simple freight forwarding with a range of value-added services (Bowen and Leinbach 2004). However, despite the major integrators (e.g. FedEx, UPS, DHL, BAX Global and alike), majority of the freight forwarders are small- to mid-size companies. Due to high capital investment, schedules and capacities are usually fixed by carriers far in advance and therefore freight forwarders decide on freight routing and book the capacity based on their forecasted demand (Chew et al. 2006).

A freight forwarder generates its profit from the difference between the price that a customer is obliged to pay for the execution of the requested service and the costs of the fulfillment of the request. Moreover, the nature of the freight forwarding industry, especially for small forwarders, is based on personal relations and long-term trust-building that requires

meeting service level expectations and consistence in the quality of service (Agnes 2000). Accordingly, forwarders are challenged to conduct their business with the minimum possible cost while satisfying the shippers' expectation in a competitive market. Achieving this goal requires a sophisticated decision making process that integrates all the related information to produce high quality decisions for freight routing to satisfy the demand in a reasonable time window with minimum cost to generate profit. The goal of this research is to address this need in freight forwarding industry. We, however, limit the scope of the research to consider only multimodal air-cargo transportation as the fastest growing mode of transportation in the U.S. The freight forwarders constitute more than 90% of air-cargo shipments (Hellermann, 2006) and play a critical role in the air mode of transportation.

Air is arguably the most competitive mode of transportation in providing the fastest and most reliable transportation service that is required in today's global supply chains. Over the past decade, there has been a consistent growth in demand for air-cargo deliveries. According to the Bureau of Transportation Statistics (BTS), in 2007, the value of air-cargo shipment goods in the US was over $1.8 trillion, a 31% increase in just five years from a survey in 2002 (Margreta et al., 2009). Futher, despite the financial crises, annual forecast reports by both Airbus (2010) and Boeing (2010) predict a 5.9% annual growth rate for global air-cargo tonnage over the next 20 years.

In response to the demand growth, the air transportation network has been steadily expanding its capacity over the past two decades through establishing new airports, offering more flights options, and investing in road connectivity. One consequence of these

developments is the expansion of service zones of airports and the overlaping of their market catchment regions. This has resulted in the creation of *Multi-Airport Regions* (MARs) where several airports accessible in a region substitute and supplement each other in meeting the region's demand for air transportation (Loo, 2008). These MARs provide alternative access options for passengers as well as air-cargo shippers and forwarders. Accessibility of multiple airports and expansion of transportation options introduce new opportunities and challenges for forwarders that in turn reemphasizes the importance of effective operational decision making for competitiveness.

On the other hand, along with the increasing trend of demand for air transportation, the time variavility measure of the air mode has steadily declined. For example, in July 2007, 28% of the flights in the U.S. domestic market arrived late, up from 19% in July 2003 (BTS, 2010). The impact of these delays is as severe for the time-sensitive air-cargo shipments (common in JIT logistics) as it is for passengers. In fact, when the International Air Transport Association (IATA) asked major shippers for their main issues in February 2008, efficiency (reducing costs) and reliability were identified as the top two issues.[1]

1.2 Research Objectives

In this dissertation, the objective is to provide an operational decision support system for air freight-forwarders for time-sensitive cargo transportation. The goal is to enable them to

[1] Bisignani, G., Plenary speech, IATA World Air Cargo Symposium, 2008.

better and predict the network variability based on historical and real-time information and respond through effective operational planning and scheduling of cargo transportation. The performance measures for the forwarders in this research are the operational costs and service level cost that is measured by the delivery tardiness penalties.

Accordingly, the objectives of this research are,

- Develop a methodology to analyze the historical flight performance, airport congestion state, and announced real-time information to estimate the air-network state at a given time in near future and how it is affecting the air-cargo shipment
- Develop stochastic dynamic as well as deterministic routing models to assist forwarders in the operational planning of air-cargo transportation on a stochastic time-dependent air-road network and enable them to plan for the variability in the stochastic and time-dependent air network.
- Design algorithms for solving the models developed. Specifically, these algorithms identify optimal (near optimal) solutions for the scheduling and routing of air-cargo on the stochastic and time dependent air and road networks.

1.3 Research Scope

In this study, we focus on middle-size freight forwarders that handle freight shipping for different shippers. The freight forwarder is responsible for collecting, sorting, consolidating and delivering time-sensitive goods from different origins to different destinations. The forwarder in this research does not provide extra services usually offered by major integrators such as

warehousing or vendor-managed inventory system. The objective of the forwarder is to deliver the freight before the deadline agreed with shipper while minimizing the operational cost; deviation from the delivery deadline is penalized.

Customer orders are received in advance (before the beginning of each day) and forwarder is responsible to collect and transport the air-cargo shipment orders. Orders are available at customer sites for pickup and they have individual destination delivery airports. The customer orders are time-sensitive with specific delivery deadline at destination airport. We assume that, due to the nature of the orders, there are no economies of scale, e.g., no air-cargo consolidation benefits.

For the long-haul transportation, forwarder relies on contracted air-carriers and is thus obligated to their schedules and capacity limitations. It is assumed that, if needed, further capacity is available to forwarder but with a price that is based on the contract between the forwarder and carrier. In this setting, the air network is stochastic, carriers' on-time performance is not guaranteed, and network disruption is possible. In other words, flights may depart later than the announced schedule or may even get canceled. Moreover, travel time for any flight arc can be different from the expected time. Accordingly, the forwarder needs to prepare to deal with the connectivity problem in intermediate ports and consider these factors in estimating the delivery time and transportation cost. It is assumed that forwarder can implement a dynamic routing policy by altering the freight path on the air network en route; however, there are capacity availability restrictions with this option and re-routing might introduce additional costs. The aforementioned dynamic routing is based on the realization of

the network status (e.g. level of congestion, incidents, and network disruptions). Therefore, in this research we study the value of the information based on the time of realization and fidelity of data.

On the road network, it is assumed that the network is deterministic. Consequently, the connectivity of the network and arc travel times are fixed and known in advance. The freight forwarder is assumed to operate a fleet of identical vehicles to perform the transportation on the road. A fixed cost is imposed for each vehicle's allocation to the pickup and delivery task and variable cost is based on arc travel by each vehicle, e.g. total traveled miles.

1.4 Novelty and Contribution of the Research

This research contributes to the existing literature of air-cargo transportation and operations research. A comprehensive literature review and detailed contributions are presented in each chapter individually. In this section, however, we provide a brief review of the highlights of the research and its contributions.

In the realm of air-cargo transportation, this research is the first work that introduces dynamic cargo routing based on real-time information availability. Considering the stochasticity of air-network, we provide a novel approach to analyze the publicly available historical data to perform a static routing to reduce the expected operational and service cost. We further enhance this approach to incorporate the real-time information, while accounting for its fidelity, to dynamically re route the cargo en route. Through a set of experimental studies and real world based case studies, we demonstrate the performance of this approach in terms of reducing total cost including service level costs.

In addition, this is the first study that provides operational algorithm to implement the concept of alternative access airport policy. This algorithm enables forwarders to increase their competitiveness and reduce their cost by providing a decision support system to expand their options in a multiple airport region.

In terms of contribution to the vehicle routing literature, we introduce a new class of pickup and delivery problem that generalize a many-to-many pickup and delivery problems by considering time dependence and pickup-delivery pairing dependence of the delivery costs. In terms of methodological contribution, we introduce the approach of successive subproblem solving to address the common issues of homogeneous subproblems which result from (Lagrangian) problem decomposition of many vehicle routing problems with identical vehicles. This approach is demonstrated to be very competitive in solving large scale problem instances in reasonable time and with optimality (or near optimality) compared with alternative methods.

1.5 Organization of the Dissertation

In addressing the freight forwarders problem, this dissertation is organized as follows. In Chapter 2, we study the dynamic routing of air-cargo on the air network. In Chapter 3, we consider the short-haul transportation of air-cargo by studying its routing on the road network. The dissertation summary and conclusion are presented in the last chapter.

In Chapter 2, we address the problem of dynamic routing of time-sensitive air-cargo using real-time information on stochastic air-network. We present a procedure to estimate the network parameters including flight departure delays and travel times from historical data based on a origin and destination airport for a given operation day. A static routing policy is

developed through stochastic dynamic programming to minimize the expected operational and delivery tardiness costs. Next, we provide an approach to analyze the real-time information (accounting for their fidelity) to estimate the network parameters and respond by dynamic re routing of the cargo if necessary to minimize the objectives. The performance of the algorithm is evaluated through a set of real-world based case studies.

In Chapter 3, the problem of air-cargo pickup and delivery problem with alternative access airports is studied. We introduce a mixed integer mathematical program for customer order pickup scheduling, fleet routing and allocations, and assignment of customer orders to flights available a multiple regional airports. We decompose the problem based on identical vehicles using Lagrangian decomposition and then develop a successive subproblem solving approach to solve the problem. The performance of this innovative approach is tested through a set of experimental problems and a case study based on the Southern California region.

Chapter Two

Dynamic Routing of Time-Sensitive Air-Cargo Using Real-Time Information

2.1 Introduction

Over the past decade, the unprecedented growth in the global trade has further increased the importance of just-in-time (JIT) logistics and contributed to the growth of the air-cargo industry. According to a recent study for The International Air Cargo Association, the global air-cargo industry carried 100 billion ton-miles with a direct revenue exceeding $50 billion in 2005 (Kasarda et al., 2006). The biennial World Air Cargo Forecast by Boeing forecasts that the world air-cargo traffic will grow at a rate of 5.8% per year over the next 20 years (Boeing, 2008). This growth is accompanied by steady increase in flight delays. For example, in July 2007, 28% of the flights in the U.S. domestic market arrived late, up from 19% in July 2003 (Bureau of Transportation Statistics, 2010). The impact of these delays is as severe for the time-sensitive air-cargo shipments (common in JIT logistics) as it is for passengers. In fact, when the International Air Transport Association (IATA) asked major shippers for their main issues in February 2008, efficiency (reducing costs) and reliability were identified as the top two issues.[1] Facing these challenging trends, freight forwarders and shippers must plan and manage their

[1] Bisignani, G., Plenary speech, IATA World Air Cargo Symposium, 2008.

routes more effectively to improve the delivery performance of air-cargo. Internet companies, such as “Flightstats.com”, “Flightview.com”, “Pathfinder-web.com” and “Flightexplorer.com”, provide historical and real-time flight on-time performance data to improve in-advance planning and real-time management of routes. Further, “Pathfinder-web.com” also provides static routes based on such factors as weather/airport status and on-time statistics. The dynamic route planning for a time-sensitive air-cargo by leveraging the available historical and real-time air-network congestion information is the subject of this study.

A freight forwarder (forwarder in short), upon receiving a time-sensitive shipment, has three options: shipping via (1) an integrator’s (e.g., FedEx, UPS, DHL) express or next-flight-out service, (2) a mixed belly (e.g., United Airlines, Delta Airlines, American Airlines) or combination carrier (e.g., Lufthansa Cargo AG, Korean Air), and (3) chartered/dedicated freighter. Clearly, the forwarder’s decision depends on the reward/penalty structure of the agreement with the shipper as well as on the attributes of the shipment such as size (weight and volume), value density, commodity type (e.g., hazmat), origin and destination, contracted capacity with carriers and so on. In this study we are considering shipments for which chartering dedicated freighter is not economically feasible. Accordingly, the forwarder in this study considers only integrators’ express and next-flight-out service (cost effective for shipments less than 70-150 lbs) and the mixed belly or combination carrier option which provides broader network

coverage with more frequent flight connectivity and significantly lower costs.[2] Furthermore, a shipment route involving multiple carriers, and possibly the integrator, provides the greatest schedule and route flexibility leading to the shortest delivery lead-time. This study is motivated by practical applications affecting different industries. Since the beginning of 2000, automotive OEMs (e.g., GM and Ford) have been shifting their sourcing from domestic facilities to Canada, Mexico and overseas (Klier and Rubenstein, 2008). This has not only increased the supply chain transportation lead-times but also increased the supply chain sourcing risks. Supply disruptions caused by various reasons, such as quality defects and incorrect shipments (quantity, part mix), can halt the assembly processes in multiple facilities. The disruption of an assembly line is estimated to cost $60-100K/hour in a medium-sized finished vehicle assembly plant.[3] In response, the OEMs often resort to expedited shipment by either chartering a freighter or a cargo helicopter for time-definite delivery, which can cost $100Ks depending on the origin-destination and freighter availability. These incidents are routine and OEMs have chartered aircrafts to ship products such as wheels, power trains and transmissions.

The logistic disruptions also arise when a time insensitive and surface divertible cargo becomes a time-sensitive cargo requiring air shipment. Freight forwarders regularly draw shipments from intermodal facilities (e.g. ports, airports, rail terminals) and forward it to the consignees (with or without break bulk). However, due to the late arrival of the vessel or the

[2] For instance, the shipping rate for an LD2 container with dimensions (61.5×60.4×64) inches and weight 1,228 lbs from Cleveland to Seattle on 22 March 2010 with UPS is $4,9K-$8,5K depending on service type and is $933 for Delta Cargo (Source: www.ups.com, http://www.delta.com/business_programs_services/delta_cargo/).

[3] Based on interviews with the managers at Ford MP&L and GM Supply Chain department.

congestion at the intermodal facility, there occur excessive delays such that the cargo becomes no longer suitable for surface diversion (e.g. trucking) and needs to be air shipped. For instance, the Target Logistics, a freight forwarding company in California, US, often experiences delays due to the congestion at the port of Long Beach, California. A container shipment arriving from East Asia may require some of its contents to be air shipped next-flight-out if the delay is excessive. When such an incident occurs, the Target Logistics explores options for the best outbound flight from the regional airports (Los Angeles, Ontario, Oakland, San Diego) by trading off the delivery lead-time with the cost. In addition to considering the flight availability, cost, and size restrictions, the Target Logistics also accounts for the road traffic congestion to the airport and its other shipments and classes for that day. Another practical application is the air-cargo shipments during peak seasons (e.g. Christmas Day) where the demand for both the passenger and the cargo transportation exceeds the supply. C.H. Robinson, a leading third party logistics (3PL) company, provides air-cargo freight forwarding services to manufacturing companies, such as 1st and 2nd Tier automotive suppliers in Michigan, through the Detroit Metropolitan Airport (DTW). Whereas the air-cargo demand is stable and the contracted carrier capacity is sufficient during regular months, C.H. Robinson cannot meet the requested service levels in high demand seasons. For example, during December months, C.H. Robinson determines the flight routes, which are less likely to be congested, and books same-day flights with mixed carriers for its time-sensitive shipments.

The main goal of this study is to investigate the benefits of dynamic (online) routing of a time-sensitive air-cargo on the air network from an origin airport to a destination airport while

accounting for the real-time and historical information (e.g., delays, cancellations, capacity availability) to optimize a given shipment criteria (e.g., cost, delivery lead-time). We study the problem from a freight forwarder's perspective for two reasons. First, more than 90% of air-cargo shipments are handled through freight forwarders (Doganis, 2002). In comparison, shippers sending freight directly with carriers/integrators account for only a small fraction (approximately 5-10%) of total airfreight volume (Althen et al., 2001). Second, due to the industry practice of capacity contracts, the freight forwarders have access to cargo capacity from multiple carriers at favorable terms and rates (Hellermann, 2006). We also note that, in most instances, a static route may be the best option since it is not only the least cost option but can also provide short delivery lead-times. However, for highly time-sensitive shipments and in the absence of routes with short lead-times (or the routes are subject to delays), dynamic routing can provide short delivery lead-times with affordable costs. The approach presented in this study allows freight forwarders to effectively make these trade-off decisions. The proposed approach is a Markov decision process (MDP) model for dynamic routing that differs from other MDP formulations in the literature. Our contribution is three fold. First, we propose a novel departure delay estimation model based on the real-time delay announcement and historical data. Secondly, we provide a dynamic routing model on the air network that differs from those on traditional road networks such that it considers scheduled departures and effect of stochastic travel times and departure delays. The dynamic routing model incorporates the proposed departure delay estimation model. Finally, through experimental studies and real-world case studies, we show that the proposed dynamic routing model can provide significant

savings for freight forwarders. These savings depend on the severity of delays, variability of travel times, availability and accuracy of real-time delay announcements as well as availability of flight alternatives. Lastly, we note the distinction between this paper's problem, freight forwarders' dynamic routing of air-cargo through available flights to improve the overall delivery performance of a single shipment, and the broader and more strategic problem of carriers or integrators planning of their fleet routes and schedules. The later problem concerns an asset owner's (carrier, integrator) operations planning to improve operating performance as well as utilization of aircraft fleet and other assets (Yan et al. 2006, Tang et al. 2008).

The rest of the paper is organized as follows. Survey of relevant literature is given in Section 2. Modeling the dynamic routing of air-cargo and delay estimation is presented in Section 3. Section 4 presents the results of an experimental study conducted to investigate the benefits of dynamic routing and accurate real-time flight status information. Two case study applications of the proposed approach are discussed in Section 5. Finally, Section 6 offers concluding remarks and proposes avenues for future research.

2.2 Literature Review

The problem investigated in this study relates to multiple research streams. The proposed dynamic routing formulation and solution approach is closest to the stochastic time-dependent shortest path problems (STD-SP) and hence we restrict our review to those studies with stochastic and time-dependent arc travel costs. In terms of application, this study also relates to the literature on the estimation of flight departure/arrival delays and cancellations/diversions which is briefly reviewed in the end.

The shortest-path problems are referred as STD-SP when arc costs follow a known probability distribution which is also time-dependent. Hall (1986) studied the STD-SP problems and showed that the optimal solution has to be an 'adaptive decision policy' (ADP) rather than a static route. In an ADP, the node to visit next depends on both the node and the time of arrival at that node, and therefore the classical SP algorithms cannot be used. Hall (1986) employed the dynamic programming (DP) approach to derive the optimal policy. Bertsekas and Tsitsiklis (1991) proved the existence of optimal policies for STD-SP. Later, Fu and Rilett (1998) modified the method of Hall (1986) for problems where arc costs are continuous random variables. They showed the computational intractability of the problem based on the mean-variance relationship between the travel time of a given path and the dynamic and stochastic travel times of the individual arcs. They also proposed a heuristic in recognition of this intractability. Bander and White (2002) modeled a heuristic search algorithm AO* for the problem and demonstrated significant computational advantages over DP, when there exists known strong lower bounds on the total expected travel cost between any node and the destination node. Fu (2001) estimated immediate arc travel times and proposed a label-correcting algorithm as a treatment to the recursive relations in DP. Waller and Ziliaskopoulos (2002) suggested polynomial algorithms to find optimal policies for stochastic shortest path problems with one-step arc and limited temporal dependencies. Gao and Chabini (2006) designed an ADP algorithm and proposed efficient approximations to time and arc dependent stochastic networks. An alternative routing solution to the ADP is a single path satisfying an optimality criterion. For identifying paths with the least expected travel (LET) time, Miller-Hooks

and Mahmassani (1998) proposed a modified label-correcting algorithm. Miller-Hooks and Mahmassani (2000) extended this algorithm by proposing algorithms that find the expected lower bound of LET paths and exact solutions by using hyperpaths.

All of the above studies on STD-SP assume deterministic time dependence of arc costs, with the exception of Waller and Ziliaskopoulos (2002) and Gao and Chabini (2006). However, the change in the cost of traversing an arc over-time can be stochastic as in the flight departure delays. Psaraftis and Tsitsiklis (1993) is the first study to consider stochastic temporal dependence of arc costs and to suggest using real-time information en route. They considered an acyclic network where the cost of outgoing arcs of a node is a function of the environment state of that node and the state changes according to a Markovian process. They assumed that the arc's state is learned only when the vehicle arrives at the source node and that the state of nodes are independent. They proposed a DP procedure to solve the problem. Azaron and Kianfar (2003) extended Psaraftis and Tsitsiklis (1993) by evolving the states of current node as well as its forward nodes with independent continuous-time semi-Markov processes for ship routing problem in a stochastic but time invariant network. Kim et al. (2005a) studied a similar problem as in Psaraftis and Tsitsiklis (1993) except that the information of all arcs are available real-time. They proposed a dynamic programming formulation where the state space includes states of all arcs, time, and the current node. They stated that the state space of the proposed formulation becomes quite large, making the problem intractable. To address the intractable state-space issue, Kim et al. (2005b) proposed state space reduction methods. Thomas and White (2007) study a similar problem as in Kim et al. (2005a) but also consider the amount of

time that an observed arc has spent in a particular state. All these studies consider routing on unscheduled transport networks where there is no schedule induced or switching delays at the nodes as in scheduled networks or multimodal transportation, respectively. There are few studies on the routing problem on multimodal networks with time-dependent arc weights (e.g., cost or travel time). Ziliaskopoulos and Wardell (2000) proposed a time-dependent intermodal optimum path algorithm for deterministic multimodal transportation networks while accounting for delays at mode and arc switching points. Opasanon and Miller-Hooks (2001) proposed the stochastic variation of the approach by Ziliaskopoulos and Wardell (2000) where the mode transfer delays and arc travel times are stochastic and time varying. However, this study assumes independence over time for all probability distributions. Our proposed dynamic routing model differs from earlier models in the STD-SP literature by accounting for the scheduled departures, the effect of stochastic travel times and departure delays. In addition, it admits the real-time announced information on the status of flights and makes routing decisions and updates the delay distributions based on this online information.

The estimation of flight departure/arrival delays and cancellations/diversions has been the subject of several studies (Mueller and Chatterji 2002, Chatterji and Sridhar 2005, Tu et al. 2008). These studies can be categorized into analytical (e.g. queuing), statistical (e.g. regression models) and simulation approaches that vary by computational efficiency and level of detail. For example, the delay and cancellation component in the Federal Aviation Administration (FAA) NAS Strategy Simulator takes a macroscopic approach and obtains approximations of delay based on the aggregate values of input parameters, namely traffic demand and airport

capacity. The majority of delay estimation approaches proposed in the literature predict cancellations and delays at the system level rather than for each individual flight. The only two studies considering the traveler's perspective (e.g. passenger) are Wang and Sherry (2007) and Tien et al. (2008). Whereas Wang and Sherry (2007) estimate delays at a flight level, Tien et al. (2008) propose a model that estimates overall averages across multiple flights. Tien et al. (2008) consider passenger trip scenarios by explicitly accounting for probability of flight cancellation, distribution of flight delay (if not cancelled), and probability of missing a connecting flight. In Section 3.1, we adopt the traveler's perspective approach taken in Wang and Sherry (2007) and Tien et al. (2008) and propose a delay estimation model accounting for flight disruption and recovery scenarios and using historical data to estimate the probabilities. Our model differs from the two studies in that it incorporates real-time information updating while accounting for the fidelity of real-time delay announcement.

2.3 Dynamic Air-Cargo Routing

Let $G \equiv (N, A)$ be the directed graph of an air network with a finite set of nodes $n \in N$ representing *airports* and a set of arcs $l \in A$ representing connecting *flights* between the airports. Since there can be multiple flights between any airport pairs, we designate each flight with a distinct arc. In particular, let $A_l \subseteq A$ denote the set of flights between airports n' to n'' where $l = (n', n'')$, then $i \in A_l$ denotes a unique flight from n' to n''. In the remainder of this work, we refer to these flights as arcs. A dynamic routing problem on this air network is concerned with departing from the *origin node* (n_0) and arriving to the *destination node*

(n_d) via a series of airport/flight selection decisions. The goal is to find an optimal routing policy that minimizes a total cost criterion.

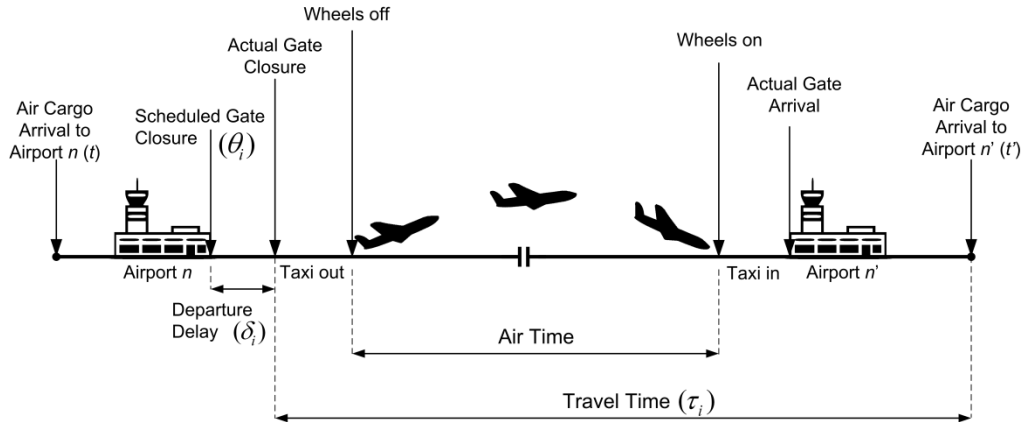


Figure 2.1. Time sequence of the air-cargo arriving to airport *n* at time *t*, departing at $\theta_i + \delta_i$ and arriving to the airport *n'* at time *t'*=$\theta_i + \delta_i + \tau_i$.

The flight arcs have three parameters affecting the flight selection decisions which are illustrated through the time sequence depiction in Figure 2.1. First parameter is the stochastic *travel time* of arc i (τ_i) which is measured as the duration from the gate closure at the origin airport until the unloading of the air-cargo at the destination airport. This duration includes taxi-out at the origin airport, air time (e.g., flight duration), taxi-in at the destination, and unloading time. The second parameter is the *scheduled departure time of flight* (θ_i). Node arrival prior to θ_i results in waiting until departure. Whereas the scheduled departure times are exactly known, the arrival time to the airport node is unknown making the waiting time at the node a stochastic variable. For the purpose of notational clarity and without loss of generality, we assume any cargo processing times (e.g. security checks, processing prior to being loaded onto the aircraft) are already accounted for in the scheduled departure time θ_i. Alternatively, θ_i can be considered as the scheduled cut-off time for flight i for air-cargo acceptance. The final

parameter, absent from most network routing models, is the stochastic departure delay (δ_i) corresponding to an uncontrollable waiting time at the origin node of an arc (flight i) before traveling through it. Therefore, the total waiting time for an air-cargo of flight i arriving to the airport at time t is jointly determined by the waiting due to scheduled departure time $\max\{t - \theta_i, 0\}$ and the departure delay δ_i. Accordingly, if the flight has not departed past the scheduled departure time, the actual departure time depends on δ_i, which is stochastic. Once the flight has departed, the arc becomes unavailable. This temporal change in arc availability is another attribute that distinguishes this problem setting from the other STD-SP problems.

The departure delay (δ_i) is attributable to a multitude of factors that can be classified as the congestion at the origin and destination airports, weather, equipment (mechanical failures, late pushback tug, etc.), personnel (unavailable flight crew or gate agents, etc.), ground operations, passenger/cargo processing/loading delays, unscheduled maintenance and so forth (Mueller and Chatterji, 2002). The departure delay can be negative, zero or positive. The cases $\delta_i = 0$ and $\delta_i > 0$ indicate on-time and late departures, respectively. We adapt "*DepDelayMinutes*" definition of the Bureau of Transportation Statistics (BTS) where the departure delay is defined as the difference between scheduled and actual departure time and early departures are set to *0* and regarded as on-time departures. Accordingly, we consider only the non-negative departure delays in our routing model for three reasons. First, the early departures durations are very small compared to late departure delays and thus the effect on the routing policy decisions is minimal. For instance in 2010, the average early departure delays for all flights outbound from Detroit, Atlanta, Memphis, New York (LaGuardia), Minneapolis,

Charlotte and Dallas airports were -3.8, -3.8, -4.0, -5.4, -4.0, -4.4, and -3.7 minutes which are negligible compared to the average late departure delays of 31.9, 32.4, 33.2, 48.5, 26.2, 29.8, and 32.7 minutes, respectively (BTS, 2010). Secondly, early departures are only possible once all the cargo is loaded (or passengers have boarded). This can only happen if the capacity is full or if the routed cargo is already loaded on the plane. In the former case, the flight is unavailable due to insufficient capacity and need not be considered in routing. In the latter case, we already selected this flight and considering its negative departure possibility would only further support the inclusion of the flight in the routing policy. Lastly, only the late departure delay information is announced in real-time (i.e. early departures are not announced).

Most carriers accept cargo reservations in advance, e.g., in hours, which is sufficient for a forwarder to book a flight while en route or at the preceding airports. These booking cut-off times (a.k.a. closeout or lockout times) are typically 30-60 minutes for shipments under 100 lbs and 1-2 hours for larger shipments depending on the carrier and airport. The cut-off times for transfers can range between 30 minutes to several hours, depending on the connection type (domestic or international), carrier and airport operations, and on whether the cargo is loose or containerized in Unit Load Devices (ULDs). During transshipment of air-cargo from one aircraft to another, the forwarders are subject to the line-up area check-in time (Nsakanda et al., 2004). This line-up area is the final sequencing stage of shipments in ULDs or pallets before the loading onto an aircraft. The latest check-in time depends on the carrier, aircraft size and airport operations. Nsakanda et al. (2004) report on terminal cut-off time of 45 minutes as the latest time to send an ULD or a cart to the staging area. In our model, we consider the carrier cut-off

times for the initial loading at the origin airport and the line-up area cut-off times for transfers at the intermediate airports. We assume that the forwarder can freely revise, at some cost/penalty if necessary, its booking decisions prior to arriving at a node subject to the cut-off times.[4] However, upon arriving at a node, the final flight selection decision is made and then the air-cargo is loaded on the aircraft. We assume that there is no recourse decision at that node once the air-cargo is loaded meaning the flight decision is permanent. This is a reasonable assumption since the freight forwarders often do not have the flexibility to get their cargo loaded and unloaded at a short notice due to physical constraints.

Any flight at a given time can be in one of the two states: available or unavailable for loading the air-cargo. The unavailable flights are those that are departed, diverted, cancelled (due to insufficient load levels, bad weather conditions, operational failures, etc.) or no-longer accepting cargo (e.g., past cut-off time or insufficient capacity). Sometimes, the flight delays can be lengthy and we consider delays larger than a threshold level (ξ) as *excessive delays*.[5] The availability of a flight is random and cannot be fully guaranteed while the cargo is en route, so we rely on probability estimates from the historical data on flight cancellations and diversions that are publicly available from the BTS and the FAA's Operations Network (OPSNET). It is also

---

[4] The U.S. Bureau of Customs and Border Protection (CBP) and Canada Border Services Agency (CBSA) require freight forwarders to transmit air-cargo and conveyance data several hours in advance for both inbound and outbound shipments. However, for short-haul distances, this requirement is prior to time of departure ("wheels up") of aircraft for first U.S. or Canadian airport of arrival and is thus not limiting the changes in flight routes.(*Source:*http://www.cbp.gov/xp/cgov/trade/automated/automated_systems/ams/camir_air/, http://www.cbsa-asfc.gc.ca/prog/aci-ipec/menu-eng.html)

[5] Delays longer than a threshold typically lead to cancellation or other recovery methods, rather than delays subsequent flights (AhmadBeygi et al., 2008)

possible that not all flights outgoing from an airport will have cargo space available. The availability of cargo space is further affected by the seasonality and trends in air-cargo supply and demand volumes. In the absence of real-time information on the availability of cargo space for short-term booking, we account for unavailability through historically estimated probabilities. In Section 3.2, we incorporate the flight unavailability due to cancellation, diversion and lack of cargo space. Section 3.1 presents a delay prediction model which considers real-time announced delay information and its fidelity. Given that this real-time information is broadcast by the carriers, airports and FAA, they reflect the best information available from the delay and cancellation estimation processes used in practice.

2.3.1 Modeling Departure Delay

In this section, we first describe the distribution of the departure delay given the real-time announced delay information. Then, we present the delay modeling approach used in the dynamic air-cargo routing model.

Let's denote the density and cumulative distribution functions of the departure delay for a flight i with $\psi(\delta_i)$ and $\Psi(\delta_i)$, respectively. Let α_i denote the "on-time" departure probability of flight i, i.e. $\psi(\delta_i = 0) = \alpha_i$ and δ_i follows any continuous distribution for delayed flights $\delta_i > 0$. This distinction between delayed and on-time flights allows for empirical estimation of the delay distributions by fitting common continuous distributions such as Exponential and Weibull. For routing purposes, we assume that the flight departure delay is bounded with a *finite delay* (ξ) such that after ξ the flight is *considered* as unavailable. Provided that ξ is chosen sufficiently large, any flight that is delayed longer than ξ but eventually departed is not only of

little value for dynamic routing but is also considered an outlier (Tu et al, 2008). Further, as per the definition of the BTS, early departures are regarded as on-time departures and δ_i is set to 0. Accordingly, we have $\Psi(\delta_i) = 1 - \alpha_i$ for $0 < \delta_i < \zeta$ and $\psi(\delta_i) = 0$ for $\delta_i < 0, \delta_i \geq \zeta$.

At any given time t, the decision maker has access to real-time information on the departure delay $\hat{\delta}_i(t)$ for $i = 1,2 \ldots, |A|$ as forecasted by the carriers and airports. This information is referred as the *announced departure delay* and is assumed imperfect. To simplify the notation, we will suppress the time from the announced delay and use $\hat{\delta}_i$ instead. Given the announced delay $\hat{\delta}_i$, the distribution $P(\delta_i|\hat{\delta}_i)$ represents the degree of accuracy in the departure delay announcement, e.g. $P(\delta_i = \hat{\delta}_i|\hat{\delta}_i) = 1$ corresponds to the case of perfect information. However, once the real-time announcement ($\hat{\delta}_i$) on the departure delay (δ_i) is revealed, we assume that the information is tail conditionally accurate such that the flight will not depart earlier than the announced departure delay, i.e. $P(\delta_i) = 0$ for $\delta_i \leq \hat{\delta}_i$. Note that if there is no announcement, then either the flight departs on time or will be delayed without an announcement. In the latter case, the announced delay is considered as a zero delay announcement, e.g. $\hat{\delta}_i = 0$. We assume announced delays can be updated but are non-decreasing with time, i.e. $\hat{\delta}_i(t_1) \leq \hat{\delta}_i(t_2)$ for $t_1 \leq t_2$.

Given the historical data on announced and actualized delays, one can estimate the conditional probability of the actualized delay given the announced delay. The estimation of $P(\delta_i|\hat{\delta}_i)$ requires the availability of sufficient historical data on the actualized departure delays and the associated announced delays. For any given flight, these historical data sets are usually sparse considering the effect of other determining factors such as seasonality (e.g., time of the

day, day of the week, month) and non-recurring events (e.g., weather conditions, special days). Accordingly, we instead approximate this distribution by considering the *intervals* for the announced delay, $\hat{\delta}_i \in (v_i^r, v_i^{r+1})$ for $r = 1, \dots, m$, where v_i^r and v_i^{r+1} define the upper and lower bound on the departure delay for interval r, respectively. Note that $v_i^{r_1} \leq v_i^{r_2}$ for $r_1 \leq r_2$ and $v_i^1 = 0$ and $v_i^{m+1} = \xi$. The delay intervals (v_i^r, v_i^{r+1}) can be determined through a bi-variate clustering method (e.g. Gaussian Mixture Model clustering), which can then be used to estimate the joint distribution of announced and actualized departure delay, e.g. $P(v_i^{r'} \leq \delta_i \leq v_i^{r'+1}, v_i^r \leq \hat{\delta}_i \leq v_i^{r+1})$. Hence, prior to receiving any departure delay information (e.g. delay announcement) and before the scheduled departure time, the delay distribution of a delayed flight i satisfies $\sum_r P_r = 1$ where $P_r = P(v_i^r \leq \delta_i \leq v_i^{r+1})$. Given that the announced delay information is tail conditionally accurate, we have $P(\delta_i) = 0$ for $\delta_i \leq v_i^r$ and $v_i^r \leq \delta_i \leq \hat{\delta}_i$ where $\hat{\delta}_i \in (v_i^r, v_i^{r+1})$. Therefore, number of intervals (m) determines the announcement fidelity, e.g. the larger the m, the more accurate the announcements are.

Based on announced delay intervals, for a given flight i, we calculate the probability density function of departure delay given departure delay announcement $\hat{\delta}_i$, where $\hat{\delta}_i \in (v_i^r, v_i^{r+1})$ for a given $1 \leq r \leq m$ as,

$$P\left(\delta_i \middle| v_i^r \leq \hat{\delta}_i \leq v_i^{r+1}\right) = \sum_{r'>r}^{m} P\left(\delta_i \middle| v_i^{r'} \leq \delta_i \leq v_i^{r'+1}\right) P\left(v_i^{r'} \leq \delta_i \leq v_i^{r'+1} \middle| v_i^r \leq \hat{\delta}_i \leq v_i^{r+1}\right) + P(\delta_i | \hat{\delta}_i \leq \delta_i \leq v_i^{r+1}) P\left(\hat{\delta}_i \leq \delta_i \leq v_i^{r+1} \middle| v_i^r \leq \hat{\delta}_i \leq v_i^{r+1}\right). \quad (1.)$$

Note that $P\left(v_i^{r'} \leq \delta_i \leq v_i^{r'+1} \middle| v_i^r \leq \hat{\delta}_i \leq v_i^{r+1}\right) = 0$ for $r' < r$ and the last term in (1) corresponds to the case $r' = r$. From the Bayes' rule, we have,

$$P\left(v_i^{r'} \le \delta_i \le v_i^{r'+1} \middle| v_i^r \le \hat{\delta}_i \le v_i^{r+1}\right) = \frac{P\left(v_i^{r'} \le \delta_i \le v_i^{r'+1}, v_i^r \le \hat{\delta}_i \le v_i^{r+1}\right)}{P\left(v_i^r \le \hat{\delta}_i \le v_i^{r+1}\right)} \quad \text{for } r' > r, \tag{2.}$$

$$P\left(\hat{\delta}_i \le \delta_i \le v_i^{r+1} \middle| v_i^r \le \hat{\delta}_i \le v_i^{r+1}\right) = \frac{P\left(\hat{\delta}_i \le \delta_i \le v_i^{r+1}, v_i^r \le \hat{\delta}_i \le v_i^{r+1}\right)}{P\left(v_i^r \le \hat{\delta}_i \le v_i^{r+1}\right)}. \tag{3.}$$

We calculate $P\left(\delta_i | v_i^{r'} \le \delta_i \le v_i^{r'+1}\right)$ in (1) for $r' > r$ and $\delta_i \in (v_i^{r'}, v_i^{r'+1})$ as,

$$P\left(\delta_i | v_i^{r'} \le \delta_i \le v_i^{r'+1}\right) = \frac{P\left(\delta_i, v_i^{r'} \le \delta_i \le v_i^{r'+1}\right)}{P\left(v_i^{r'} \le \delta_i \le v_i^{r'+1}\right)} = \frac{\psi(\delta_i)}{\Psi\left(v_i^{r'+1}\right) - \Psi\left(v_i^{r'}\right)}. \tag{4.}$$

For $\delta_i \notin (v_i^{r'}, v_i^{r'+1})$, we have $P\left(\delta_i | v_i^{r'} \le \delta_i \le v_i^{r'+1}\right) = 0$.

Similar to the derivation of (4), the density in the second term of (1) for $\delta_i \in (\hat{\delta}_i, v_i^{r+1})$ is expressed as,

$$P(\delta_i | \, \hat{\delta}_i \le \delta_i \le v_i^{r+1}) = \frac{\psi(\delta_i)}{\Psi(v_i^{r+1}) - \Psi\left(\hat{\delta}_i\right)}, \tag{5.}$$

and is 0 if $\delta_i \notin (v_i^{r'}, v_i^{r'+1})$.

Hence, the conditional delay distribution is as follows:

$$P\left(\delta_i | v_i^r \le \hat{\delta}_i \le v_i^{r+1}\right)$$
$$= \psi(\delta_i)\left(\frac{P\left(\hat{\delta}_i \le \delta_i \le v_i^{r+1} \middle| v_i^r \le \hat{\delta}_i \le v_i^{r+1}\right)}{\Psi(v_i^{r+1}) - \Psi(\hat{\delta}_i)} + \sum_{r'>r}^{m} \frac{P\left(v_i^{r'} \le \delta_i \le v_i^{r'+1} \middle| v_i^r \le \hat{\delta}_i \le v_i^{r+1}\right)}{\Psi\left(v_i^{r'+1}\right) - \Psi\left(v_i^{r'}\right)}\right). \tag{6.}$$

For *m=1* (e.g. $v_i^1 = 0$ and $v_i^2 = \xi$) and $\hat{\delta}_i = 0$, the expression in (6) is $\psi(\delta_i)$. For notational simplicity, we define conditional delay distribution of flight i at time t with announced delay $v_i^r \le \hat{\delta}_i \le v_i^{r+1}$ as:

$$q_i\left(\delta_i, \hat{\delta}_i\right) = P\left(\delta_i = t - \theta_i \middle| v_i^r \le \hat{\delta}_i \le v_i^{r+1}\right), \tag{7.}$$

Note that we suppress t for $q_i\left(\delta_i, \hat{\delta}_i\right)$ in (7) and assume that it will be clear from the context. We further define the cumulative probability that the delayed flight i departs at or

after t as:

$$Q_i(t, \hat{\delta}_i) = \sum_{t'=t}^{\theta_i+\xi} q_i(\delta_i = t' - \theta_i, \hat{\delta}_i), \tag{8.}$$

In the next section, we describe our dynamic programming model for air-cargo routing.

2.3.2 Dynamic Programming Model for Air-Cargo Routing

The objective of the dynamic air-cargo routing model is to minimize the expected cost criteria for a trip originating at origin n_o and concluding at destination n_d. The cost criteria can be a function of the service level (e.g. delivery time), a penalty function measuring earliness/tardiness of arrival time to the final destination, itinerary cost or a weighted combination of these criteria. We assume that the forwarder has already booked an itinerary (called static path) and thus secured the cargo space availability on this path. As long as the forwarder does not deviate from the static path, there are no additional flight booking and handling costs; otherwise there is a one-time penalty for breaching the booking contract.

Consider a flight path p between (n_o, n_d) where $p = (i_1, i_2, \dots, i_k, \dots, i_K),\ k = 1,2, \dots, K$ is defined as sequence of flights such that $i_k \in A_{l_k}$ where $l_k \equiv ({n'}_{l_k}, {n''}_{l_k})$ and ${n'}_{l_{k+1}} \equiv {n''}_{l_k}$. Note that ${n'}_{l_1} \equiv n_o$ and ${n''}_{l_K} \equiv n_d$. Let p_S indicate the static path. Denote the set $I(n,t) \subseteq A$ as the set of flights scheduled to depart from node n with departure times $\theta_i \leq t$ for $\forall i \in I(n,t)$. Each node on a flight path is a decision stage (or epoch) at which a routing decision (i.e., which flight to select next) is to be made. Let n_k be the airport location of k^{th} decision stage, t_k is the time at k^{th} decision stage where $t_k \in \{1, \dots, T\},\ T > t_K$. Note that we are discretizing the planning horizon. Since the objective of our dynamic air-cargo routing model

can be expressed as an additive function of the cost of individual stages on the flight path, the dynamic flight selection problem can be modeled as a dynamic programming model.

The state of the system at k^{th} decision stage is denoted by $\Omega_k = \Omega(n_k, t_k, \widehat{\Delta}_k, z_S^k)$. This state vector is composed of the state of the air-cargo and flight network and is thus characterized by the current node (n_k), the current node arrival time (t_k), and the *announced departure delay state* of all flights at time t_k at stage k, i.e., $\widehat{\Delta}_k = \{\hat{\delta}_i,\ \forall i \in A\}$, and static route indicator, i.e., $z_S^k = 1$ if the flights are selected from the static route until stage k and $z_S^k = 0$ otherwise. After the air-cargo is loaded on to a flight, there is a chance that the flight becomes unavailable (e.g., cancelled) or forwent without a penalty due to excessive delay. In either case, the forwarder is faced with the task of choosing another flight. It can be shown that an optimal policy decision is to account for not only the first flight choice but collectively all recourse flights. Therefore, we define the action space for the state Ω_k as *the set of all orderings of all the available flights* scheduled to depart from airport n_k denoted with $I^o(n_k, t_k)$. Note that this list accounts for all the restrictions experienced by the freight forwarders such as the restriction of the aircraft for certain ULD classes and unavailability of spot capacity on a certain flight at t_k.

At every decision stage, the air-cargo freight forwarder evaluates the alternative flight orderings from the "current" node based on the expected cost-to-go. The expected cost-to-go at a given node with the selection of a flight ordering is the expected total cost of the flight ordering selected and the cost-to-go from the next node. Let $\pi = \{\pi_1, \pi_2, \dots, \pi_K\}$ be the policy of the routing and is composed of policies for each of the $K - 1$ decision stages. For a given state Ω_k, the policy $\pi_k(\Omega_k)$ is a deterministic Markov policy that chooses an ordering of flights

departing from node n_k, i.e., $\pi_k(\Omega_k) \in I^o(n_k, t_k)$. Therefore, the expected total cost for a given policy vector $\pi = \{\pi_1, \pi_2, \dots, \pi_K\}$ is as follows:

$$F\left(n_o, t_1, \widehat{\Delta}_1, 1|\pi\right) = E_{\Delta_k}\left\{g_{K+1}(\Omega_{K+1}) + \sum_{k=1}^{K} g_k(\Omega_k, \pi_k(\Omega_k), \Delta_k)\right\}, \quad (9.)$$

where $\left(n_o, t_1, \widehat{\Delta}_1, 1\right)$ is the starting state of the system and the Δ_k is the actualized departure delay vector at stage k. The single stage cost $g_k(\Omega_k, \pi_k(\Omega_k), \Delta_k)$ is cost of the flight ordering selected given the actualized departure delay Δ_k. The $g_{K+1}(\Omega_{K+1})$ is the penalty function based on the earliness/tardiness of arrival time to the final destination n_d. Then, the minimum expected total cost can be found by minimizing $F_o\left(n_o, t_1, \widehat{\Delta}_1, 1\right)$ over the policy vector $\pi = \{\pi_1, \pi_2, \dots, \pi_K\}$ as follows:

$$F^*\left(n_o, t_1, \widehat{\Delta}_1, 1\right) = \min_{\pi=\{\pi_1,\pi_2,\dots,\pi_K\}} F\left(n_o, t_1, \widehat{\Delta}_1, 1|\pi\right). \quad (10.)$$

The corresponding optimal policy is then,

$$\pi^* = \min_{\pi=\{\pi_1,\pi_2,\dots,\pi_K\}} F\left(n_o, t_1, \widehat{\Delta}_1, 1|\pi\right). \quad (11.)$$

Hence, the Bellman's cost-to-go equation can be expressed as follows (Bertsekas, 2005):

$$F^*(\Omega_k) = \min_{\pi_k} E_{\Delta_k}\left\{g_k\left(\Omega_k, \pi_k(\Omega_k), \widehat{\Delta}_k\right) + F^*(\Omega_{k+1})\right\} \quad \forall k = 1, \dots, K. \quad (12.)$$

We now derive the $F^*(\Omega_k)$ in (12). Consider the k^{th} decision stage where the air-cargo has arrived to node n_k at t_k. An outbound flight from n_k can be in either available or unavailable state at t_k. Let γ_i denote the steady-state probability that flight i is not cancelled or diverted and has also sufficient capacity.[6] Then the probability that the i^{th} flight is available at t_k and can depart with the air-cargo, $P_i\left(t_k, \hat{\delta}_i\right)$, is calculated as follows,

[6] Note that we also include in this probability the cases where the flight can be delayed longer than the threshold ξ.

$$P_i(t_k, \hat{\delta}_i) = \gamma_i Q_i(t_k, \hat{\delta}_i). \quad (13.)$$

Therefore, the chance of loading the cargo on a flight increases with the cumulative probability that the delayed flight i departs at or after t given the announcement $\hat{\delta}_i$. For notational simplicity, we also define $\bar{P}_i(t_k, \hat{\delta}_i) = 1 - P_i(t_k, \hat{\delta}_i)$. Note that equation (13) assumes that the cancellation/diversion and delay decision processes are complementary, e.g., on-time or delayed if not cancelled. When the on-time departures are not possible with reasonable delay (due to a late arrival, mechanical, weather, congestion, or staffing issue), the carrier or airport then opts to cancel (Jarrah et al. 1993, Rupp and Holmes 2006).

For a given state Ω_k, $C_i(\Omega_k)$ is the cost of selecting flight i and depends on whether the route deviates from the static path. The $C_i(\Omega_k)$ can be in one of the three cases (*I*,*II*, or *III*),

$$C_i(\Omega_k) = \begin{cases} 0 & if\ i = p_S(k)\ and\ z_S^k = 1, \quad (Case\ I) \\ L_i & if\ z_S^k = 0, \quad (Case\ II) \\ H + L_i & if\ i \neq p_S(k)\ and\ z_S^k = 1, \quad (Case\ III) \end{cases} \quad (14.)$$

where, $p_S(k)$ is the k^{th} flight in the static route, L_i is the air-cargo and handling fare of flight i, and H is the penalty cost of forgoing the static itinerary, e.g. the air-cargo booking price for the static itinerary. The case (*I*) corresponds to maintaining the static path by choosing flight i from the static path. The case (*II*) corresponds to the scenario where the route has already deviated from the static path and thus L_i is incurred for using flight i. Case *III* corresponds to deviating from the static path in stage. In cases (*I*) and (*II*), we assume that once the route deviates from the static path, all future flights are booked with full booking fee. In lieu, one can assume flights on static path can be used at a cost less than full fee $\hat{L}_i < L_i$, i.e., for case (*II*), $C_i(\Omega_k) = L_i$ if $i \neq p_S(k)$ and $z_S^k = 0$, and, $C_i(\Omega_k) = \hat{L}_i$ if $i = p_S(k)$ and $z_S^k = 0$.

We note that certain privileges such as the availability of discount fares and waivers for route changes are captured through H and L_i.

Subsequently, the cost-to-go of flight i at time t_k with departure delay information $\widehat{\Delta}_k$ is calculated as,

$$f_i\left(n_k, t_k, \widehat{\Delta}_k, z_S^k\right) = C_i(\Omega_k) + \sum_{\delta_i}\sum_{\tau_i} q_i\left(\delta_i, \hat{\delta}_i\right) P(\tau_i|\delta_i)\, F^*\left(n_{k+1}, \theta_i + \delta_i + \tau_i, \widehat{\Delta}_{k+1}, z_S^{k+1}\right), \quad (15.)$$

where C_i stage k cost of choosing flight i, δ_i is the stochastic departure delay satisfying $\hat{\delta}_i \le \delta_i \le \xi$, the $\theta_i + \delta_i + \tau_i$ is the stochastic arrival time to n_{k+1}, $q_i\left(\delta_i, \hat{\delta}_i\right)$ is the conditional departure delay probability in (10), probability $P(\tau_i|\delta_i)$ is the conditional probability of the travel time given the departure delay, and $z_S^{k+1} = 1$ if $z_S^k = 1$ and $i = p_S(k)$ and $z_S^{k+1} = 0$ otherwise.

Let $\pi_k \in I^o(n_k, t_k)$ denote a flight ordering of all the available flights at t_k outgoing from node n_k. This ordering is determined based on the cost-to-go of individual flights, i.e. $(j) < (i)$ if $f_{(j)}\left(n_k, t_k, \widehat{\Delta}_k, z_S^k\right) \le f_{(i)}\left(n_k, t_k, \widehat{\Delta}_k, z_S^k\right)$, where (i) and (j) are the rankings of flights i and j, respectively. We can calculate the probability of departing with flight j in the flight ordering π_k as $P_{(i)}\left(t_k, \hat{\delta}_{(i)}\right)\left[\prod_{(j)<(i)} \bar{P}_{(j)}\left(t_k, \hat{\delta}_{(j)}\right)\right]$ which considers that all higher ranked flights are unavailable. Then, the expected cost-to-go of the flight ordering π_k at time t_k with departure delay information $\widehat{\Delta}_k$ from node n_k is,

$$F\left(n_k, t_k, \widehat{\Delta}_k, z_S^k|\pi_k\right) = M\prod_{i\in\pi_k} \bar{P}_i\left(t_k, \hat{\delta}_i\right) + \sum_{(i)\in\pi_k} P_{(i)}\left(t_k, \hat{\delta}_{(i)}\right)\left[\prod_{(j)<(i)} \bar{P}_{(j)}\left(t_k, \hat{\delta}_{(j)}\right)\right] f_{(i)}\left(n_k, t_k, \widehat{\Delta}_k, z_S^k\right). \quad (16.)$$

Here, M is a large delivery failure penalty cost paid by the forwarder to the shipper if the shipment is not delivered beyond a threshold delay. The agreement between the air-cargo

forwarder and shipper carries performance guarantee clauses that usually restrict this penalty. Clearly, as M increases, the routing decisions become more conservative, e.g. choose airports with more flights or flight availabilities.

The expression in (12) is calculated as,

$$F^*(\Omega_k) = \min_{\pi_k} \Bigg\{ \sum_{(i)\in\pi_k} P_{(i)}\big(t_k, \hat{\delta}_{(i)}\big) \Bigg[\prod_{(j)<(i)} \bar{P}_{(j)}\big(t_k, \hat{\delta}_{(j)}\big) \Bigg] C_{(i)}(\Omega_k) + M \prod_{i\in\pi_k} \bar{P}_i\big(t_k, \hat{\delta}_i\big) + \sum_{(i)\in\pi_k} P_{(i)}\big(t_k, \hat{\delta}_{(i)}\big) \Bigg[\prod_{(j)<(i)} \bar{P}_{(j)}\big(t_k, \hat{\delta}_{(j)}\big) \Bigg] \Bigg(\sum_{\delta_i} \sum_{\tau_i} q_i\big(\delta_i, \hat{\delta}_i\big)\, P(\tau_i|\delta_i)\, F^*\big(n_{k+1}, \theta_i + \delta_i + \tau_i, \widehat{\Delta}_{k+1}, z_S^{k+1}\big) \Bigg) \Bigg\} \quad \forall k = 1, \dots, K. \tag{17.}$$

The backward induction approach is often used to solve $F^*(\Omega_k)$ for an optimal policy π^* offline, i.e. before the trip starts. However, the size of the state space is $O\big(2|N|Tm^{|A|}\big)$ makes the offline solution strategy prohibitive for $m \geq 2$. For instance, let's consider the scenario where there are $|N| = 10$ airports each with 8 outbound flights on the average and the air-cargo trip duration is $T = 216$ time units (e.g., 18 hours discretized with 5 minute time intervals). Whereas we have $2|N|Tm^{|A|}$ =4,320 states for $m = 1$, the size of the state-space grows to $2|N|Tm^{|A|} \cong 4 \text{ x } 10^{27}$ for $m = 2$. Instead, we solve for $F^*(\Omega_k)$ using the backward induction algorithm online which has the complexity equivalent to the case with $m = 1$. The departure delay information for all flights $\widehat{\Delta}_k$ is available at the time of decision t_k in epoch k. Since the permanent flight selection decision (i.e. an ordering of flights) is made and cargo is committed at the time of node arrival (t_k), the only departure delay information available and used for this decision is $\widehat{\Delta}_k$. Therefore, the online backward induction approach uses stationary departure delay information $\widehat{\Delta}_k$ in making a decision at a node. In the next decision epoch

$k+1$, the flight selection decision is made based on the new departure delay information $\widehat{\Delta}_{k+1}$.

2.4 Experimental Study

Experimental study investigates the effect of such problem parameters as accuracy of announced delay information, distribution parameters of the departure delay, effect of travel time variability, and number of air-connections on various performance criteria (e.g., expected cost and delivery reliability).

The experimental study is based on five problem configurations ($\mathcal{N}_0$,$\mathcal{N}_1$,$\mathcal{N}_2$,$\mathcal{N}_3$,$\mathcal{N}_4$) as illustrated in Figure 2.2 together with the problem parameters. The parameters for every flight are the probability of on-time departure (α), the expected departure delay (μ) if the flight is delayed, the scheduled departure time (θ), and the mean (τ) and standard deviation (σ) of the Gaussian travel time distribution. In all configurations, the origin airport is *A* (origin), the destination airport is *D* (destination), and there are two alternative intermediate airports (*B* and *C*) with inbound flights from the origin airport. Furthermore, the expected total trip time of going through *B* or *C* is same for all three networks.

The $\mathcal{N}_0$ configuration represents the baseline configuration from which the other network configurations are constructed. In the baseline, the flights' travel times are deterministic; i.e. $\sigma_{AB} = \sigma_{CD} = \sigma_{AC} = \sigma_{BD} = 0$. The mean travel times for flights between the same airport pair are same as shown in Figure 2.2, e.g. $E(\tau)$=200 for flights 5, 6, and 9 between *B* and *D* in $\mathcal{N}_0$. The $\mathcal{N}_1$ and $\mathcal{N}_2$ configurations are identical to the baseline except for the standard deviation of the flights' travel times. In the $\mathcal{N}_1$ configuration, coefficient of variation

(CV) for flights' travel times are set at 5%, i.e., $\sigma_{AB} = \sigma_{CD} = 5$ and $\sigma_{AC} = \sigma_{BD} = 10$. In the $\mathcal{N}_2$ configuration, CV is set at 20%, i.e., $\sigma_{AB} = \sigma_{CD} = 20$ and $\sigma_{AC} = \sigma_{BD} = 40$. The $\mathcal{N}_3$ network configuration is identical to the baseline except that direct flights from *B* to *D* are replaced with one-stop flights connecting at node *E.* The $\mathcal{N}_4$ network differs from the baseline at airport *C*, where we consider six scenarios, e.g. $[S_1, S_2, S_3, S_4, S_5, S_6]$, for the departure delay distribution of the outgoing flights (7, 8, 10). Note that the expected delay of flights (7, 8, 10) are identical in all six scenarios. We assume that, at t_0=95, the cargo is processed and ready for loading onto the first available flight. Further, the due date is set at *T*=100, e.g., the cargo requires expedited shipment. In order to better understand the effect of parameters and without loss of any generality, we consider total trip time as the performance measure and assume there are no cancellations and capacity constraints.

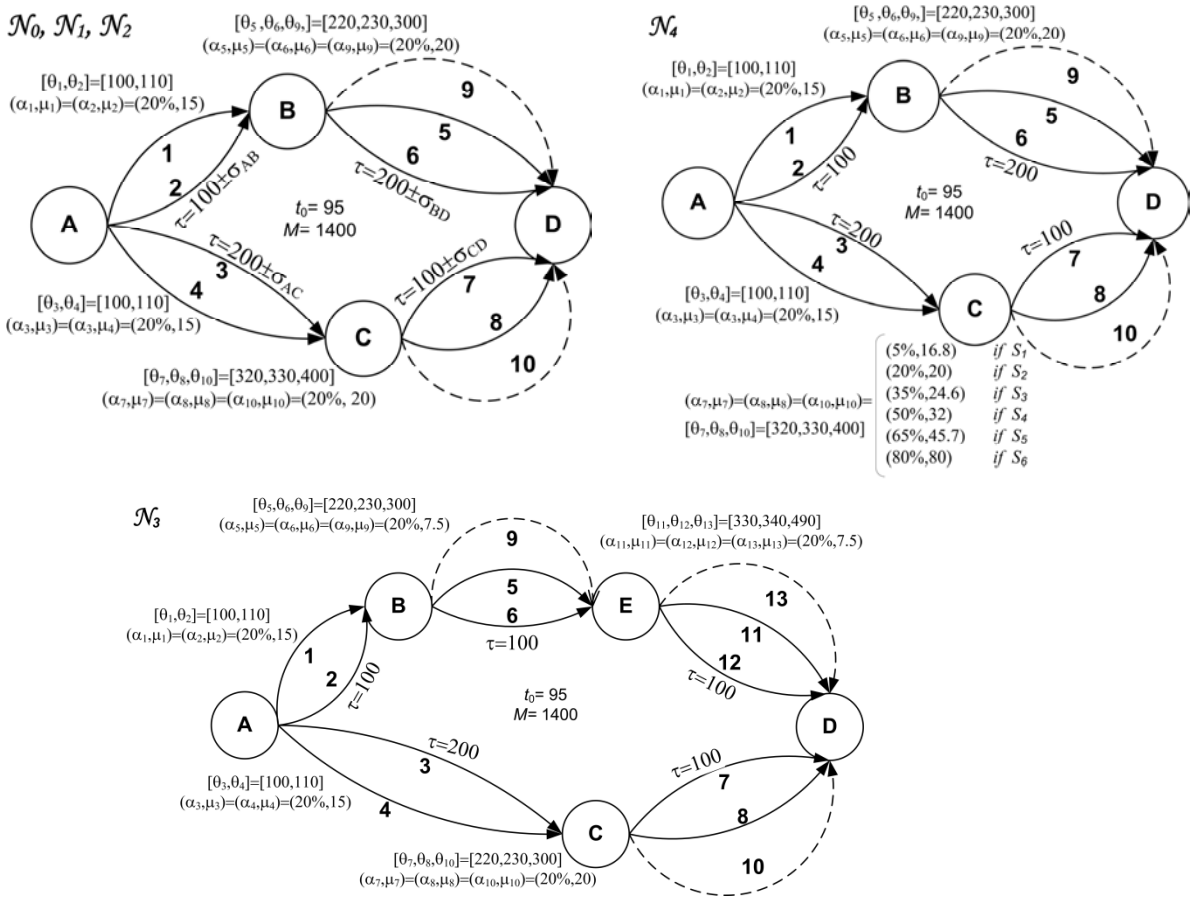


Figure 2.2. Network structure and parameters for five problem configurations ($\mathcal{N}_0$ to $\mathcal{N}_4$).

For each configuration, we first derive the *three routing policies* (dynamic, static and dynamic with perfect information) and then simulate each policy for all 20,000 delivery run samples. The static policy is a fixed flight path determined based on the expected departure delays; accordingly, the recourse flights are only selected if a flight in the path becomes unavailable. The dynamic policy under perfect information is determined by a priori knowledge of all realizations. We define measure (ρ) as dynamic policy's improvement over static policy as a percentage of total possible under perfect information.

$$\rho = 100 \times \frac{F_s^*(\Omega_1) - F^*(\Omega_1)}{F_s^*(\Omega_1) - F_{PI}^*}, \qquad (18.)$$

where $F_S^*(\Omega_1)$, $F^*(\Omega_1)$ and F_{PI}^* denotes the expected costs of the static policy, dynamic policy, and dynamic policy with perfect information, respectively. For configurations $\mathcal{N}_0$, $\mathcal{N}_3$, and $\mathcal{N}_4$, we sample only departure delays (actual and announced delay) and, for configurations $\mathcal{N}_1$ and $\mathcal{N}_2$, we also sample flight travel time (τ). For consistence, we use the same actualized and announced departure delay information in simulating the static policy, dynamic policy and dynamic policy with perfect information.

Figure 2.3 presents the distribution of flight paths for all problem configurations. $\mathcal{N}_0$ with *m*=1, dynamic policy is almost indifferent between *B* and *C* and tends to choose early flights out of airport *A*. With increasing *m*, the dynamic policy begins choosing secondary flights (e.g., flights *2* and *4*). This is attributable to the instances where the announced delay for early flight makes the secondary flight desirable. In contrast, the static policy commits to the flight path *(1,5)* connecting through node *B*. Whenever static policy misses the connecting flight *5*, it chooses the next flight out, e.g. either *6* or *9*. In cases of $\mathcal{N}_1$ and $\mathcal{N}_2$, we observe that the travel time variability notably affects the dynamic policy's path selection. For *m=1*, routes through *B* are more preferred in $\mathcal{N}_1$ since arrival time variability at node *B* is less than node *C* and therefore less chance of missing flights. With *m=2* and *m=5*, the dynamic policy begins selecting routes through *C* as it can now better manage the risk of missing flights departing from *C*; e.g. select *C* only if departing flights from *C* are delayed. The dynamic policy's route choice in $\mathcal{N}_2$ is similar to $\mathcal{N}_1$ except that later flights departing from *B* and *C* are selected more. Therefore, we conclude as the travel time variability increases, the dynamic policy selects similar routes as the static policy. For $\mathcal{N}_3$ with *m*=1, the flight path with the least number of connections (i.e., passing

through *C*) is preferred due to higher chance of getting on an early flight at *C* than *B*. With increased announcement accuracy, the dynamic policy sometimes chooses the most preferable path through *B*, which constitutes all the early flights departing from *B* and *E*. In comparison, the static policy commits to the flight path *(1,5,11)* connecting through the nodes *B* and *E*. However, whenever static policy misses the flights *5* at node *B* or *11* at node *E*, it chooses the next flight out, e.g. either *6* or *9* at node *B* and either 12 or 13 at node *E*.

In the case of $\mathcal{N}_4$, Figure 2.3 illustrates the effect of changing the delay distribution of all flights departing from airport *C* with *m*=2. These distributions share the same expected delay and S_2 is identical to $\mathcal{N}_0$ with *m*=2. As the expected value of delay distribution for delayed flights increases (or decreases) from that of S_2, the dynamic policy prefers the flight paths going through *C* more. The reason for preferring *C* more with $S_{3\text{-}6}$ is the availability of flight *8*, in essence, provides the dynamic policy a *truncation on the delay distribution* experienced by flight *7*. In summary, the choice of flight paths depends on the policy used. The static policy tradeoffs the tardiness of a fixed path with the risk of missing a connecting flight. On the other hand, the dynamic policy exploits both the real-time departure delay information (whenever available) and the multiplicity of flights departing from connecting airports.

Figure 2.4a indicates that the rate of improvement with increased accuracy is diminishing for $\mathcal{N}_0$ to $\mathcal{N}_3$. Further, the dynamic policy can achieve the majority of performance improvement even with some level of real-time delay information. An increase in travel time variability decreases the dynamic policy's performance improvement over the static alternative, e.g. $\mathcal{N}_0$ versus $\mathcal{N}_1$ and $\mathcal{N}_2$. The effect of delay distribution is illustrated in Figure 2.4b. For S_5, the

on-time departure probability is very high and there is some level of truncation of the experienced delay at the airport *C*, and thus dynamic policy is better performing than S_1. With increasing *m*, the dynamic policy's performance is increasing and is robust with respect to the delay distribution due to the truncation effect on the experienced delay in airport *C*.

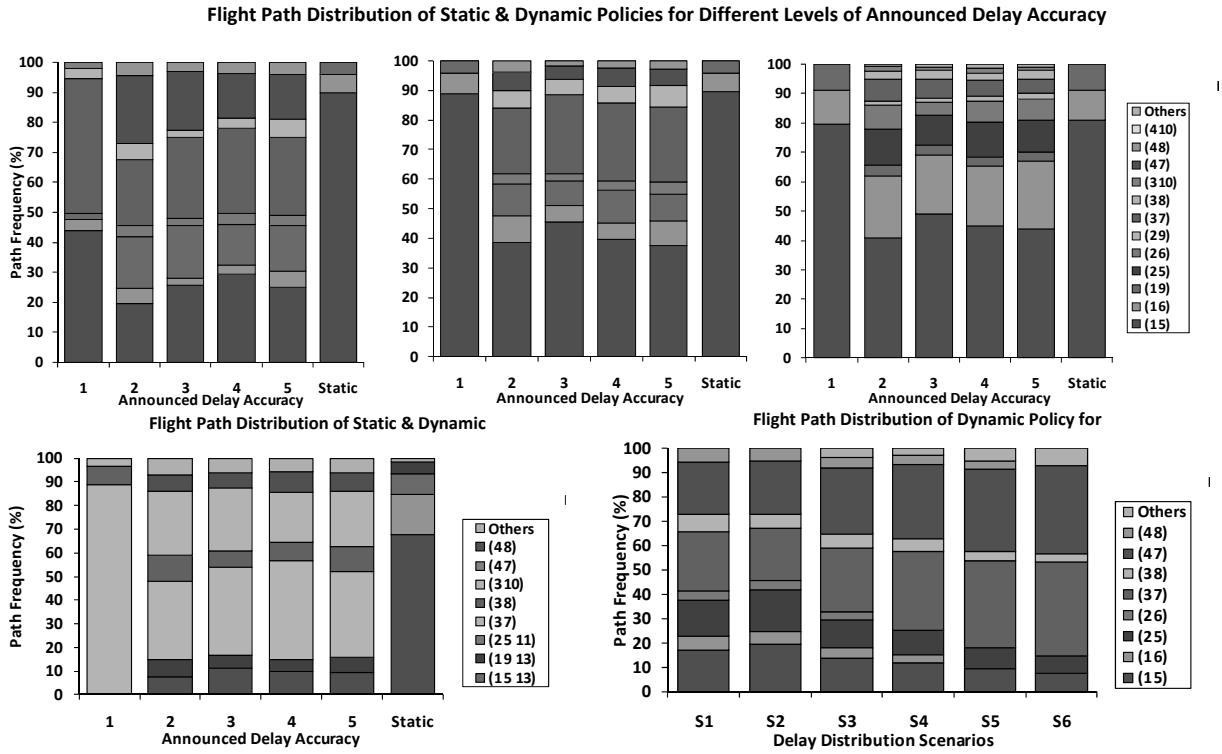


Figure 2.3. Flight path distributions of static and dynamic policies for different levels of announced delay accuracy ($\mathcal{N}_0$, $\mathcal{N}_3$), travel time variation ($\mathcal{N}_1$, $\mathcal{N}_2$), and different delay distributions ($\mathcal{N}_4$).

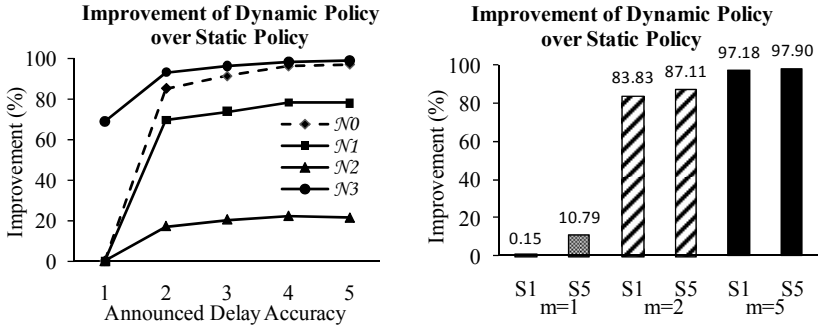


Figure 2.4. Improvement (ρ) of dynamic policy over static policy for $\mathcal{N}_0$, $\mathcal{N}_1$, $\mathcal{N}_2$, $\mathcal{N}_3$ (a) and $\mathcal{N}_4$ (b)

Another important performance measure for the shippers and freight forwarders is the delivery reliability, i.e., the percentage of shipments arriving on time. Figure 2.5 shows the conditional expected tardiness for $\mathcal{N}_0$, $\mathcal{N}_3$, and $\mathcal{N}_4$ at different levels of announced delay accuracy and delivery due dates. Figure 2.5a and 5b illustrate that increasing information accuracy improves tardiness performance. Case for $\mathcal{N}_3$ is similar to $\mathcal{N}_0$, but the difference in static and dynamic policy tardiness is more remarkable. In the case of $\mathcal{N}_4$, the conditional expected tardiness of two policies with no real-time information is similar and insensitive to the delay distribution (Figure 2.5c). Further, with the increased level of announced delay accuracy, the effect of the delay distribution on the conditional expected tardiness diminishes.

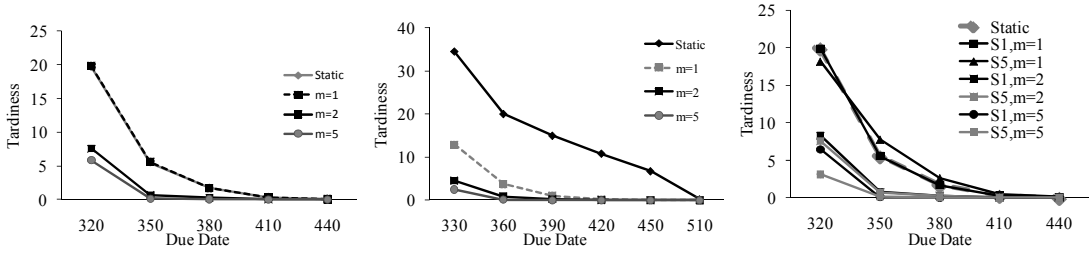


Figure 2.5. Conditional expected tardiness for different due date levels: (a)$\mathcal{N}_0$, (b)$\mathcal{N}_3$,and (c)$\mathcal{N}_4$.

2.5 Case Studies

In this section, we first briefly describe the estimation of delay and travel time distribution model parameters using real-world data sources. Next, we describe two case study applications and discuss their analysis results.

2.5.1 Estimation of Flight Departure Delay and Travel Time

The flight i is delayed with probability $(1-\alpha_i)$ and the corresponding departure delay (δ_i) has non-negative and continuous probability density $\psi(\delta_i)$ and cumulative density $\Psi(\delta_i)$. We estimate these probabilities using the publicly available historical databases. The departure delay depends on a number of factors such as seasonality (e.g. time of the day, etc.), origin and destination airports, carrier, weather, special days and other non-recurring events. The databases of the BTS and the OPSNET provide detailed multi-year historical on-time departure and departure delay information on all the US domestic flights and major US airports. The data in both the BTS and the OPSNET are either aggregated at the facility level or available only for the passenger flights. Since our routing model is applicable to both dedicated carriers as well as passenger carriers, we assume that departure delay for cargo carrying flight can be approximated with the delay data for mixed passenger/cargo flights. This assumption can be justified by considering the fact that in both cases the flights are affected by similar factors (Mueller and Chatterji 2002, Chatterji and Sridhar 2005). These delay data are extracted for each combination of the determining factors to estimate the most accurate non-parametric delay distribution for each flight. However, due to small sample sizes, we aggregated the data by selecting the *origin airport*, *destination airport*, *month of the year* and *time of the day* as the factors to be included. These factors are identified as most significant by conducting multiple analysis of variance tests.

In case study applications, we considered the month of June in 2009. First, we estimated the percentage of on-time departures (α_i) and the cancellation and diversion percentages (γ_i).

We considered those flights with delays in excess of $\xi = 90$ minutes as cancelled. After filtering out the on-time departures, cancellations and diversions from the data collected, we estimated the departure delay distributions. While any of the non-parametric estimation techniques are suitable, we considered various common distributions for presentation purposes. The goodness of fit tests of common distributions indicated that the distribution of departure delays follow exponential distribution which is also used by some of the earlier studies (Long et al. 1999, Hansen and Bolic 2001). We note that the proposed method is independent of the distribution, e.g. empirical or other common densities, such as Bi-Weibull in Tien et al. (2008), can be used if they provide better fit. Despite the aggregation over the statistically non-influential factors, the size of the data set was small for some flights and the goodness-of-fit tests were not conclusive. Accordingly, we further aggregated the data by using agglomerative hierarchical clustering to cluster the departure hours based on their average departure delay. In most cases, the clustering of the departure hours into two clusters is found satisfactory. Figure 2.6 illustrates the steps of this procedure for flights from the La Guardia Airport (LGA) to Chicago O'Hare International Airport (ORD) in June 2009. The hierarchical clustering identified two clusters: one cluster with lower departure delay means (hours in 6h00 to 16h00 except 11h00) and the other higher departure delay means (11h00 and hours in 17h00-20h00). Figure 2.7 illustrates the frequency plots of the data in two clusters of LGA-ORD flights.

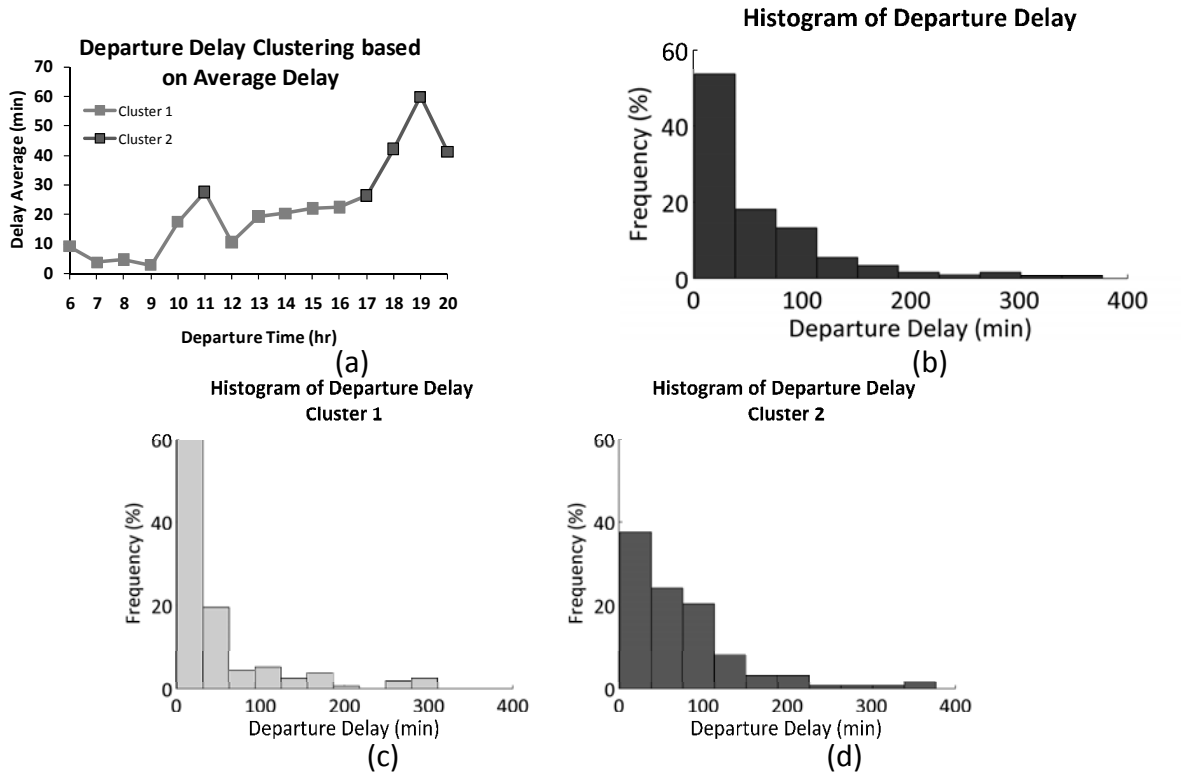


Figure 2.6. Departure hour clustering for LGA to ORD (a); departure delay frequency of LGA-ORD flights in June 2009 for all departure times (b); frequency plots for two clusters (c,d).

Using the same database, we also estimated the travel time distributions conditional on the departure delay. Figure 2.7 illustrates the joint and marginal distributions of LGA-ORD flights in June 2009 for the cluster with higher departure delay. For this particular cluster, the actual travel time and departure delay are found to be statistically independent. For those instances with significant dependence, we use conditional travel time distribution.

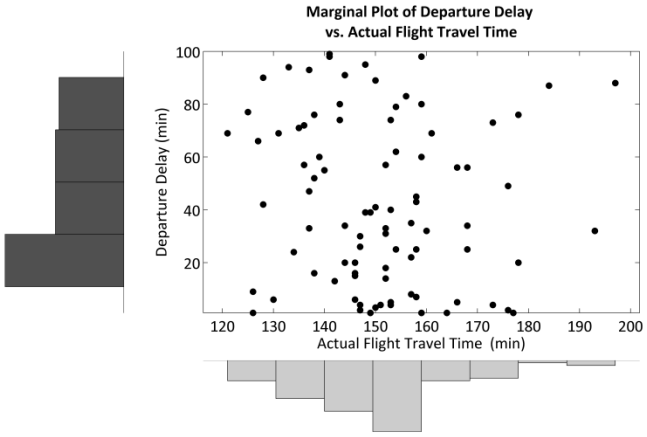


Figure 2.7. Marginal distribution of departure delay and travel time for LGA-ORD flights for cluster 2.

The departure delay announcement policies vary from carrier to carrier and from airport to airport. Since our goal in this study is to investigate the effect of dynamic routing using real-time delay information, we studied the case study problem for different availability and accuracy levels of delay information. The departure delay for each flight is generated according to the distributions estimated.

2.5.1.1 Case Study I: La Guardia Airport (LGA) to Seattle–Tacoma International Airport (SEA)

We consider the routing of the time-sensitive air-cargo from LGA to SEA at 6:00 a.m. on Friday June 12, 2009. We consider two air-carriers: Northwest airlines (NW) and American Airlines (AA).[7] Since there are no direct flights, we chose three potential connecting airports for this problem: ORD, DTW, and Minneapolis International Airport (MSP). We then established the air-network by extracting the data from the OPSNET and the BTS databases and estimated

[7] As January 2010, the Northwest Airlines merge to the Delta Airlines was completed; however, to be able to use the historical data, we evaluate the case study for June 2009.

problem parameters as shown in Table 2.1. The departure times are based on US Eastern Time. According to the extracted data, the distribution for departure delay of delayed flights is estimated by the exponential distribution. The flight travel times are found to be independent of the departure delay and their distributions are estimated by Gaussian distributions. We further assume that the freight forwarder can load onto the cargo connecting flights, e.g. no cargo space restriction.

Table 2.1. LGA to SEA time-sensitive air-cargo routing case study.

Flight Label	Carrier	Flight Num.	From	To	Sch. Dep.	Departure Delay		Duration (min)	
						Mean (min)	On-time (%)	Mean	Std. Dev.
1	AA	303	LGA	ORD	6:55	35	58	145	5
2	AA	305	LGA	ORD	7:29	41	48	149	15
3	NW	541	LGA	DTW	6:00	6	43	117	13
4	NW	533	LGA	DTW	7:25	21	60	117	10
5	NW	1424	LGA	MSP	6:00	28	51	182	12
6	NW	1646	LGA	MSP	8:15	26	53	185	16
7	AA	1843	ORD	SEA	9:30	19	56	263	12
8	AA	313	ORD	SEA	12:15	25	64	256	12
9	NW	211	DTW	SEA	9:33	44	53	296	13
10	NW	215	DTW	SEA	12:19	27	66	291	10
11	NW	167	MSP	SEA	12:45	18	61	224	11

We solved the air-cargo routing problem for different departure delay announcement polices, e.g. by varying m. For each m category, we generated 20,000 samples of flight departure delay and announced delay information for all flights in Table 2.1. We determined the routing solution for each sample using static policy, dynamic routing policy, dynamic routing with perfect information based on total trip time using delivery failure penalty *M*=1400 minutes. Figure 2.8a illustrates the routing choice of static and dynamic policies. With *m*=1, the static and dynamic policies prefer paths through Detroit (flights 1 then 7) and Chicago (flights 3 then 9), respectively. While the expected travel time of path (1-7) is less than path (3-9), the

chance of missing flight 7 is higher than missing flight 9. The dynamic policy trades off this risk in favor of shortest path. Consequently, it misses the flight 7 in about 25% of the time and continues by flight 8. However, as *m* increases, the dynamic policy mostly substitutes the path (1-8) with path (3-9) in such announcement scenarios where flight 7 is expected to be missed.

Figure 2.8 presents the distribution of the delivery times. For *m=1*, the static policy's single path choice leads to single mode distribution of delivery times. In comparison, the paths (1-7) and (1-8) corresponds to the two modes of the dynamic policy's distribution. The dynamic policy exploits the availability of departure delay information and chooses earlier but riskier flights. In comparison, the static policy chooses a path of flights with a high probability of being available. Figure 2.9a illustrates that the expected value of the static and dynamic policy distributions are identical for the case *m=1*. With increased announcement accuracy (*m=2*), we note that the dynamic policy's distribution shifts towards left as a result of choosing path (3-9) more than before. This corresponds to about 70% performance improvement over the static policy (Figure 2.9b). With *m=5*, the frequency of long trip durations is minimized and the performance improvement is about 86%. The ability to change the flight decisions online provides the dynamic policy the ability to choose the earlier flights with recourse options. Therefore, the dynamic policy is not only superior in the expected sense but can also provide early delivery performance which cannot be attained by a static policy. Whereas the earliest delivery for the static policy is at 822 minutes, the dynamic policy can attain deliveries as early as at 792 minutes.

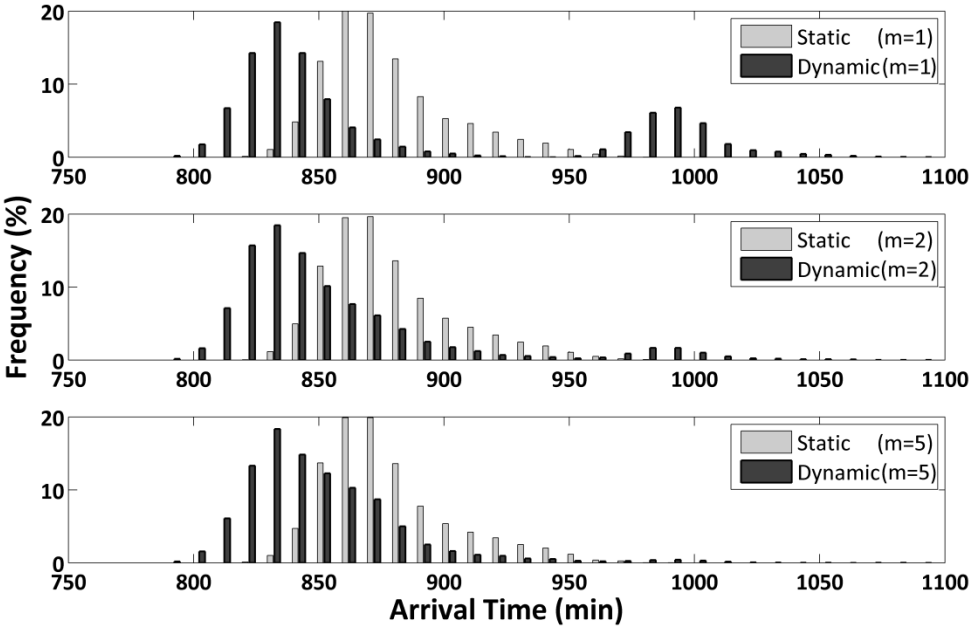
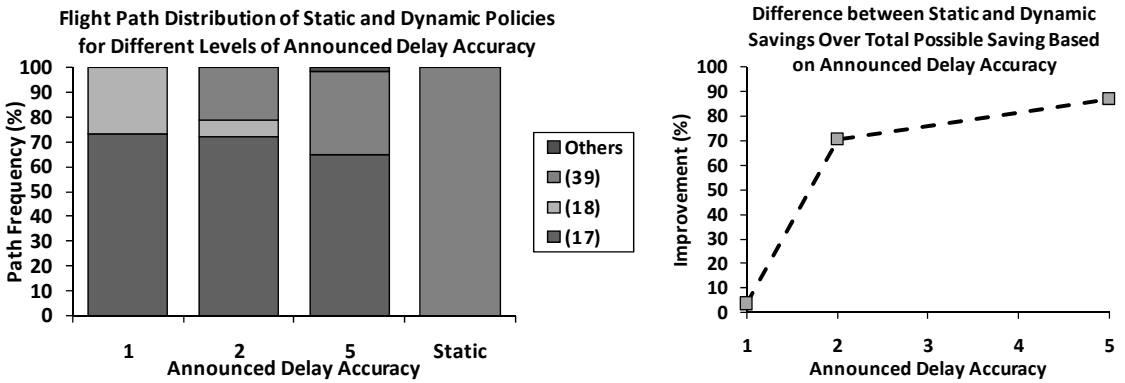


Figure 2.8. Travel time distributions for different announcement accuracy levels (LGA-SEA case study).



(a) (b)

Figure 2.9. LGA to SEA case study, flight path frequency (a) and improvement (b)

This case study illustrates that dynamic policy can significantly enhance the routing performance, especially when the information accuracy is high. It also illustrates that dynamic policy may not lead to significantly better results in the case of limited route alternatives and low information accuracy.

2.5.1.2 Case Study II: La Guardia (LGA) to Dallas Fort Worth International Airport (DFW)

This case study considers the routing of a time-sensitive air-cargo from LGA to DFW at 6:00 a.m. on Tuesday, June 16, 2009. We consider three air-carriers: US Airways (US), Delta Airlines (DL) and NW. Table 2.2 presents problem parameters and distributions.

Table 2.2. LGA to DFW time-sensitive air-cargo routing case study.

Flight Label	Carrier	Flight Num.	From	To	Sch. Dep.	Departure Delay		Duration (min)	
						Mean (min)	On-time (%)	Mean	Std. Dev.
1	DL	19	LGA	ATL	6:00	42	45	174	14
2	DL	533	LGA	ATL	7:00	36	77	170	13
3	DL	423	LGA	ATL	8:00	14	83	165	15
4	NW	541	LGA	DTW	6:00	6	43	117	6
5	NW	533	LGA	DTW	7:25	21	60	117	10
6	NW	172	LGA	MEM	6:30	14	84	163	13
7	NW	509	LGA	MEM	11:00	22	77	165	17
8	NW	1424	LGA	MSP	6:00	28	51	182	12
9	NW	525	LGA	MSP	8:15	26	53	185	16
10	US	1455	LGA	CLT	6:30	13	93	112	8
11	US	1021	LGA	CLT	8:10	18	93	115	18
12	DL	1915	ATL	DWF	8:40	15	62	141	9
13	DL	1917	ATL	DWF	9:45	8	41	138	10
14	DL	1919	ATL	DWF	11:00	16	60	139	15
15	DL	1941	ATL	DWF	12:25	28	68	145	35
16	NW	1193	DTW	DWF	9:31	15	50	171	13
17	NW	407	DTW	DWF	12:40	22	58	174	15
18	NW	1148	MEM	DWF	15:20	33	69	93	9
19	NW	1125	MSP	DWF	10:15	7	81	153	11
20	NW	1167	MSP	DWF	15:45	14	41	152	13
21	US	1772	CLT	DWF	9:25	13	47	159	24
22	US	1101	CLT	DWF	11:24	29	33	160	21

ATL: Atlanta International Airport; **MEM**: Memphis International Airport; **CLT**: Charlotte International Airport

Figure 2.10a shows the flight path frequency for the two policies. With *m*=1, the static and dynamic policies prefer paths through Atlanta (flights 1 then 12) and Charlotte (flights 10 then 21), respectively. The expected travel time of path (1-12) is less than path (10-21). However, the flights 12 and 13 are missed in Atlanta with recourse to flights 13 and 14. With more accurate information on departure delay, e.g. *m=2* and *m=5*, the dynamic policy reduces the missed flights 13 and 14 by following path (10-21) when advantageous. Figure 2.10b

illustrates that the expected performance improvement of dynamic routing policy over static policy. Clearly, even for limited information accuracy, dynamic can realize 20% performance improvement. However, unlike the previous case study, the upside potential of the improved accuracy is limited to 59%. This result demonstrates that even though the dynamic policy can provide significant benefit with limited information accuracy, its upside potential might be limited.

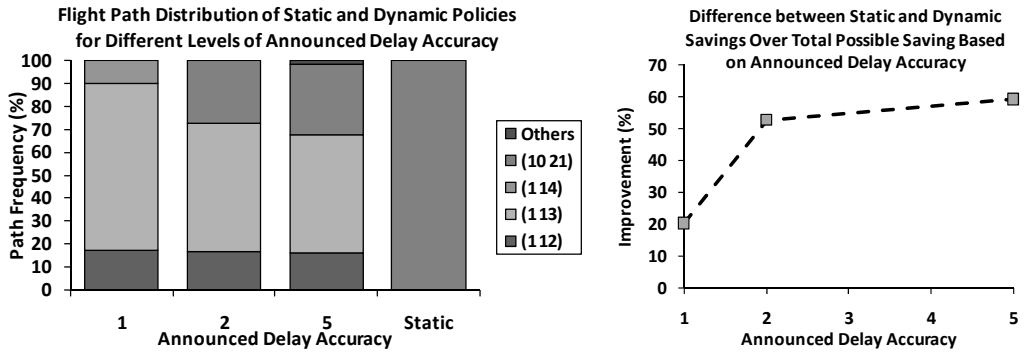


Figure 2.10. Flight path distribution (a) and improvement of dynamic policy over static policy (b).

Another advantage of the dynamic policy over the static policy is the reliability of delivery, a key performance metric in the air-cargo industry. Figure 2.11 shows the conditional tardiness as percentage of tardy deliveries and average total tardiness for different due dates. For *m=1*, while the average tardiness performances of both policies are similar, the dynamic can cut down the late deliveries by up to half for certain due dates, e.g. only 17% of deliveries are tardy with dynamic policy compared to 32% with static policy for due date at 737 minutes. However, this performance improvement fluctuates with different due dates and could be insignificant at certain due dates, e.g. at due dates 760 and 707. With increased information accuracy, the dynamic policy reduces the frequency of tardy deliveries and total average

tardiness, e.g. the percentage of late deliveries at 737 minutes is only 13% for dynamic policy with *m*=5.

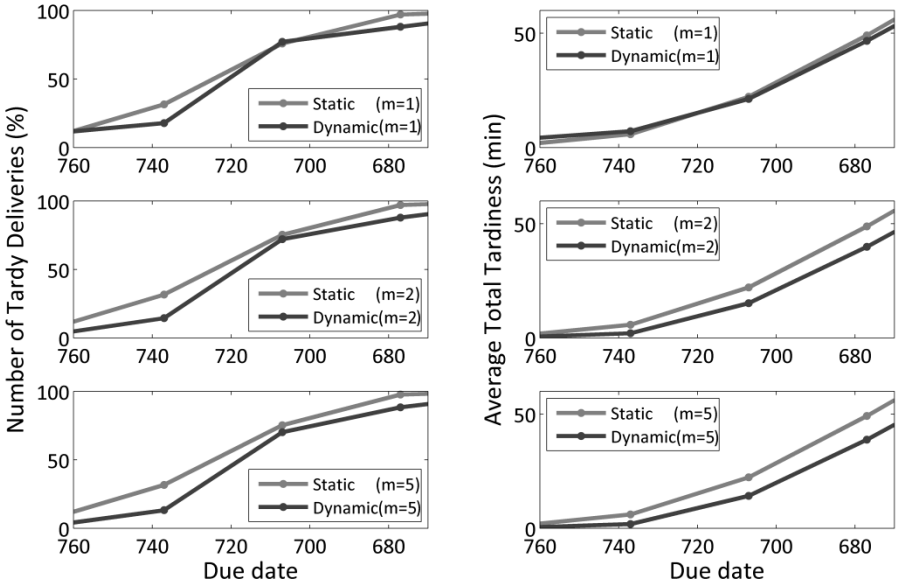


Figure 2.11. Conditional Tardiness: percentage of tardy deliveries (left) and average total tardiness (right)

2.6 Conclusions and future research

We studied the air-cargo routing problem from the freight forwarders perspective and investigated the benefits of dynamic routing for the shipment of time-sensitive air-cargo given a shipment criterion subject to the availability of flights and travel time variability. We further examined the effect of real-time flight information accuracy on the dynamic routing performance. The contributions of this paper to the literature are a novel dynamic routing model which accounts for the scheduled departures, the effect of stochastic travel times and departure delays and a novel departure delay estimation model based on the real-time announced delay information and historical delay distributions.

We developed a departure delay estimation approach for the dynamic air-cargo routing based on conditional probability models. The proposed delay estimation model accounts for the unavailability of a flight due to late arrival of the cargo and uses both historical and real-time departure delay information. We then formulated a dynamic routing Markov decision problem with a novel action space definition. The action space consists of not only the first flight choice but collectively all recourse flights at an airport node. A set of controlled experiments is conducted to investigate the effect of delay information accuracy, departure delay distribution, travel time variability and topology of flight network on the expected cost and delivery reliability. Lastly, we presented two case study applications using real flight network and departure delay data. The results show that dynamic policy is able to not only improve the expected delivery performance but also increase the delivery reliability. Further, the departure delay information is critical for realizing the full potential of dynamic routing. However, the majority of the improvements can be attained even with little real-time information availability and accuracy.

Chapter Three:

Air Cargo Pickup and Delivery Problem with Alternative Access Airports

3.1. Introduction

This paper considers a freight forwarder's problem of selecting air cargo flight itineraries to a given set of heterogeneous customers and, simultaneously, planning the pickup and airport delivery schedule of customer loads. The air cargo flight itinerary options for each customer consist of a set of flights departing from the origin airport(s) and arriving to the destination at different times. For each customer, the forwarder selects an itinerary considering flight and delivery service level related costs, such as tardiness penalties. Given the air cargo itinerary assignments, the forwarder performs the customer pickup and airport deliveries via a fleet of trucks originating from a depot. In this paper, we formulate and develop an efficient solution approach for freight forwarders to concurrently plan the air cargo flight itinerary selection and pickup and delivery scheduling of multiple customer loads to minimize the total cost of air and road transportation and service.

Over the past decade, there has been consistent growth in demand for air cargo deliveries. According to the Bureau of Transportation Statistics (BTS), in 2007, the value of air cargo shipment goods in the US is over $1.8 trillion, a 31% increase in just five years (Margreta et al., 2009). Annual forecast reports by both Airbus (2010) and Boeing (2010) predict a 5.9%

annual growth rate for global air cargo tonnage over the next 20 years. In response to this growth, the air transportation network has been steadily expanding its capacity over the past two decades. However, this capacity expansion through new airports, offering more flights options, and investing in road connectivity cause the service zones of airports to expand and overlap. This has resulted in the creation of *Multi-Airport Regions* (MARs) where several airports accessible in a region substitute and supplement each other in meeting the region's demand for air transportation (Loo, 2008). These MARs provide alternative access options for passengers as well as air cargo shippers and forwarders. For instance, air travelers consider MARs in a region and select airports and flights primarily based on airport access time, flight itinerary options, and frequency factors (Basar and Bhat, 2004). These factors are also important concerns for the air cargo transportation. The shippers are mainly concerned with the on-time delivery performance and the shipping costs, and thereby leave the flight itinerary decisions to forwarders. The freight forwarders, intermediaries between shippers and carriers, constitute more than 90% of air cargo shipments (Hellermann, 2006). In the case of MAR, the forwarders decide on which origin airport to use given the flight itinerary options and costs. Their decisions are primarily based on such factors as airport accessibility, proximity to the origin of the loads, flight itinerary options (e.g., frequency, destinations). Hall (2002) proposed the *Alternative Access Airport Policy* (AAAP) where considering multiple airports (and subsequently flight itinerary options) in a MAR can be beneficial to reduce truck mileage, decrease sorting and handling costs, improve delivery service level, and avoid congestion on both road and air network. The author discussed the merits of AAAP for air cargo transportation using the case study of the Southern California region.

In this paper, we consider a freight forwarder's operational implementation of AAAP for air cargo transportation. While Hall (2002) outlined and discussed the advantages of the AAAP, to the best of our knowledge, there is no study on its modeling and implementation. We model the forwarder's problem of selecting flight itineraries for a given set of air cargo customers, picking up their loads via a fleet of vehicles and then delivering to the airports in the region. One decision component in this problem is the flight itinerary assignment of the air cargo of different customers that are geographically dispersed in the MAR. These decisions are driven by the availability of flight itinerary options, cargo drop-off cutoff times, destination arrival times, flight itinerary costs, and tardiness penalties. The other decision component is the multi-vehicle routing to pick up customer loads and deliver to the airports prior to the starting time of the selected flight itineraries. These routing decisions are affected by the locations (depot, customers and airports), starting times of the selected flight itineraries, and the vehicle fleet size. This operational implementation of AAAP generalizes the Many-to-Many Pickup and Delivery Problems (M-M-PDP) in several aspects. For instance, the delivery cost of customer air cargo is both destination and time dependent. We hereafter refer to this problem as *PDP with Assignment and Time-Dependent delivery cost (ATD-PDP)*. Our contribution in this research is three fold. First, we model the operational implementation of AAAP for freight forwarders which generalizes several known pickup and delivery problems in terms of model structure and objective function. Second, we develop a novel and highly efficient solution method based on the Lagrangian decomposition. Finally, we present the results of a case study implementation of the AAAP in a Southern California MAR.

The rest of this paper is organized as follows. We briefly describe the relevant literature

in Section 2. In Section 3, we present the problem formulation, network transformation and preprocessing. The solution method is developed and properties such as convergence are discussed in Section 4. In Section 5, we report on the results of the computational study with experimental problem instances and a case study implementation. Section 6 concludes with discussion and future research directions.

3.2 Related Literature

The freight forwarder's operational implementation of the AAAP is closely related to the pickup and delivery problem. *Pickup and delivery problems* have been extensively studied in past decades; for a comprehensive survey see (Berbeglia et al., 2007; Berbeglia et al., 2010; Laporte, 1992, 2009; Parragh et al., 2008a, b; Toth and Vigo, 2001). Generally, the PDP involves routing a fleet of vehicles to satisfy a set of transportation requests between the given origins and destinations. In the PDP, all the origin pickups must precede the destination deliveries and be performed by the same vehicle. Moreover, each route must start and terminate at the same location (i.e., depot). The PDP usually considers capacitated vehicles and the goal is to minimize criteria related to a travel measure. The travel measure can be as simple as the total travel distance for urban commercial vehicles (Miguel Andres, 2007) or more complex as the total excess riding time over the direct ride time in passenger transportation (Diana and Dessouky, 2004). The PDP can be classified into two categories: transportation between customers and the depot, and transportation between the pickup and delivery locations (Parragh et al. 2008a). The proposed problem is in the latter category, which can be further classified into paired and unpaired pickup and delivery locations.

In the paired PDP, also known as *One-to-One* PDP (1-1-PDP), the load picked up from a customer location can only be delivered to one of the delivery locations. Some customers, however, may share the same delivery location. In the *stacker-crane problem* (SCP), unit loads of non-identical commodities have to be transported from the origin to the destination using a unit capacity vehicle (see Frederickson, 1978). In the *Vehicle Routing Problem with Pickup and Delivery* (VRPPD), the unit capacity requirement of SCP is relaxed and replaced with a set of constraints based on the load properties (e.g., weight, volume, or unit count). A special case of the VRPPD is the *VRPPD with Time Windows* (VRPPDTW) where visiting the pickup or delivery location is only allowed during a time window. While the VRPPD generally concerns goods transportation, the *dial-a-ride problem* (DARP) addresses the passenger transportation and therefore includes additional side constraints (e.g., maximum ride time, time windows, or service quality). Accordingly, the objective function measures customers (in)convenience; see (Cordeau and Laporte, 2007) for a comprehensive survey on the modeling and solution algorithms for DARP.

In comparison, the unpaired PDPs, also known as *Many-to-Many* PDP problems (M-M-PDP), consider the case where any commodity can be picked up and delivered to delivery locations that accept the commodity. The M-M-PDP was initiated with Anily and Hassin (1992) that introduced the *swapping problem* (SP) for moving *n*-commodity objects between customers with a single unit capacity vehicle. In the SP, each customer supplies one type of commodity and demands a different type. In addition to the *n*-commodity case of the SP, there are several other single commodity problems that are studied under the M-M-PDP where picked up loads are homogenous. Hernandez-Perez and Salazar-Gonzalez (2004a, b, 2007)

introduced and studied the *one-commodity pickup and delivery traveling salesman problem* (1-PDTSP). The 1-PDTSP is the more general case of the *Q-delivery traveling salesman problem* (Q-DTSP) by Chalasani and Motwani (1999) and the *capacitated traveling salesman problem with pickup and deliveries* (CTSPPD) by Anily and Bramel (1999). In the 1-PDTSP, a single vehicle, starting from a depot, transports goods from pickup nodes to delivery nodes without exceeding the vehicle capacity; the objective is to minimize the total traveling cost. Q-DTSP and CTSPPD are special cases of 1-PDTSP where the pickup and deliver quantities are all one unit and the vehicle capacity is restricted (i.e., Q units). Hernandez-Perez and Salazar-Gonzalez (2009) later extend their 1-PDTSP to the *Multi-Commodity One-to-One Pickup and Delivery Traveling Salesman Problem* (m-PDTSP); however, with this extension, the problem is not an M-M-PDP anymore.

The proposed problem is essentially a PDP as it consists of transporting loads from customer sites (pickup locations) to the airports (delivery locations) in the MAR. The depot is both the origin and destination of the vehicles; however, it is neither pickup nor a delivery point. The proposed problem differs from the 1-1-PDP in that a customer load can be accepted by more than a single delivery location (airport). Further, it differs from the general M-M-PDP in that the delivery cost of customer loads is time and destination dependent. Moreover, the delivery cost structure is different than those proposed for PDPs. Accordingly, we denote this problem as the PDP with Assignment and Time-Dependent delivery cost (ATD-PDP). The use of term " assignment" indicates that the delivery cost of a customer's load depends on the airport and flight itinerary selected. The proposed problem's characteristics have not been studied in the literature and, to the best of our knowledge, this is the first research on PDPs with

assignment and time dependent delivery costs. The proposed problem is clearly an *NP*-hard problem in the strong sense as it coincides with the VRPPD when there is only one airport and a single itinerary (accepted by all customers), which departs late enough to complete all pickups and delivery to the airport prior to the departure.

3.3 Model Formulation

In this section, we develop the model formulation of the ATD-PDP. We first discuss the time dependent delivery cost. Next, we describe the graph transformation and present the mixed integer programming model formulation. Last, we introduce and discuss pre-processing steps and valid inequalities to strengthen the formulation.

Let $G_o = (V_o, E_o)$ be an undirected graph representing the network topology of the problem where V_o is the set of nodes and E_o is the set of connecting edges. The set V_o consists of the depot d, the set of customers (pickup locations) C, and the set of airports (delivery locations) H; i.e., $V_o = \{d\} \cup C \cup H$. Let K be the set of uncapacitated homogeneous vehicles (trucks) that originate from the depot and operate during the depot's opening (θ_d^{op}) and closing hours (θ_d^{cl}). A cost c_{ij} and a travel time t_{ij} is associated with each edge $\forall (i,j) \in E_o,\ i \neq j$ of the network, where $c_{ij} \geq 0$ and $t_{ij} \geq 0$. We assume that the triangle property holds for the travel times and travel costs; i.e., we have $t_{ij} + t_{jg} \geq t_{ig}$ and $c_{ij} + c_{jg} \geq c_{ig},\ \forall i,j,g \in V_o$. Note that, if needed, the fixed cost of utilizing a vehicle can be captured by adjusting the cost parameter $c_{dj},\ \forall j \in C \cup H$. We further assume that travel time t_{ij} is deterministic and time-independent. Without loss of generality, we assume that there are no time windows for customers' pickups and the service (e.g., loading and unloading)

times are negligible. The formulation can be easily extended to incorporate these considerations as the methods presented do not rely on their absence. Let R_h be the set of flight itinerary options available at airport $h \in H$ on the day of operation. The cost of assigning a flight itinerary $r \in R_h$ to customer i is F_{ir}^h, which accounts for the flight cost of the carrier as well as the delivery service level related costs, such as tardiness penalties. The starting time of the flight itinerary r is denoted by Q_r^h (i.e., the cargo drop-off cutoff time for the first flight of the itinerary). We only consider those flights that can be used on the day of operation, e.g., $Q_r^h \geq \theta_d^{op}$.

3.3.1 Time Dependent Delivery Cost

In assigning the customer i's cargo to a flight itinerary $r \in R_h$, the freight forwarder accounts for the airport h arrival time. The assignment is feasible only if the airport delivery time (t) is on or before the flight itinerary starting time, i.e., $t \leq Q_r^h$. When a customer's load is delivered to an airport h at time t and there are no flights available, $t > \max_{r \in R_h}\{Q_r^h\}$, then the air cargo is assigned to a recourse flight itinerary $r_0 \notin R_h$, e.g., a next day itinerary. We assign a penalty cost $F_{i0}^h > F_{ir}^h \ \forall r \in R_h$, for airport delivery after the departure time of the last flight on the day of operation. Accordingly, we define the time dependent *airport delivery cost* of delivering customer i's load to airport h at time t, $f(h,i,t)$, as follows:

$$f(h,i,t) = \begin{cases} \min\limits_{r \in R_h}\{F_{ir}^h | t \leq Q_r^h\} & if \ \ t \leq \max\limits_{r \in R_h}\{Q_r^h\} \\ F_{i0}^h, & otherwise \end{cases}$$

The definition above indicates that for each customer, not all the itinerary options need to be considered and we can identify the potential set of itinerary options that are dominated

by at least another itinerary option from the same airport. The flight itineraries that are dominated for all customers are removed from further consideration. The flights itineraries that are dominated only for a subset of customers are preprocessed such that their assignment to that subset of customers is precluded. Lemma 1 provides the conditions necessary to identify the dominated flight itineraries from airport h for customer i.

Lemma 1. Given two flight itineraries $r, r' \in R_h,\ r \neq r'$, itinerary r is dominated by itinerary r' if (a) $F^h_{ir'} \leq F^h_{ir}$ and $Q^h_r < Q^h_{r'}$ or (b) $F^h_{ir'} < F^h_{ir}$ and $Q^h_r \leq Q^h_{r'}$. Moreover, if (c) $F^h_{ir'} = F^h_{ir}$ and $Q^h_{r'} = Q^h_r$, considering either one is sufficient.

Proof. The proof is evident from the definition of $f(h,i,t)$.■

Upon the elimination of dominated itineraries, the following corollary states that there exist no two flight itineraries for customer i at airport h that either depart at the same time or have the same cost.

Corollary 1. After eliminating the dominated flight itineraries, there are no two flight itineraries such that $r, r' \in R_h,\ r \neq r'$ in $f(h,i,t)$ with $Q^h_{r'} = Q^h_r$ or $F^h_{ir'} = F^h_{ir}$.

Theorem 1 characterizes the airport delivery cost function after eliminating the dominated itineraries.

Theorem 1. Airport delivery cost function $f(h,i,t)$ based on non-dominated flight itineraries is a non-decreasing step function with discontinuities at every $Q^h_r\ \forall r \in R_h$.

Proof. Let us first consider single flight itinerary case where $f(h,i,t) = F^h_{ir},\quad \forall t \leq Q^h_r$ and F^h_{i0} otherwise. Since $F^h_{i0} > F^h_{ir}$, the $f(h,i,t)$ is a step-function, which is non-decreasing and has a single discontinuity at Q^h_r. In the case of more than one flight itinerary, let us consider any two itineraries $r, r' \in R_h$. From Lemma 1 and Corollary 1, we have $Q^h_{r'} < Q^h_r$

and $F_{ir'}^h < F_{ir}^h$. Therefore, for any two delivery times t_1 and t_2 where $t_1 < t_2$, we have $f(h,i,t_1) \leq f(h,i,t_2)$. In this case, $f(h,i,t)$ is a non-decreasing step-function with discontinuities at Q_r^h and $Q_{r'}^h$. The case for more than two itineraries follows from the induction. Thus, the airport delivery cost function is a non-decreasing step-function with discontinuities at the starting times of the non-dominated itineraries. ■

Figure 3.1 illustrates a typical airport delivery cost function at airport h for two customers $i, j \in C$. There are two flight itinerary options available $r = 1$ and 2. While customer i can use both $r = 1$ and 2, customer j can only use the flight itinerary $r = 1$ and its load cannot be shipped by itinerary $r = 2$, e.g., destination of itinerary $r = 2$ is different than the customer j's destination. Note that airport delivery after Q_2^h for customer i (Q_1^h for customer j) will result in the penalty cost of F_{i0}^h (F_{j0}^h for customer j).

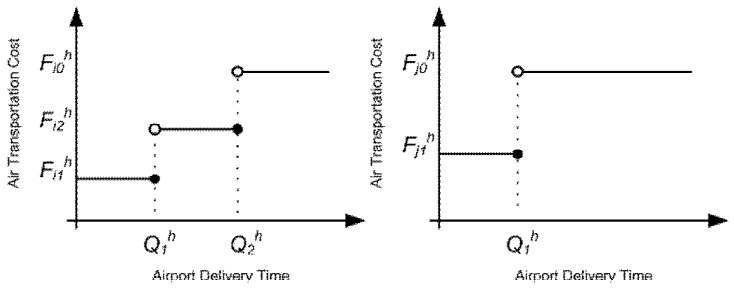


Figure 3.1. Illustrative airport h delivery cost function for customers i, $j \in C$; customer i has two flight itinerary options (left) and customer j has a single flight itinerary option (right).

Given the time-independent and deterministic edge travel times, we can infer the following two corollaries from Theorem 1.

Corollary 2. Waiting at any node or delaying any airport delivery is suboptimal for ATD-PDP.

Corollary 3. All used vehicles start at their earliest time from the depot.

3.3.2 Graph Transformation

In ATD-PDP, each airport can be visited multiple times by a vehicle to deliver loads from different customers. Consequently, a feasible solution may not be a Hamiltonian cycle. Hence, in modeling the ATD-PDP, we need to keep track of the order of these airport visits for each vehicle by introducing additional variables. Moreover, another set of additional variables is needed to handle the step-function characteristic of the airport delivery cost. To reduce the complexity of the ATD-PDP's formulation and eliminate the need for these additional variables, we perform a graph transformation of the original network graph $G_o = (V_o, E_o)$.

We now describe important properties of the optimal solutions of ATD-PDP used in the graph transformation. The first property relates to preemption, which is the act of temporarily leaving the previously picked up load at a location that is not its destination for retrieving it for delivery at a later time.

Lemma 2. There is an optimal solution for ATD-PDP that is non-dominated by a solution with preemption.

Proof. First, vehicles have no capacity restrictions to motivate preemptive solutions. In addition, a preemptive solution potentially prolongs deliveries by introducing additional node visits, shown to be suboptimal in Corollary 2. For any solution with preemption, we can identify a similar solution without preemption where the return visit for picking up the dropped load is eliminated while the remainder of the decisions remains the same. Since this elimination does not increase the airport arrival time, then, from Theorem 1, the non-preemptive solution has

same or better objective function than that of the preemptive solution.■

Corollary 4. In ATD-PDP, there is an optimal solution where all customer nodes are visited at most once.

Based on the above corollary, we can restrict the visit of each customer to at most once. The airport nodes, in contrast, can be visited more than once by each vehicle. However, the following theorem establishes that each vehicle visits an airport only once for each itinerary.

Theorem 2. There exists an optimal solution of ATD-PDP where each vehicle delivers customers' load to an airport for each flight itinerary only once.

Proof. Consider an optimal solution in which a set of customers (S) are assigned to a given flight itinerary r' at airport h. Assume that these customers are delivered to the airport by one vehicle but in two visits at times t_1 and t_2 consecutively, where $t_1 < t_2$. Clearly, $t_1 < t_2 \leq Q_{r'}^h$. Let us denote the set of delivered customers at each visit as two distinctive and non-empty sets of S_1 and S_2 respectively; i.e., $S_1 \cup S_2$. To prove the theorem it is sufficient to show that moving all the customers in set S_1 to set S_2 will results in a feasible solution with the objective value the same as the optimal objective value. First, since set S_2 is not empty and vehicles are uncapacitated, the proposed solution is feasible. Moreover, since the same itinerary is used the objective value is the same as the original optimal value. In other words, although the vehicle may still visit the airport at time t_1 for other itineraries, since S_1 is empty in the proposed solution, itinerary r' is used only once in the second visit.■

Theorem 2 states that we can restrict the solution of ATD-PDP to those solutions

where each flight itinerary requires at most one visit to the airport. The following corollary establishes that we only need to consider visits to an airport h equal to the number of flight itineraries $\forall r \in R_h$ plus an additional visit for the recourse flight $r_0 \notin R_h$.

Corollary 5. In ATD-PDP, there is an optimal solution where any airport h is visited, at most, $|R_h| + 1$ times.

We use this property to perform the graph transformation. In our graph transformation scheme, we partition each airport node h into $|R_h| + 1$ nodes, each node representing a single flight itinerary. In the remainder, we refer to these nodes as *flight nodes*.

Let $G = (V, E)$ be the transformed graph of the original graph$G_o = (V_o, E_o)$. In this transformation, each airport node $h \in H$ is replaced by $|R_h| + 1$ flight nodes, $|R_h|$ nodes each corresponding to a flight itinerary plus another node for the recourse flight. Consequently, the airport set H is replaced with a new set of flight nodes $r \in R$, where $|R| = \sum_{h \in H} |R_h| + |H|$. The geographical locations of the flight nodes are identical to that of their respective airport nodes. Then, we have $V = \{d\} \cup C \cup R$. The cost of assigning flight itinerary $r \in R_h$ to customer i (F_{ir}^h) is replaced with the delivery cost (F_{ir}) to flight node $r \in R$. Note that we are using the same index r for itineraries and flight nodes. Further, we introduce a hard upper time window Q_r for flight node r, i.e., it cannot be visited after Q_r. The flight node for recourse flights has the delivery cost of F_{i0} and upper time window of infinity.

As for the edges, we replace the airport $\forall h \in H$ edges $\forall (j, h) \in E_o$, $\forall j \in V_o \backslash \{h\}$ with new flight node edges $(j, r) \in$, $\forall j \in V, \forall r \in R_h$ and assign edge travel times $t_{jr} = t_{jh}$ and costs $c_{jr} = c_{jh}$. Similar procedure is repeated for the outgoing links. In addition, a new set of links interconnecting the flight nodes are added with zero travel time and cost for the flight

nodes generated from the same airport. The transformed graph $G = (V, E)$ inherits all the edges connecting depot to customers and customers to customers.

While a feasible solution in the original graph may not be a Hamiltonian cycle, the same solution is represented with one or more Hamiltonian cycles on the sub-graphs of the transformed graph. Indeed, any solution in graph G can be easily transferred back to a solution in original graph G_o by collapsing the flight nodes back to their original airport node. Figure 3.2 illustrates the graph transformation on a network with 5 customers and 2 airports, each with 2 flights. In the feasible solution illustrated in Figure 3.2a, loads from customer(s) $\{1\}$,$\{2,3,4\}$, and $\{5\}$ are assigned to flight itineraries $r2$ at airport $H1$, $r3$ at airport $H2$, and $r4$ at airport $H2$, respectively. While vehicle 1's trip is a Hamiltonian cycle, vehicle 2 visits the airport $H2$ twice. In the transformed graph in Figure 3.2b, this solution is represented in a single Hamiltonian cycle as vehicle 1 visiting flight node $r1$ and vehicle 2 visiting flight node $r3$ and then subsequently $r4$. In Figure 3.2b, the shaded flight nodes correspond to flight nodes for recourse flights.

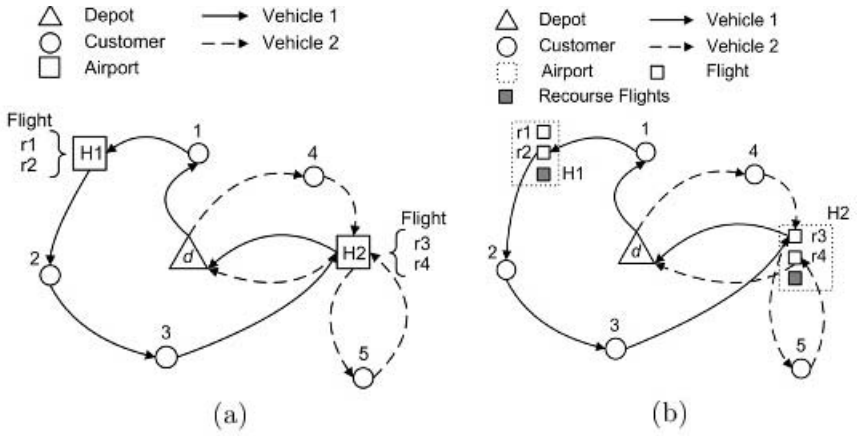


Figure 3.2. Illustration of a sample feasible solution in the original (a) and transformed (b) graphs.

The graph transformation eliminates the need for additional variables for tracking the

order of vehicle visits to airports as well as handling the step-function characteristic of the time dependent delivery cost. This transformation further reduces the complexity of the ATD-PDP's formulation. In particular, it allows network preprocessing and introducing valid inequalities to strengthen the formulation as described in Section 3.4.

3.3.3 Formulation

The objective of the ATD-PDP is to pick up all customer loads, assign loads to flight itineraries and deliver loads to the airports on time while minimizing the total cost. We now formulate the ATD-PDP using the transformed graph as a mixed-integer programming model. Let x_{ij}^k denote the binary decision variable indicating whether vehicle k travels from node i directly to node j. Let y_{ir}^k be the binary decision variable indicating whether the load of customer i is shipped by flight itinerary $r \in R$ with vehicle $k \in K$. The arrival time of the vehicle k at node $j \in V$ is denoted as a_j^k. For the depot, we set $a_d^k = \theta_d^{op}$ for any vehicle $k \in K$. The formulation of the ATD-PDP, labeled (MP), is as follows.

$$(\text{MP}) \qquad z_{MP}^* = \min_{x,y} \sum_{k\in K} \left[\sum_{i\in V} \sum_{j\in V\setminus\{i\}} c_{ij} x_{ij}^k + \sum_{i\in C} \sum_{r\in R} F_{ir} y_{ir}^k\right] \qquad (1)$$

Subject to

$$\sum_{k\in K} \sum_{r\in R} y_{ir}^k = 1 \qquad \forall i \in C \qquad (2)$$

$$\sum_{j\in V\setminus\{i\}} x_{ij}^k \le 1 \qquad \forall i \in V,\ \forall k \in K \qquad (3)$$

$$\sum_{i\in V\setminus\{j\}} x_{ij}^k + \sum_{i\in V\setminus\{j\}} x_{ji}^k = 0 \qquad \forall j \in V,\ \forall k \in K \qquad (4)$$

$$\left(x_{ij}^k - 1\right)M + a_i^k + t_{ij} \le a_j^k \qquad \forall i \in V,\ \forall j \in V\setminus\{d, i\},\ \forall k \in K \qquad (5)$$

$$a_i^k \ge \theta_d^{op} \qquad \forall i \in C,\ \forall k \in K \qquad (6)$$

$$a_r^k \leq min\{\theta_d^{cl} - t_{rd}, Q_r\} \qquad \forall r \in R,\ \forall k \in K \tag{7}$$

$$(y_{ir}^k - 1)M + a_i^k + t_{ir} \leq a_r^k \qquad \forall i \in C,\ \forall r \in R,\ \forall k \in K \tag{8}$$

$$2y_{ir}^k \leq \sum_{j \in V \setminus \{i\}} x_{ij}^k + \sum_{j \in V \setminus \{r\}} x_{rj}^k \qquad \forall i \in C,\ \forall r \in R,\ \forall k \in K \tag{9}$$

$$y_{ir}^k, x_{ij}^k \in \{0,1\} \quad a_i^k, a_r^k \geq 0 \qquad \forall i, j \in V | i \neq j, \quad \forall r \in R, \quad \forall k \in K \tag{10}$$

The objective (1) minimizes the total cost of delivery including flight itineraries, service level and road travel cost. Constraint set (2) ensures that every customer's load is assigned to a flight itinerary. Constraint set (3) guarantees that each node is visited at most once by each vehicle. Constraint set (4) is the flow conservation at each node for each vehicle. Constraint sets (5) and (6) calculate the arrival time at every node while also preventing sub-tours. Constraint set (7) prohibits visiting a flight node after the starting time of the flight itinerary while ensuring that the vehicle can also return to the depot before the depot's closing time. Constraint set (8) guarantees that a customer load pickup precedes its delivery to the selected flight node. Constraint set (9) ensures that both pickup and delivery of a customer load is performed by a same vehicle. Constant M is a big number corresponding to arrival times and can be calculated as summation of all the links' travel times.

For brevity, let $J_k(x, y)$ denote the objective function for vehicle k and $J(x, y)$ denote objective for all vehicles.

$$J_k(x, y) = \sum_{i \in V} \sum_{j \in V \setminus \{i\}} c_{ij} x_{ij}^k + \sum_{i \in C} \sum_{r \in R} F_{ir} y_{ir}^k \qquad \forall k \in K \tag{11}$$

$$J(x, y) = \sum_{k \in K} J_k(x, y) \tag{12}$$

3.3.4 Network Preprocessing and Valid Inequalities

We strengthen the formulation (MP) by network preprocessing and introducing valid inequalities. First preprocessing step is to tighten the upper and lower bounds on the node arrival times. For a customer node, the earliest arrival time ($\underline{a}_i^k$) is attained via a direct travel from the depot,

$$a_i^k \geq \ \underline{a}_i^k = \theta_d^{op} + t_{di} \quad \forall i \in C, \quad \forall k \in K, \tag{13}$$

and, the latest arrival time ($\overline{a}_i^k$) is the latest time that allows a vehicle to pick up the customer load, deliver it to a flight node, and return to the depot before the closing time,

$$a_i^k \leq \overline{a}_i^k = \max_{r \in R}\{\min(Q_r, \theta_d^{cl} - t_{rd}) - t_{ir}\ \} \qquad \forall i \in C,\ \forall k \in K. \tag{14}$$

For a flight node, the earliest arrival time ($\underline{a}_r^k$) is attained by the shortest travel from the depot after visiting a customer,

$$a_r^k \geq \ \underline{a}_r^k = \theta_d^{op} + \min_{i \in C}(t_{di} + t_{ir}) \qquad \forall r \in R, \quad \forall k \in K. \tag{15}$$

The latest arrival time to a flight node ($\overline{a}_r^k$) is already included in (MP) as the constraint set (7). Next preprocessing step is determining the lowest value for M in constraints (5) and (8). In particular, we replace M with edge specific M_{ij},

$$M_{ij} = \overline{a}_i^k + t_{ij} \qquad \forall i \in V,\ \forall j \in V \backslash \{d, i\},\ \forall k \in K. \tag{16}$$

The final preprocessing step is the elimination of the inadmissible edges that can never be traversed in a feasible solution. We remove edges $(i, d)\ \forall i \in C$ since vehicles must return to the depot empty. The edges $(d, r)\ \forall r \in R$ are eliminated since vehicles leave the depot empty. Lastly, we remove any edge $(i, j)\ \forall i, j \in V$ if $\underline{a}_i^k + t_{ij} > \overline{a}_j^k$, i.e., the vehicle cannot traverse an edge if it cannot arrive its destination before the latest allowed arrival time.

We tighten the constraints set (5) by using the following lifting scheme from Desrochers and Laporte (1991) by taking the reverse arcs into account.

$$x_{ji}^{k}\left(M - t_{ij} - t_{ji}\right) + \left(x_{ij}^{k} - 1\right)M + a_i^k + t_{ij} \leq a_j^k \quad \forall i, j \neq i \in V\backslash\{d\}, \forall k \in K \qquad (17)$$

In addition, we introduce the following cut set that ensures that a vehicle visits a customer only if it delivers the customer's load to a flight node.

$$\sum_{j \in V\backslash\{i\}} x_{ij}^{k} = \sum_{r \in R} y_{ir}^{k} \quad \forall i \in C, \quad \forall k \in K \qquad (18)$$

3.4 Methodology

First, we briefly present the standard Lagrangian Decomposition approach. Next, we introduce the Successive Subproblem Solution (SSS) method for solving ATD-PDP problem using the (MP) formulation with preprocessing and valid inequalities described in Section 3.4. We also provide convergence results and a method to estimate the bound used in subgradient optimization to improve the convergence to quality primal feasible solutions.

3.4.1 Standard Lagrangian Decomposition Approach

The standard Lagrangian Decomposition (LD) approach is commonly used for formulations composed of two or more intertwined subproblems that are easier to solve independently through specialized algorithms. In fact, the LD approach is commonly used for the vehicle routing problems (Kohl and Madsen, 1997). The (MP) formulation is a candidate for LD approach since constraints (2) are the only coupling constraints for vehicles and the rest of the constraints and the objective is separable by vehicle. Hence, by relaxing the constraints (2) through Lagrangian relaxation, (MP) can be decomposed to $|K|$ subproblems, each

corresponding to a single vehicle.

The Lagrangian relaxation of MP with respect to constraints (2) results in the following relaxed problem (LR),

$$\text{(LR)} \qquad \Phi(\lambda) = \min_{(x,y)\in\Omega}\left[\sum_{k\in K} J_k(x,y) + \sum_{i\in C} \lambda_i\left(1 - \sum_{k\in K}\sum_{r\in R} y_{ir}^k\right)\right], \qquad (19)$$

where $\lambda = \left(\lambda_1, \cdots, \lambda_{|C|}\right) \in \Re^{|C|}$is the vector of Lagrangian multipliers associated with constraints (2). The set denotes all feasible solutions of the (LR). Then, the Lagrangian Dual (LD) problem maximizes the (LR) solution, which is a lower bound on z_{MP}^*.

$$\text{(LD)} \qquad \Phi_{LD}^* = \max_{\lambda}\left(\Phi(\lambda)\right). \qquad (20)$$

The set splits into $|K|$ disjoint subsets, i.e. $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_{|\mathrm{K}|}$, where each Ω_{k} is defined by constraints (3)−(10) for a given $k \in K$. Further, the objective of (LD) is additive, thus leading to the following decomposition,

$$\Phi(\lambda) = \sum_{k\in K} \Phi_k(\lambda) + \sum_{i\in C} \lambda_i. \qquad (21)$$

where $\Phi_k(\lambda) = \min_{(x,y)\in\Omega_{\mathrm{k}}} L_k(\lambda, x, y)$ and $L_k(\lambda, x, y) = J_k(x,y) - \sum_{i\in C}\sum_{r\in R} \lambda_{\,i} y_{ir}^k$ is the Lagrangian function of the k^{th} subproblem. To solve (LD), we solve the primal subproblem $\Phi_k(\lambda)$ for each vehicle k at the low-level and update the Lagrangian multipliers at the high-level, e.g., using subgradient optimization (Conejo et al., 2006; Fisher, 2004; Geoffrion, 1974). The optimization at both levels is performed iteratively until the dual solution converges.

However, since the vehicles are homogeneous, the subproblems $\Phi_k(\lambda)$ are identical; i.e. $\Omega_1 = \Omega_2 = \cdots = \Omega_{|\mathrm{K}|}$. Hence, all subproblems have the same optimal solution with identical objective value. Accordingly, solving (LD) is equivalent to solving the following,

$$|K|\max_{\lambda}\left(\min_{(x,y)\in\Omega_k} L_k(\lambda,x,y)\right)+\sum_{i\in C}\lambda_i \tag{22}$$

where $k \in K$ is any one of the subproblems. This case of identical subproblems presents challenges in the solution process. In particular with discrete decisions, it leads to oscillating dual solutions, affecting the convergence rate. Further, the solutions converged are primal infeasible and provide a lower bound on z^*_{MP} that can be weak. Lastly, the primal infeasibility of the solutions requires integration with an exact (heuristic) method such as branch-and-bound (Lagrangian heuristic) to obtain optimal (good quality) solutions (Kohl and Madsen, 1997).

3.4.2 Successive Subproblem Solving Method

We adapt the Successive Subproblem Solving (SSS) method to avoid the challenges associated with the standard Lagrangian Decomposition method due to the identical subproblems. This approach is introduced by Zhai et al. (2002) to solve the unit commitment problem in electrical power generator scheduling. The SSS approach extends and improves over the standard Lagrangian Decomposition method by addressing the dual solution oscillation. However, it does not guarantee either the primal feasibility or the quality of feasible solutions. We address these issues in Section 4.4 by developing a modified variable target value method for subgradient optimization for SSS approach.

In SSS, we introduce an absolute penalty term that helps to reduce the oscillation and constraint violations more rapidly. Accordingly, the Lagrangian function is revised to the following augmented form,

$$\hat{L}(\omega,\lambda,x,y)=\sum_{k\in K}J_k(x,y)+\sum_{i\in C}\lambda_i\left(1-\sum_{k\in K}\sum_{r\in R}y_{ir}^{k}\right)$$

$$+\omega\sum_{i\in C}\left|1-\sum_{k\in K}\sum_{r\in R}y_{ir}^{k}\right|, \tag{23}$$

where $\omega>0$ is the penalty parameter. The revised dual problem (PS) and dual function $\Phi(\lambda)$ are then expressed as,

$$\text{(PS)}\qquad \Phi_{PS}^{*}(\omega)=\max_{\lambda}\widehat{\Phi}(\omega,\lambda)=\max_{\lambda}\left(\min_{(x,y)\in\Omega}\hat{L}(\omega,\lambda,x,y)\right). \tag{24}$$

The $\Phi_{PS}^{*}(\omega)$ is the optimum dual solution with penalty weight ω. The optimum solution (PS) can be either a feasible or infeasible solution to the original problem (MP). If the solution is feasible, it can be shown that it is also optimum, i.e., no duality gap $\Phi_{PS}^{*}(\omega)=z_{MP}^{*}$. Following theorem establishes that the $\Phi_{PS}^{*}(\omega)$ is a lower bound on the primal optimum solution z_{MP}^{*}.

Theorem 3. For any ω and $\lambda,\widehat{\Phi}(\omega,\lambda)\le\Phi_{PS}^{*}(\omega)\le z_{MP}^{*}\le J(x,y)$.

Proof. By definition from (1) and (24), we have $z_{MP}^{*}\le J(x,y)$ and $\widehat{\Phi}(\omega,\lambda)\le\Phi_{PS}^{*}(\omega)$, respectively. Let (x^{*},y^{*}) be the primal optimum solution to problem (MP) and λ^{*} denote the optimum multipliers. The primal optimum solution is feasible, thus, satisfies constraint set (2). Accordingly, we have

$$z_{MP}^{*}=\sum_{k\in K}J_k(x^{*},y^{*})+\sum_{i\in C}\lambda_{\mathrm{i}}^{*}\left(1-\sum_{k\in K}\sum_{r\in R}y_{ir}^{k*}\right)+\omega\sum_{i\in C}\left|1-\sum_{k\in K}\sum_{r\in R}y_{ir}^{k*}\right|=\hat{L}(\omega,\lambda^{*},x^{*},y^{*}).$$

From the definition (24),

$$\Phi_{PS}^{*}(\omega)=\min_{(x,y)\in\Omega}\hat{L}(\omega,\lambda^{*},x,y)\ \le\hat{L}(\omega,\lambda^{*},x^{*},y^{*})=z_{MP}^{*}.\ \blacksquare$$

Revised Lagrangian function (23) cannot be decomposed into k subproblems due to the penalty term. Hence, to calculate the subgradient of $\widehat{\Phi}(\omega,\lambda)$ with respect to λ, we now

need to solve the integrated low-level problem $\min_{(x,y)\in\Omega}\hat{L}(\omega,\lambda,x,y)$, which is computationally inefficient. Revised Lagrangian function in (23), however, can be reformulated as an additive function. Let us redefine the Lagrangian function for k^{th} vehicle as follows:

$$\hat{L}_k(\omega,\lambda,x,y) = J_k(x,y) - \sum_{i\in C}\sum_{r\in R}\lambda_i y_{ir}^k + \omega\sum_{i\in C}\left|q_k(i) - \sum_{r\in R} y_{ir}^k\right| , \quad (25)$$

where,

$$q_k(i) = 1 - \sum_{\substack{s\in K\\ s\neq k}}\sum_{r\in R} y_{ir}^s . \quad (26)$$

It can be verified that the Lagrangian function (23) can be expressed in terms of $\hat{L}_k(\omega,\lambda,x,y)$ and $q_k(i)$ as follows:

$$\hat{L}(\omega,\lambda,x,y) = \hat{L}_k(\omega,\lambda,x,y) + \sum_{\substack{s\in K\\ s\neq k}} J_s(x,y) + \sum_{i\in C}\lambda_i q_k(i) . \quad (27)$$

Since (27) is additive, we can now solve the (PS) in parts, e.g., for each vehicle. The subproblem for vehicle k is then defined as follows:

$$(\mathrm{PS_k}) \quad \hat{\Phi}_k(\lambda) = \min_{(x,y)\in\Omega_k} \hat{L}_k(\omega,\lambda,x,y) . \quad (28)$$

The variable $q_k(i)$ is fixed for the k^{th} subproblem. The $q_k(i)$ links subproblem k to other subproblems by conveying the information about customer i's assignment to other vehicles. Hence, the solutions of the subproblems are likely to be different from each other, thus alleviating the issues associated with identical subproblems.

In solving (PS), the SSS method solves the vehicle subproblems one at a time, while calculating the $q_k(i)$ $\forall i \in C$ using the solution from other vehicles. The SSS method updates the Lagrangian multipliers after solving any of the subproblems. Note that this is needed since solving subproblems one after another using the same multipliers improves the $\hat{L}(\omega,\lambda,x,y)$ at a decreasing rate because the subgradient directions are not being updated. In SSS, the

Lagrangian multipliers are updated using the surrogate subgradient (SSG) approach introduced by Zhao et al. (1999). The standard subgradient approach requires solving all subproblems to obtain the subgradient direction (Geoffrion, 1974; Fisher, 2004). In the SSG approach, however, the solution to only one of the subproblems is sufficient to obtain a proper surrogate subgradient direction. Let g_i^j denote the surrogate subgradient for customer i at any iteration j and is calculated as,

$$g_i^j = 1 - \sum_{k \in K} \sum_{r \in R} \left(y_{ir}^k\right)^j \qquad \forall i \in C. \tag{29}$$

We first introduce the notation used in the SSS method and then present its algorithmic steps.

Notation:

$\left(x^j, y^j\right)_k$: solution of k^{th} subproblem at iteration j

$\left(x^j, y^j\right)$: solution at iteration j

$\hat{\lambda}^0$: initial Lagrangian multipliers, i.e., $\hat{\lambda}^0 = \left\{\hat{\lambda}_i^0, \forall i \in C\right\}$

λ^j: Lagrangian multipliers at iteration j, i.e., $\lambda^j = \left\{\lambda_i^j, \forall i \in C\right\}$

g^j: surrogate subgradients at iteration j, i.e., $g^j = \left\{\sum_{i \in C} g_i^j, \forall i \in C\right\}$

δ^j: step-size at iteration j

L_ω^j: Lagrangian function value at iteration j with penalty ω, i.e., $L_\omega^j = \hat{L}\left(\omega, \lambda^j, x^j, y^j\right)$

β: step-size update parameter, $0 < \beta < 1$

α: initialization factor for Lagrangian multipliers

ε: threshold for Lagrangian multiplier convergence criteria, $\varepsilon > 0$

SSS Procedure:

Initialization.

I.1. Given $\hat{\lambda}^0$, e.g., $\hat{\lambda}^0 = 0$, solve (LD) using (22) and obtain (x^0, y^0)

I.2. Calculate,

$$\lambda_i^0 = \alpha\left(1 - \sum_{k\in K}\sum_{r\in R}\left(y_{ir}^k\right)^0\right) \qquad \forall i \in C, \tag{30}$$

where, $0 < \alpha < \left(\Phi_{PS}^*(\omega) - \hat{L}(\omega, 0, x^0, y^0)\right) / \sum_{i\in C}\left\|1 - \sum_{k\in K}\sum_{r\in R}\left(y_{ir}^k\right)^0\right\|^2$

I.3. Calculate $L_\omega^0 = \hat{L}(\omega, \lambda^0, x^0, y^0)$ and update Lagrangian multipliers:

$$\lambda^1 = \lambda^0 + \delta^0 g^0,$$

where $0 < \delta^0 = \beta(\Phi_{PS}^*(\omega) - L_\omega^0)/\|g^0\|^2$ and $0 < \beta < 1$. Set $j = 1$.

Step 1. Subproblem Solution:

1.1. For $k = 1,2, \ldots, |K|$, Repeat:

1.1.a. Solve subproblem $(\mathrm{PS_k})$ in (28) by setting

$$\left(x^j, y^j\right)_s = \left(x^{j-1}, y^{j-1}\right)_s \text{for} s \in K \;\; s \neq k$$

to obtain $\left(x^j, y^j\right)_k$.

1.1.b. If the following improvement condition is satisfied,

$$L_\omega^j = \hat{L}\left(\omega, \lambda^j, x^j, y^j\right) < \hat{L}\left(\omega, \lambda^j, x^{j-1}, y^{j-1}\right), \tag{31}$$

where $\left(x^j, y^j\right) = \left(x^j, y^j\right)_k \cup \left\{\left(x^{j-1}, y^{j-1}\right)_s | s \in K, s \neq k\right\}$,

then go to Step 2, otherwise continue with the next k.

1.2. Set $\left(x^j, y^j\right) = \left(x^{j-1}, y^{j-1}\right)$.

Step 2. Subgradient Optimization:

2.1. Update Lagrangian multipliers :

$$\lambda^{j+1} = \lambda^j + \delta^j g^j,$$

where $0 < \delta^j = \beta\left(\Phi_{PS}^* - L_\omega^j\right)/\left\|g^j\right\|^2$ and $0 < \beta < 1$.

Step 3. Check the stopping criteria.

3.1. If $\left\|\lambda^{j+1} - \lambda^j\right\| \leq \varepsilon$, then go to Step 4; otherwise set $j = j + 1$ and return to Step 1.

Step 4. Terminate with solution $\left(x^j, y^j\right)$.

The SSS method is initialized by solving (LD) to obtain initial solutions to estimate the starting values for Lagrangian multipliers. The bounding of α in the initialization ensures that $L_\omega^0 = \hat{L}(\omega, \lambda^0, x^0, y^0) < \Phi_{PS}^*(\omega)$. This inequality is important for convergence analysis as explained in the next section. The subproblems in Step 1 are sequentially solved until the improvement condition in (31) is attained. In each subproblem solution, the previous iteration's solutions are used to calculate $q_k(i)\ \forall i \in C$. When none of the vehicle k's subproblem solution satisfies (31), then the previous iteration's solution is maintained. The multipliers are updated using the surrogate gradient in Step 2.1. The SSS method terminates when multipliers converge.

The SSS method requires $\Phi_{PS}^*(\omega)$. This value, however, is generally unknown in advance and needs to be estimated. A poor underestimation may result in convergence to a primal infeasible solution with large duality gap (see Theorem 4). In the standard Lagrangian method, the value used in place of $\Phi_{PS}^*(\omega)$ is an overestimation of z_{MP}^*, which affects the convergence rate. However, the solutions converged are either primal infeasible or optimal (Held et al., 1974). In comparison, SSS method, using an overestimation of $\Phi_{PS}^*(\omega)$, may converge to a primal feasible but not optimal solution. Hence, SSS differs from the standard

Lagrangian method, as it may converge to a suboptimal primal feasible solution without a feasibility recovery heuristic. The reason for this is that the SSS minimizes the augmented Lagrangian relaxation in (25) by solving decomposed subproblems in Step 1. The bound estimate of $\Phi_{PS}^{*}(\omega)$ in SSS is therefore critical affecting both the convergence rate and the solutions converged, i.e., primal feasible or infeasible. We present the bound estimation procedure in Section 4.2.2.

3.4.2.1 Convergence Analysis

In this section, we provide convergence results for SSS method with subgradient optimization using $\Phi_{PS}^{*}(\omega)$. The following theorem establishes that the Lagrangian function value at each iteration of SSS underestimates the optimal solution to the (PS).

Theorem 4. (Solution Bounding) For a given ω, at each iteration i, $L_{\omega}^{j} < \Phi_{PS}^{*}(\omega)$.

Proof. For $j = 0$, the condition on α suffices. In the case of $j \geq 1$, from (31) we have,

$$L_{\omega}^{j} = \hat{L}\left(\omega, \lambda^{j}, x^{j}, y^{j}\right) \leq \hat{L}\left(\omega, \lambda^{j}, x^{j-1}, y^{j-1}\right).$$

Further,

$$\hat{L}\left(\omega, \lambda^{j}, x^{j-1}, y^{j-1}\right) = \hat{L}\left(\omega, \lambda^{j-1}, x^{j-1}, y^{j-1}\right) + \hat{L}\left(\omega, \lambda^{j}, x^{j-1}, y^{j-1}\right) - \hat{L}\left(\omega, \lambda^{j-1}, x^{j-1}, y^{j-1}\right)$$

$$= L_{\omega}^{j-1} + \sum_{i \in C}\left(\lambda_{i}^{j} - \lambda_{i}^{j-1}\right)\left(1 - \sum_{k \in K}\sum_{r \in R} y_{ir}^{k}\right) = L_{\omega}^{j-1} + \ \delta^{j-1}\left\|g^{j-1}\right\|^{2}.$$

From the definition of δ^{j} in Step 3 of SSS procedure we have, $L_{\omega}^{j} \leq L_{\omega}^{j-1} + \beta\left(\Phi_{PS}^{*}(\omega) - L_{\omega}^{j-1}\right)$. Since $\beta < 1$, we obtain, $L_{\omega}^{j} < L_{\omega}^{j-1} + \Phi_{\mathrm{PS}}^{*}(\omega) - L_{\omega}^{j-1} \leq \Phi_{\mathrm{PS}}^{*}(\omega)$. ■

The following lemma states that the search direction of the Lagrangian multipliers in any iteration is always a proper direction, i.e., $\left(\lambda^{*} - \lambda^{j}\right)g^{j} > 0$.

Lemma 3. (Direction). Let λ^{*} be the optimal multiplier vector, then $\Phi_{\mathrm{PS}}^{*}(\omega) - L_{\omega}^{j} \leq$

$(\lambda^* - \lambda^j)g^j,\ \forall j.$

Proof. Based on (23) and (24), we have

$$\Phi^*_{\mathrm{PS}}(\omega) = \hat{\Phi}(\omega, \lambda^*) = \hat{L}(\omega, \lambda^*, x^*, y^*) \le \hat{L}\left(\omega, \lambda^*, x^j, y^j\right) = L^j_\omega + \hat{L}\left(\omega, \lambda^*, x^j, y^j\right) - L^j_\omega$$

$$= L^j_\omega + \left(\lambda^* - \lambda^j\right)g^j \ .$$

Last step follows from the definition of g^j in (29) and Lagrangian function $\hat{L}(\omega, \lambda, x, y)$ in (23). From Theorem 4, we have $\Phi^*_{\mathrm{PS}}(\omega) - L^j_\omega > 0$, thus the theorem's result follows.■

The convergence of the Lagrangian multipliers is established by the following theorem.

Theorem 5. (Convergence) In the SSS algorithm, the Lagrangian multipliers are converging; i.e,

$$\left\|\lambda^* - \lambda^{j+1}\right\|^2 < \left\|\lambda^* - \lambda^j\right\|^2 \quad \forall j,$$

where λ^* is the optimal multiplier vector.

Proof. From (32) we have

$$\left\|\lambda^* - \lambda^{j+1}\right\|^2 = \left\|\lambda^* - \lambda^j - \delta^j g^j\right\|^2$$

$$= \left\|\lambda^* - \lambda^j\right\|^2 + \left(\delta^j\right)^2\left\|g^j\right\|^2 - 2\delta^j\left(\lambda^* - \lambda^j\right)g^j.$$

Using result from Lemma 3, we have,

$$\left\|\lambda^* - \lambda^{j+1}\right\|^2 \le \left\|\lambda^* - \lambda^j\right\|^2 + \left(\delta^j\right)^2\left\|g^j\right\|^2 - 2\delta^j\left(\Phi^*_{\mathrm{PS}}(\omega) - L^j_\omega\right).$$

Then, from the definition of δ^j in Step 2 of SSS procedure,

$$\left\|\lambda^* - \lambda^{j+1}\right\|^2 \le \left\|\lambda^* - \lambda^j\right\|^2 - \delta^j\left(\Phi^*_{\mathrm{PS}}(\omega) - L^j_\omega\right),$$

and using the result of Theorem 4, we obtain $\left\|\lambda^* - \lambda^{j+1}\right\|^2 \le \left\|\lambda^* - \lambda^j\right\|^2$. ■

Increasing the penalty parameter improves the quality of the solution converged as

established by the following theorem.

Theorem 6. For any two penalty weight ω_1 and ω_2, where $0 < \omega_1 < \omega_2$,

$$\widehat{\Phi}(\omega_1, \lambda) \leq \widehat{\Phi}(\omega_2, \lambda) \leq \Phi^*_{\mathrm{PS}}(\omega_2) \leq z^*_{MP}.$$

Proof. From (23), we have $L^j_{\omega_2} - L^j_{\omega_1} = (\omega_2 - \omega_1)\left|\sum_{i \in C} g^j_i\right| \geq 0$.Thus, $L^j_{\omega_2} \geq L^j_{\omega_1}$. Subsequently from (24), we have,

$$\min L^j_{\omega_2} \geq \min L^j_{\omega_1} \ ,$$

$$\max \widehat{\Phi}(\omega_2, \lambda) \geq \max \widehat{\Phi}(\omega_1, \lambda) \ ,$$

$$\Phi^*_{\mathrm{PS}}(\omega_2) \geq \Phi^*_{\mathrm{PS}}(\omega_1).$$

From Theorem 5, we already have $\Phi^*_{\mathrm{PS}}(\omega_2) \leq z^*_{MP}$.■

While Theorem 6 states that the solution quality of SSS improves with penalty parameter, we note that choosing ω very large may cause ill-conditioning and numerical instability.

3.4.2.2 Bound Estimation: Variable Target Value Method

The SSS procedure uses an estimate of $\Phi^*_{\mathrm{PS}}(\omega)$ for the surrogate subgradient optimization. Rather than using a static estimate, we dynamically change this estimate in order to obtain a good quality primal feasible solution. Specifically, we modify the variable target value method (VTVM) presented in Lim and Sherali (2006) and incorporate backtracking to improve the target value estimation. Since the SSS method can converge to primary feasible but suboptimal solution, we integrated a backtracking phase within the VTVM to improve the quality of the feasible solution.

We modify the SSS method by replacing $\Phi^*_{PS}(\omega)$ with a dynamically adjusted estimate

Φ_{PS}^{j} (target value). Analogous to Theorem 3, it can be shown that $L_{\omega}^{j} < \Phi_{PS}^{j}$ holds true for each iteration j. In choosing the estimate Φ_{PS}^{j}, the goal is to approximate z_{MP}^{*} as close as possible. In standard VTVM method, the target value Φ_{PS}^{j} is increased as long as the convergence rate is satisfactory and then decreased to close in on an optimal solution. In our adaptation, we increase the target value Φ_{PS}^{j} with a controlled rate until we find a primal feasible solution. Finding a primal feasible solution, as explained in Section 4.2.1, indicates that the target value is an overestimation of $\Phi_{PS}^{*}(\omega)$. This primal feasible solution, however, maybe a low quality suboptimal solution. Therefore, with a backtracking phase, we revise the latest target value to obtain a better primal feasible solution. Specifically, after encountering with a primal feasible solution, we return back to a past iteration where the target value underestimates the current solution's objective value. Then, the modified SSS repeats the iteration with a smaller step size in an effort to find an improved primal feasible solution.

We first provide the notation used in VTVM with backtracking and then present the modified steps of the SSS procedure. Next, we briefly discuss the convergence behavior of the SSS with backtracking. Note that we replace $\Phi_{PS}^{*}(\omega)$ with Φ_{PS}^{j} in the remainder steps of the SSS procedure.

Notation for SSS with Backtracking VTVM:

Φ_{PS}^{j}: target value at iteration j

$\overline{\Phi}_{PS}^{j}$: upper bound on the optimal solution at iteration j

$\underline{\Phi}_{PS}$: lower bound on the optimal solution value

(x^{*}, y^{*}): an optimal solution to (MP)

Δ_j: accumulated improvements since the last Lagrangian function improvement until the beginning of iteration j

ε_j: acceptance tolerance that the current incumbent value L_ω^j is close to the target value Φ_{PS}^j in iteration j

σ: acceptance interval parameter

η_j: fraction of cumulative improvement that is used to increase the target value in iteration j

$\varepsilon_{\mathrm{GAP}}$: optimality gap threshold

Modified Steps of the SSS Procedure with Backtracking VTVM:

Initialization. Execute Steps I.1, I.2, I.3 of the original SSS procedure, and,

I.4. Set $\Phi_{PS}^{j=1} = \underline{\Phi}_{PS}$, $\overline{\Phi}_{PS}^{j=1} = +\infty$, $\eta_{j=1} = 0.35$, $\sigma = 0.2, \Delta_{j=1} = 0$, and $\varepsilon_{j=1} = \sigma\left(\Phi_{PS}^{j=1} - L_\omega^{j=0}\right)$.

Step 1. Subproblem Solution & Backtracking:

1.1. For $k = 1,2, \ldots, |K|$, Repeat:

1.1.a. Solve subproblem ($\mathrm{PS_k}$) in (28) by setting $\left(x^j, y^j\right)_s = \left(x^{j-1}, y^{j-1}\right)_s$ for $s \in K, s \neq k$ and obtain $\left(x^j, y^j\right)_k$. Denote $\left(x^j, y^j\right) = \left(x^j, y^j\right)_k \cup \left\{\left(x^{j-1}, y^{j-1}\right)_s | s \in K, s \neq k\right\}$.

1.1.b. If $\left(x^j, y^j\right)$ is primal feasible, then

i. Set $(x^*, y^*) = (x^j, y^j)$,

ii. Set $\overline{\Phi}_{PS}^j = L_\omega^j = J\left(x^j, y^j\right)$,

iii. Set algorithm parameters, variables, and solutions back to iteration v, i.e.,

where $v = max\left\{l: \Phi_{PS}^l < \overline{\Phi}_{PS}^j = L_\omega^j\right\}$,

iv. Set $j := v$,

v. Set $\beta = \beta/2$ and repeat iteration j with updated Lagrange multipliers $\lambda^j = \lambda^{j-1} + \delta^{j-1} g^{j-1}$.

1.1.c. If the following improvement condition is satisfied,

$$L_\omega^j = \hat{L}\left(\omega, \lambda^j, x^j, y^j\right) < \hat{L}\left(\omega, \lambda^j, x^{j-1}, y^{j-1}\right),$$

where $\left(x^j, y^j\right) = \left(x^j, y^j\right)_k \cup \left\{\left(x^{j-1}, y^{j-1}\right)_s | s \in K, s \neq k\right\}$,

then go to Step 2, otherwise continue with the next k.

1.2. Set $\left(x^j, y^j\right) = \left(x^{j-1}, y^{j-1}\right)$.

Step 2. Subgradient Optimization & VTVM:

2.1. If $L_\omega^j > \Phi_{PS}^j - \varepsilon_j$, then

2.1.a. Update the target value $\Phi_{PS}^{j+1} = \min\left\{L_\omega^j + \varepsilon_j + \eta_{\mathrm{j}} \Delta_{\mathrm{j}}\ , \overline{\Phi}_{PS}^j\right\}$,

2.1.b. Update the threshold $\varepsilon_{j+1} = \sigma\left(\Phi_{PS}^{j+1} - L_\omega^j\right)$,

2.1.c. Reset $\Delta_{\mathrm{j}} = 0$,

2.1.d. Update $\eta_{\mathrm{j}+1} = min\{2\eta_{\mathrm{j}}, 1\}$,

otherwise set $\Phi_{PS}^{j+1} = \Phi_{PS}^j$, $\overline{\Phi}_{PS}^{j+1} = \overline{\Phi}_{PS}^j$, $\eta_{\mathrm{j}+1} = \eta_{\mathrm{j}}$ and $\varepsilon_{j+1} = \varepsilon_j$.

2.2. Update $\Delta_{\mathrm{j}+1} = \Delta_{\mathrm{j}} + \left(L_\omega^j - L_\omega^{j-1}\right)$

2.3. Update Lagrangian multipliers:

$$\lambda^{j+1} = \lambda^j + \delta^j g^j,$$

where $0 < \delta^j = \beta\left(\Phi_{PS}^j - L_\omega^j\right) / \left\|g^j\right\|^2$ and $0 < \beta < 1$.

Step 3. Check the stopping criteria:

3.1. If $\left(\overline{\Phi}_{PS}^{j} - \ L_{\omega}^{j}\right) \leq \varepsilon_{\text{GAP}}$ or $\left\|\lambda^{j+1} - \lambda^{j}\right\| \leq \varepsilon$, then terminate with Step 4;

otherwise set $j = j + 1$ and go to Step 1.

Step 4. Terminate with (x^*, y^*).

The SSS with backtracking VTVM initializes the target value Φ_{PS}^{j} with an underestimation $\underline{\Phi}_{PS}$ of the dual optimal value, e.g., linear programming relaxation. From Lemma 3, it can be shown that the Lagrangian multipliers provide a proper direction and thus the dual solution L_{ω}^{j} is non-decreasing. When the dual solution is primal feasible, we perform the backtracking phase in Step 1.1b. This backtracking helps improve the quality of the subsequent feasible solutions by reverting to the an iteration v satisfying $\Phi_{PS}^{v} < L_{\omega}^{j}$ and repeat the iteration j with smaller step size. As the L_{ω}^{j} closes in on the target value such that L_{ω}^{j} is within ε_j threshold of Φ_{PS}^{j}, then Step 2.1.a updates the target value based on the accumulated improvement Δ_{j} and $\overline{\Phi}_{PS}^{j}$. This update guarantees that the dual solution and the target value is separated by at least ε_j while ensuring that the target value does not exceed the upper bound. The threshold ε_j is updated in Step 2.1.b.

Choosing large values for σ increases ε_j. With higher ε_j values, we are more likely to consider that the L_{ω}^{j} is close to the target value and thus update the target value more frequently and with larger increments (Step 2.1.a). This can result in poor feasible solutions as the upper bound $\overline{\Phi}_{PS}^{j}$ might not have decreased sufficiently. In contrast, lower σ values reduce the convergence rate. The required ranges for acceptance interval parameter and fraction of cumulative improvement are $\sigma \in (0,1/3]$ and $\eta_{\text{j}} \in (0,1]$ (Lim and Sherali, 2006).

The algorithm terminates and returns the best primal feasible solution when the gap between the best feasible solution and the Lagrangian dual function value falls below the optimality gap threshold ($\varepsilon_{\mathrm{GAP}}$).

3.5 Computational Experiments

We report on the results of two computational experiments. First, we investigate the computational and solution quality performance of the proposed approach for solving the ATD-PDP. Next, we present the results of implementing AAAP in a real-world case study using the Southern California region discussed in Hall (2002). The SSS with backtracking VTVM is programmed in Matlab R2008a and integer programs are solved with CPLEX 12.1. All experimental runs are conducted on a PC with Intel(R) Core 2 CPU, 1.66 GHz processor and 1 GB RAM running on Windows XP Professional. In the following section, we report on the computational results of the two variants of the SSS method, namely *SSS with backtracking VTVM* (SSS-B-VTVM) and *VTVM base SSS without backtracking* (SSS-VTVM).

3.5.1 Evaluation of the Solution Algorithm

We generated a set of test problems varying from small to large problem scenarios. Since the ATD-PDP is a new problem, no benchmark datasets are available. In generating the data sets, we adhered to the development procedure described in Solomon (1987). The problem scenarios have one depot and one or two airports each with three flight itinerary options for each customer. The third option represents the recourse flight itinerary option. For a problem scenario with $n = |C|$ customers and $m = |H|$ airports, we first generate

$(1+m+n)$ locations from a uniform distribution over the square bounded by $[0, 10(1+m+n)] \times 0, 10(1+m+n)]$. Next, we randomly label the nodes as the depot, airports and customers to avoid any association between the location and identity of a node. The travel time between nodes is calculated as the Euclidean distances between them. The travel cost between two nodes is set equal to their travel time.

For each airport h, the departure times Q_r^h of flights are independent and identically distributed according to a uniform distribution $U[\varphi/|K|, \theta]$ where $|K|$ is the number of available vehicles; φ is the heuristic solution to a TSP problem consisting of the depot (origin), all customers and the airport (destination) and obtained through the greedy next best routing heuristic. The cost of flight itinerary options F_{ir}^h are independent and identically distributed according to a uniform distribution $U[a, b]$ where a and b are the bounds set as 100 and 600, respectively. The flight itinerary options are sorted from cheapest to most expensive and assigned to the flight itineraries based on the starting times such that cheaper itineraries start earlier.

We have conducted experiments using 5, 7, 10 and 15 customer cases. For each experiment scenario, we generated 10 independent instances and solve them using CPLEX, SSS-VTVM, and SSS-B-VTVM. Since there is no prior work on ATD-PDP, we compare the proposed methods with the CPLEX solution of (MP) as an integrated model. We restricted the solution time to 3 hours for all methods and report the best feasible solution attained within the time limit for each instance. In total, we have solved 300 problem instances using both methods. We first present the results of SSS-VTVM. In this method, we terminate the solution procedure when a primal feasible solution is found. Table 3.2 presents the comparative solution

quality and computational performance results and Table 3.1 describes the column headings. Table 3.2 optimality results are based on the gap between the best solutions found in each method and the lower bound from CPLEX. For each problem scenario, we report the average, minimum, and maximum optimality gap of the methods and the comparison of the CPU time in terms of a ratio. The CPU time ratio metric is selected since we report the performance across all the instances.

Table 3.1. Description of column headings in Table 3.2.

|C|	Number of customers
|H|	Number of airports
|K|	Number of vehicles
CPLEX Gap (%)	Gap between the best feasible solution and lower bound of CPLEX
SSS Gap (%)	Calculated as (SSS solution - CPLEX Lower Bound)/CPLEX Lower Bound.
CPU Time Ratio	The ratio of CPLEX's CPU time to SSS's CPU time
Hit	Percentage of time that SSS or CPLEX finds an optimum solution

We first consider the results without backtracking, i.e., SSS-VTVM. For small size problems with 5 and 7 customers, the CPLEX's average gap across all scenarios is 0.5% whereas the SSS's gap is 2.1%. On the average, CPLEX finds the optimum in 96% of the cases and SSS finds in 30% of the cases. While the CPLEX's solution quality performance is slightly better than that of SSS's, the difference is small. Further, SSS is able to attain good quality solutions up to 346 times faster and on the average 29 times faster.

Table 3.2. Comparative performance of CPLEX and SSS.

							SSS-VTVM							SSS-B-VTVM						
			CPLEX Gap (%)				SSS Gap (%)				CPU Time Ratio			SSS Gap (%)				CPU Time Ratio		
|C|	|H|	|K|	Ave	Min	Max	Hit	Ave	Min	Max	Hit	Ave	Min	Max	Ave	Min	Max	Hit	Ave	Min	Max
5	1	3	0.0	0.0	0.0	100	1.8	0.0	5.2	40	1	0	3	0.1	0.0	0.8	70	1	0	1
		4	0.0	0.0	0.0	100	1.3	0.0	5.8	40	1	0	2	0.2	0.0	1.5	70	1	0	1
		5	0.0	0.0	0.0	100	0.7	0.0	2.1	50	1	0	3	0.3	0.0	1.8	80	1	0	2
	2	3	0.0	0.0	0.0	100	0.8	0.0	5.6	40	25	0	146	0.1	0.0	1.2	80	11	0	57
		4	0.0	0.0	0.0	100	1.5	0.0	8.2	50	8	0	28	0.0	0.0	0.2	90	4	0	12
		5	3.3	0.0	33.0	90	2.8	0.0	7.1	20	31	1	228	0.5	0.0	2.0	50	27	0	228
7	1	3	0.0	0.0	0.0	100	1.5	0.0	6.2	20	10	1	28	0.8	0.0	4.0	30	9	0	28
		4	0.0	0.0	0.0	100	0.6	0.0	2.7	50	11	0	38	0.4	0.0	2.7	70	10	0	38
		5	0.0	0.0	0.0	100	1.9	0.1	5.7	0	9	1	34	0.5	0.0	1.8	30	7	0	34
	2	3	0.0	0.0	0.0	100	1.4	0.0	3.3	10	102	6	344	0.9	0.0	3.3	30	91	4	344
		4	1.9	0.0	14.9	80	4.2	0.0	14.9	20	54	4	346	2.9	0.0	14.9	30	50	4	346
		5	1.5	0.0	7.3	78	3.2	0.0	11.6	11	99	18	309	1.8	0.0	6.8	22	73	9	207
10	1	3	0.2	0.0	2.3	82	2.0	0.0	6.7	18	431	7	2,205	0.8	0.0	2.3	27	239	5	966
		4	0.9	0.0	5.2	70	3.7	0.1	7.5	0	446	14	1,549	1.9	0.0	6.3	10	281	14	1,549
		5	0.7	0.0	3.2	70	4.0	0.0	6.9	0	195	0	645	1.4	0.0	4.9	0	143	0	645
	2	3	2.9	0.0	9.6	46	5.0	0.0	15.6	8	172	1	753	3.4	0.0	9.9	23	81	1	252
		4	9.1	1.1	34.4	0	10.8	2.8	19.5	0	234	2	748	7.9	2.3	14.6	0	120	2	255
		5	9.5	0.9	26.1	0	11.1	3.4	20.2	0	299	27	1,581	8.8	2.6	16.1	0	145	25	688
15	1	3	4.5	0.0	35.9	23	6.7	0.3	14.1	0	520	2	1,892	5.2	0.2	14.1	0	313	2	1,382
		4	4.7	0.0	26.6	10	5.1	1.6	15.4	0	198	64	687	3.9	1.6	8.2	0	169	53	687
		5	9.6	0.0	27.6	7	8.8	3.9	20.3	0	289	5	1,048	6.4	0.0	12.8	7	195	2	1,048
	2	3	13.1	2.8	51.4	0	9.5	4.4	16.7	0	43	2	143	8.9	4.4	16.7	0	40	2	143
		4	33.0	4.0	69.1	0	15.2	7.4	25.3	0	62	1	287	12.5	3.4	25.3	0	28	1	113
		5	31.8	0.1	66.4	0	12.5	3.8	20.5	0	72	3	535	11.2	2.0	19.6	0	42	1	244

The CPLEX's gap for medium size problems with 10 customers averages 4.8% across all scenarios and an optimum is found for 45% of the cases. In comparison, the SSS has an average gap of 6.2% and finds an optimum for 4% of the cases. While the CPLEX's solution quality performance is slightly better than that of SSS, the difference is small. The SSS is able to attain good quality solutions up to 2,205 times faster and on the average 296 times faster. For the large size problems with 15 customers, the CPLEX's average gap is 16.2% with an optimality hit rate of 7% of the time. While the SSS's average gap is 9.8%, it is not able to find a verifiable optimal solution. Unlike small and medium size problem scenarios, SSS has a better average gap performance than that of CPLEX's for large size problems. As before, the SSS is much more efficient than CPLEX, e.g., up to 1,892 times faster and on the average 197 times faster.

The last seven columns of Table 3.2 present the results for SSS-B-VTVM which improves

over the solution quality performance of the SSS-VTVM through the backtracking phase. For small size problems the average gap is reduced to 1.2% and the optimality hit rate is increased to 54%. These improvements are attained without sacrificing the CPU time performance advantage over CPLEX. For medium size problems, the average gap performance of SSS-B-VTVM is better than that of the CPLEX, e.g., 4.0% versus 4.8%, respectively. While this improvement comes with reduced CPU time performance advantage, the SSS-B-VTVM is still 168 times faster than CPLEX on the average. For large size problems, the average gap performance improves slightly and is about half of that of the CPLEX, e.g., 8.0% versus 16.2%, respectively. The CPU time performance is reduced by a third but still about 131 times faster than CPLEX on the average. Across all problem instances, the CPLEX, SSS-VTVM, and SSS-B-VTVM have on the average 5.3%, 4.8%, and 3.4% optimality gap, respectively. In terms of CPU performance, SSS-VTVM and SSS-B-VTVM are on the average 138 and 87 times faster than CPLEX, respectively.

Based on the results in Table 3.2, we study the effect of number of airports, vehicles and customers on the performance of SSS-B-VTVM (Figure 3.3). The effect of the number of airports is illustrated in Figure 3.3a. With increasing number of airports, the optimality gap of SSS-B-VTVM increases at a lower rate than that of the CPLEX. For medium and large instances, the CPU performance of SSS-B-VTVM is highest with single airport and, for small instances, highest with two airports. This is because as the problem size increases, flight itinerary assignment and routing decisions become more interrelated making it difficult to solve as an integrated model. Note that the CPU time advantage of SSS-B-VTVM is significantly reduced for two airport case in the large problem instances. This is attributable to the time limit which is

mostly restrictive for CPLEX than SSS-B-VTVM.

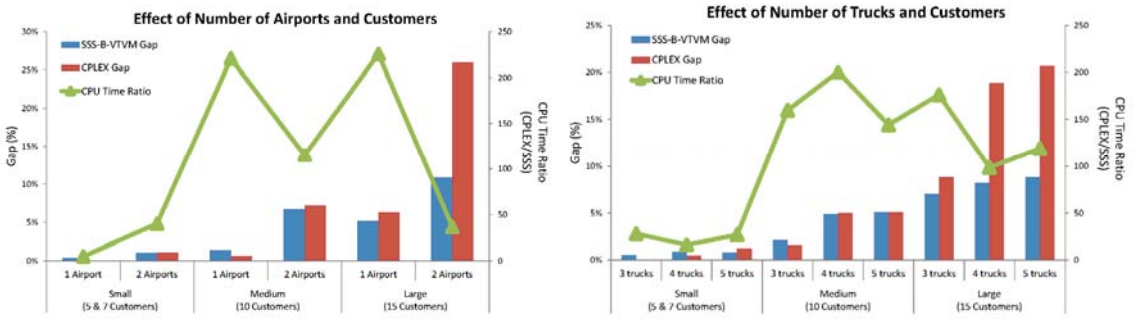


Figure 3.3. Effect of number of customers, (a) number of airports and (b) number of vehicles on the performance of SSS-B-VTVM

Figure 3.3b illustrates the effect of the number of vehicles. The gap performance of SSS-B-VTVM is robust with respect to the number of vehicles. This can be explained by the fact that additional vehicles are utilized to a lesser extent, hence their effect on the optimality is marginal. In comparison, the gap performance of CPLEX is reduced, especially, for large problems. This difference is due to the vehicle-based decomposition of SSS-B-VTVM, which is able to find quality solutions in the presence of underutilized fleet capacity. The CPU time advantage is relatively reduced, beginning with 4 vehicles in medium size problem instances. This is, indeed, a result of the time limit which makes the numerator of the CPU time ratio invariant to the number of vehicles.

3.5.2 Case Study

To assess the benefits of implementing AAAP, we conducted a case study in a Southern California MAR using real flight itinerary information and airport locations. The performance of AAAP is compared to the single airport policy where the freight forwarder can only assign

customers' air cargo loads to the flights departing from one airport.

3.5.2.1 Alternative Access Airports and Depot Locations

The Southern California MAR used in our experiments is described in Hall (2002) and illustrated in Figure 3.4. In this MAR, the Los Angeles International Airport (LAX) is the largest air-freight port. Hall (2002) suggests redirecting some of the domestic freight load to Long Beach Airport (LGB) or Ontario International Airport (ONT) to reduce the load and congestion in the LAX airport. As discussed in Hall (2002) and Chayanupatkul et al. (2004), a forwarder rarely considers more than two alternative access airports. Hence, we consider LGB and LAX as the two alternative access airports. For the location of the depot, we experimented with three location scenarios: adjacent to LAX, adjacent to LGB, and in-between LAX and LGB. We denote these depot location scenarios as DLAX, DLGB, and DMID, respectively. For the two scenarios of DLAX and DLGB, we randomly and uniformly select the depot location in a one-mile radius region with the airport in the center. For DMID scenario, we select the depot location within a one-mile radius of the city of Compton such that the travel time is identical to both the LAX and LGB airports. These regions are illustrated with dashed circles in Figure 3.4.

3.5.2.2 Customer Locations

In all experiments, the fleet size is four vehicles and there are 15 customers. We consider the scenario where the air cargo loads are time-sensitive (shipped overnight). All customer loads are available for pick-up by 7:00 pm. We generate multiple case study instances, by uniformly sampling customer locations within the MAR region, i.e., rectangular

region in Figure 3.4. The Google Maps API is used to generate the customer locations and calculate travel times. For each customer in each problem instance, we first uniformly sample a geographical coordinate (i.e., latitude and longitude) in the MAR region. Next, we determine the closest street address to this coordinate point through the Google Maps API. In case of an infeasible coordinate point (e.g., inside a lake), we re-sample for another coordinate. The travel times are estimated from the shortest paths accounting for speed limits using Google Maps API.

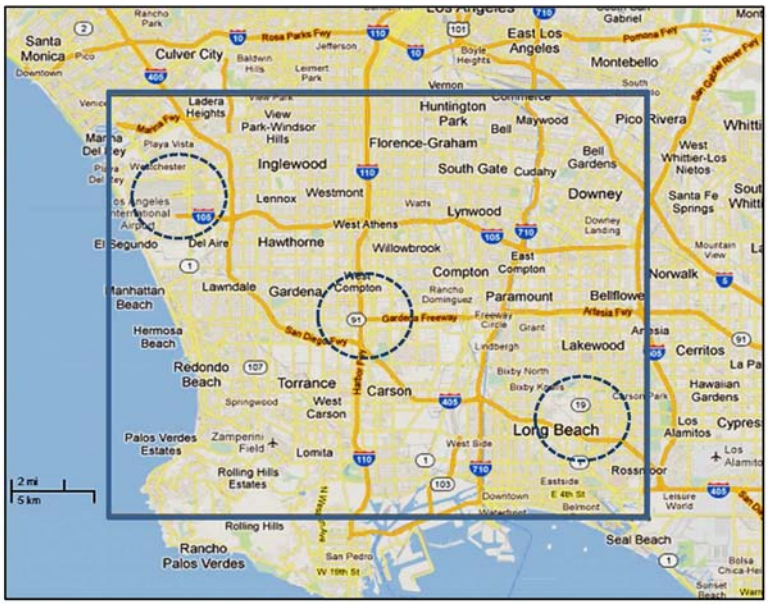


Figure 3.4. Southern California MAR used in the case study

3.5.2.3 Flight Itinerary Options

A forwarder, upon receiving a time-sensitive shipment order, can execute it via an integrator (e.g. FedEx, UPS), a mixed passenger-cargo (e.g. United Airlines, Delta Airlines, American Airlines), or a chartered/dedicated freighter. In this case study, we consider only the mixed passenger-cargo flight itinerary options, the most practiced option for small and mid-size forwarders.

Table 3.3. Case study flight itinerary options from LAX and LGB airports.

Dest.	Prob.	Origin	Airline	Departure Time		Mean Delay (min)	Mean Elapsed Time (min)
				1st	2nd		
BOS	19%	LAX	AA	22:15	07:15*	7	324
		LGB	B6	22:59	22:40*	11	324
FLL	13%	LAX	DL	21:05	09:45*	10	324
		LGB	B6	21:15	23:59	10	307
IAD	24%	LAX	AA	21:00	21:00*	11	293
		LGB	B6	21:08	21:08*	23	310
JFK	44%	LAX	AA	21:20	09:05*	10	312
		LGB	B6	21:00	08:05*	17	315

* Next day departure

We assume the final destination of a customer's air cargo is a domestic destination with direct flights from both the LAX and the LGB. Accordingly, we consider four major US airports as the destinations: Boston Logan International Airport in Massachusetts (BOS), Fort Lauderdale-Hollywood International Airport in Florida (FLL), Dulles International Airport in Dallas, Texas (IAD), and John F. Kennedy International Airport in New York (JFK). As for the airlines, we considered American Airlines (AA) and Delta Airlines (DL) for the LAX airport and JetBlue Airways (B6) for the LGB airport. In determining the cargo destination for each customer, we randomly assigned each customer's load to one of the four destinations. The probability distribution used in this assignment is based on the frequency of the outgoing flights to each destination from each airport. These probabilities are presented in the second column of Table 3.3. For each customer, there are in total four flight itinerary options, e.g., two options from each airport.

We arbitrarily selected the operation day as August $16,\ 2010$ and collected the flight

itinerary information from the BTS database[1]. The flight itinerary information, including the average departure delay and elapsed time (i.e. overall taxi-out to taxi-in time) in August 2010, is listed in Table 3.4. The departure delays are incorporated in the total delivery time by assuming flights depart late with their respective mean delay. While the first flight departure times are rather similar in two airports, the second flight departure times are notably different for some destinations. We consider the starting time of a flight itinerary as the departure time of its first flight.

3.5.2.4 Case Study Results

We evaluate the performance of different policies based on total delivery time including road and air travel times. Note that the practical implementation of the AAAP would account for forwarder's negotiated terms with air-carriers, cost structure of road transportation operations, and pricing models (Azadian et al. 2012). However, cost performance, i.e., the total delivery time, used in this case study provides ample policy comparison opportunity. Specifically, given a solution, we calculate total delivery time as the sum of road travel times by all vehicles and the total time elapsed for each customer load from the start of the operation (19:00) until its delivery time to the destination airport. We have conducted three sets of experiments corresponding to each depot location scenario (DLAX, DLGB, and DMID). In each set, we consider three different airport access policies: AAAP (with LAX & LGB), LAX only, and LGB only. For each depot location, we generated 10 problem instances and solve them with the SSS-B-VTVM algorithm under each access policy.

[1] Bureau of Transportation Statistics, U.S. Department of Transportation, Last Accessed November 2011, http://www.transtats.bts.gov/

The case study flight itinerary options in Table 3.3 show that there is no significant difference between the first flight options across the two airports. Further, the recourse flight itinerary options only differ for the loads going to BOS or FLL. Hence, in this case study, the performance differences of the three airport policies are primarily attributable to the road travel time and the small differences in the flight itinerary options. We note that the performance advantage of utilizing alternative access airports would increase when the flight itinerary options' starting times and flight itinerary durations/costs (especially for multi-leg itineraries) vary between the alternative airports. Therefore, we compare the airport policies based on the delivery time saving potential in each depot location scenario. For this, we estimate a lower bound on the total delivery time as a summation of the lower bound for flight itinerary time and road travel. The lower bound for the flight itinerary time is estimated by assigning each customer load to the cheapest itinerary accessible. The lower bound for the road travel time is calculated by solving a minimum spanning tree connecting all the nodes.

Table 4 presents the total delivery time in minutes for all problem instances in each depot location scenario and under three access policies (LAX & LGB, LGB, LAX). For AAAP policy, i.e., LAX &LGB, we report the percentage of the time that the LAX airport is selected. Last two rows in Table 3.4 present the average and standard deviations of the results. The column 'LB' denotes the lower bound on the total delivery time for each depot location scenario.

Table 3.4. Case study results for three depot location scenarios (DLGB, DMID, DLAX) and three airport access policies (AAAP, LGB, LAX)

	DLGB Depot					DMID Depot					DLAX Depot				
No	LB	AAAP	LGB	LAX	ρ	LB	AAAP	LGB	LAX	ρ	LB	AAAP	LGB	LAX	ρ
1	6,843	7,012	7,111	7,899	63%	6,800	6,989	7,108	7,362	61%	6,860	7,088	7,137	7,605	82%
2	6,776	6,981	7,181	7,358	51%	6,787	6,979	7,138	7,248	55%	6,799	6,986	7,127	7,404	57%
3	6,823	6,994	7,056	7,817	73%	6,837	7,036	7,115	7,535	72%	6,809	6,923	7,054	7,627	47%
4	6,833	7,011	7,133	7,797	59%	6,821	7,088	7,198	7,474	71%	6,860	7,075	7,107	7,664	87%
5	6,769	6,961	7,177	7,426	47%	6,808	6,999	7,077	7,561	71%	6,833	6,984	7,246	7,604	37%
6	6,765	6,932	7,080	7,501	53%	6,823	6,981	6,997	7,801	91%	6,741	6,904	7,127	7,315	42%
7	6,852	7,061	7,125	7,661	77%	6,746	6,948	7,147	7,251	50%	6,839	7,058	7,164	7,440	67%
8	6,819	7,000	7,172	7,514	51%	6,812	6,932	7,065	7,635	47%	6,781	6,988	7,181	7,461	52%
9	6,763	6,961	7,123	7,390	55%	6,832	7,042	7,158	7,614	64%	6,793	7,013	7,158	7,375	60%
10	6,804	6,996	7,151	7,797	55%	6,749	6,955	7,205	7,299	45%	6,786	6,918	7,153	7,917	36%
Ave.	6,805	6,991	7,131	7,616	58%	6,802	6,995	24%	7,121	63%	6,810	7,075	7,107	7,664	57%
Sdev.	34	35	41	201.5	10%	32	48	2%	63	14%	38	66	50	178	18%

The AAAP policy dominates the single airport policy in all depot location scenarios and in all problem instances. The AAAP's impact on the total delivery time can be assessed through the following performance measure:

$$\rho = \frac{z_{AAAP} - LB}{\min\{z_{LGB}, z_{LAX}\} - LB}\% \tag{32}$$

where z_{AAAP}, z_{LGB}, and z_{LAX} correspond to the solutions of three airport policies.

The performance measure in (32) indicates the percentage total delivery time improvement of the AAAP policy over single airport policies. In the case of DLGB depot location, the AAAP policy improves the total delivery time performance on the average by 58%. The improvements range between 47% and 77%. Similarly, for the DMID depot location, the average improvement of AAAP is 63% and the range is between 45% and 91%. In the case of DLAX depot location, the average improvement is 57% and the range is between 36% and 87%. Overall, the AAAP's improvement over single airport policies is 59% on the average across all depot locations.

In Figures 5-7, we illustrate the routes identified for each depot location and airport policy for sample problem instances. These routes are turn-by-turn routes from the Google Maps API. The labels are " 1" for the location of depot, " 2" to " 16" for the locations of 15 customers, and LAX and LGB for the airports. The label in parenthesis denotes the order of visit by the vehicle. Each color route corresponds to a unique vehicle. For instance, in Figure 3.5a, the vehicle with blue color route starts its trip from the depot located in Compton (i.e. node 1), visits customers 5, 6, 8, 3, delivers loads to LAX, and returns to the depot. Accordingly, the customer 5 is labeled 5(1), customer 6 is labeled 6(2), and so forth. In all instances, at most three of the four vehicles are used, indicating absence of recourse flight usage.

In Figure 3.5, there are three vehicles in all airport policies. In Figures 6a and 7b only two vehicles are used since the third vehicle does not provide any additional benefit in terms of improving the total delivery cost. In Figures 5c, 6c and 7c, two vehicles deliver customer loads to both the LAX and LGB airports whereas the third vehicle visits only the LGB. In Figure 3.7a, the third vehicle is used to pick up and deliver the load of only customer 5.

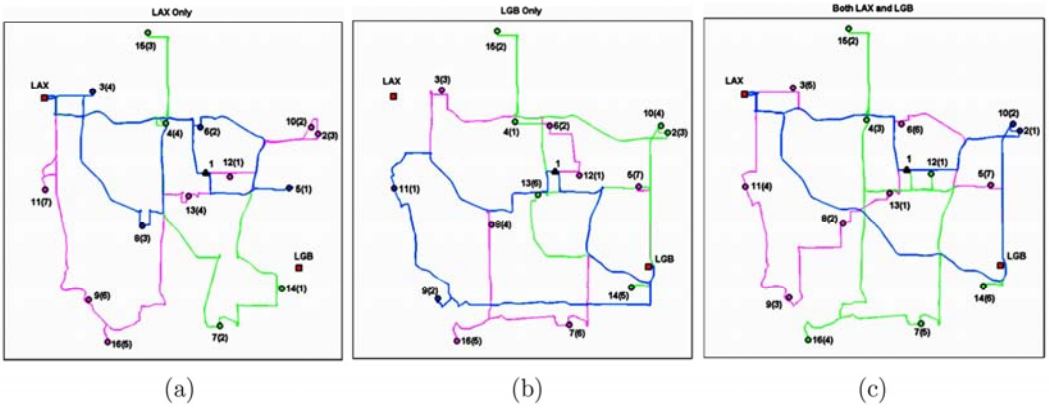


Figure 3.5. Routes for problem instance #10 with DMID depot

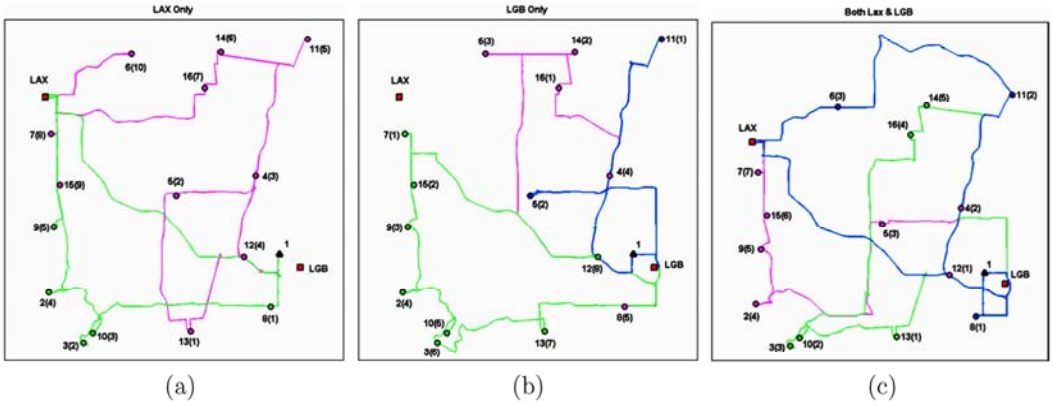


Figure 3.6. Routes for problem instance #1 with DLGB depot

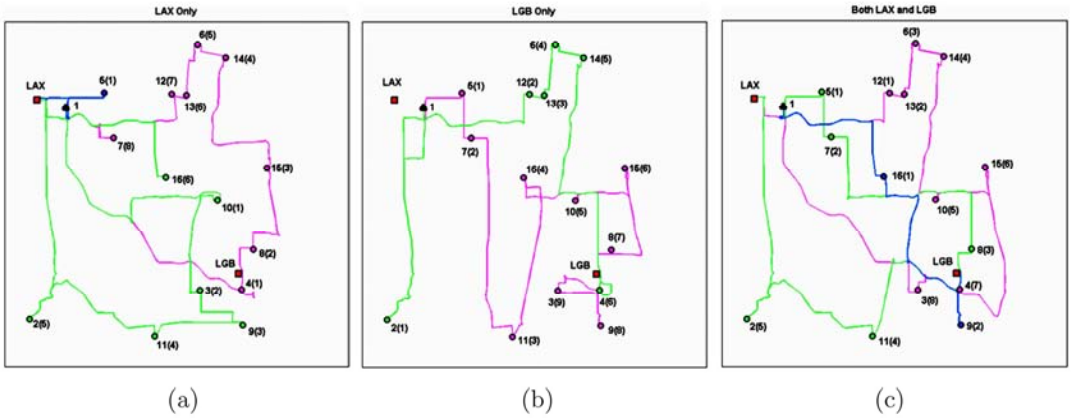


Figure 3.7. Routes for problem instance #6 with DLAX depot

3.6 Conclusion

We study a freight forwarder's operational implementation of AAAP in a MAR for air cargo transportation. The forwarder's AAAP implementation involves the task of selecting flight itineraries for a given set of heterogeneous air cargo customers, picking up their loads via a fleet of vehicles and then delivering to the airports in the region. The goal is to minimize the total cost of air and road transportation and service by simultaneously selecting the air cargo flight itinerary and scheduling pickup and delivery of multiple customer loads to the airport(s). We formulated a novel model (ATP-PDP) which extends the existing PDP models to address the case where the delivery cost is both destination and time dependent. This model is further

strengthened by preprocessing steps and special cuts. To overcome the computational complexity, we adapted an efficient solution method, SSS, based on Lagrangian decomposition. The SSS method overcomes the challenges associated with identical subproblems in standard Lagrangian decomposition and iteratively solves the ATP-PDP in parts. Since Lagrangian based methods, including SSS, can converge to a primal infeasible solution, we developed a modified variable target methodology for subgradient optimization. The integrated method, SSS-B-VTVM, converges to a primal feasible solution and the solution quality can be controlled by trading-off the quality with computational performance.

We conducted an experimental study to assess the optimality gap and CPU time performance of SSS-B-VTVM and compared with those of CPLEX. The results show that the SSS-B-VTVM yields near-optimal primal feasible solutions, i.e., on the average 3.4% optimality gap compared to 5.3% of CPLEX. Further, the SSS-B-VTVM is able to achieve this performance on the average about 87 times faster than CPLEX and more than thousand times faster for some problems. In addition, we have applied the modeling and solution methodology for a case study in a Southern California MAR and compared the AAAP performance with single airport policies considering various depot location and customer scenarios. The computational results indicate that the AAAP is able to realize savings in the order of 36% to 91% of the potential saving opportunities. This research can be extended in multiple directions. The proposed approach can be used to evaluate forwarder's locational decisions for its depot(s). Similarly, it can be used to assess the competitiveness of multiple airports in a MAR for air cargo shipments under various flight availability schedule scenarios.

Conclusion

In this research, we studied the air-cargo routing problem on both the air and road networks from the freight forwarders perspective.

In chapter two, we investigated the benefits of dynamic routing for the shipment of time-sensitive air cargo given a shipment criterion subject to the availability of flights, travel time and departure delay variability. We further examined the effect of real-time flight information accuracy on the dynamic routing performance. The contributions of this chapter to the literature are developing a novel dynamic routing model that accounts for the scheduled departures, the effect of stochastic travel times, and departure delays. In essence, the model extends the literature on dynamic routing problems subject to stochastic and time-dependent travel times with real-time information by incorporating the following aspects (1) the real-time information is inaccurate and (2) arcs on the network are schedule-based.

For the first aspect, we developed a novel departure delay estimation model based on the real-time announced delay information and historical delay distributions. For the second aspect, we formulated a dynamic routing Markov decision problem with a novel action space definition. The action space consists of not only the first choice of flight but also collectively all recourse flights at an airport node. A set of controlled experiments is conducted to investigate the effect of delay information accuracy, departure delay distribution, travel time variability and topology of flight network on the expected cost and delivery reliability. Lastly, in chapter two, we presented two case study applications using real flight network and departure delay

data. The results show that dynamic policy is able to not only improve the expected delivery performance but also increase the delivery reliability. Further, the departure delay information is critical for realizing the full potential of dynamic routing. However, the majority of the improvements can be attained even with little real-time information availability and accuracy.

There are multiple extensions possible of this study. First extension is the investigation of the effect of code sharing agreements among carriers on the dynamic routing benefits. Clearly, the code sharing increases the flight alternatives that can be selected and thus improves the performance. Another extension is to relax the assumption that flight itineraries are priced based on the individual legs and additive. Instead, the future research will consider more complex pricing of flight itineraries by accounting for the savings for booking multiple legs at a time which will restrict the flexibility of the dynamic routing.

In chapter three, we studied a freight forwarder's operational implementation of Alternative Access Airport Policy (AAAP) in a multi airport region for air-cargo transportation. The forwarder's AAAP implementation involves the task of selecting flight itineraries for a given set of heterogeneous air cargo customers, picking up their loads via a fleet of vehicles and then delivering them to the airports in the region. The goal is to minimize the total cost of air and road transportation and service by simultaneously selecting the air cargo flight itinerary and scheduling pickup and delivery of multiple customer loads to the airport(s). We formulated a novel model (ATP-PDP) which extends the existing pick up and delivery problem models to address the case where the delivery cost is both destination and time dependent. This model is further strengthened by preprocessing steps and special cuts.

To overcome the computational complexity, we adapted an efficient solution method, SSS, based on Lagrangian decomposition. The SSS method lessens the challenges associated with identical subproblems in standard Lagrangian decomposition and iteratively solves the ATP-PDP in parts. Since Lagrangian based methods, including SSS, can converge to a primal infeasible solution, we developed a modified variable target methodology for subgradient optimization. The integrated method, SSS-B-VTVM, converges to a primal feasible solution and the solution quality can be controlled by trading-off the quality with computational performance. We conducted an experimental study to assess the optimality gap and CPU time performance of SSS-B-VTVM and compared with those of CPLEX. The results show that the SSS-B-VTVM yields near-optimal primal feasible solutions, i.e., on the average 3.4% optimality gap compared to 5.3% of CPLEX. Further, the SSS-B-VTVM is able to achieve this performance on the average about 87 times faster than CPLEX and more than thousand times faster for some problems. In addition, we have applied the modeling and solution methodology for a case study in a Southern California MAR and compared the AAAP performance with single airport policies considering various depot location and customer scenarios. The computational results indicate that the AAAP is able to realize savings in the order of 36% to 91% of the potential saving opportunities. This research can be extended in multiple directions. The proposed approach can be used to evaluate forwarder's locational decisions for its depot(s). Similarly, it can be used to assess the competitiveness of multiple airports in a MAR for air cargo shipments under various flight availability schedule scenarios.

This research can be extended by studying the possible expansion and improvement to the solution algorithm for the AAAP mixed integer programming through alternative sub problem solving algorithms. Further research would test the applicability of the proposed SSS procedure for solving the general PDP and vehicle routing problems while addressing the issues associated with homogeneous subproblems. Lastly, one limitation of this research is the absence of economies of scale in cost when multiple customers' air cargo is loaded on the same flight. A fruitful future research avenue is to investigate the effect of such economic relations on the forwarder's implementation of the AAAP.

REFERENCES

Agnes, P., 2000. The "End of Geography" in Financial Services? Local Embeddedness and Territorialization in the Interest Rate Swaps Industry. *Economic Geography* 76 (4):347-366.

Ahmad Beygi, S., Cohn, A., Guan Y., Belobaba, P., 2008. Analysis of the potential for delay propagation in passenger airline networks. *Journal of Air Transport Management* 14(5), 221-236.

Airbus, 2010. Global Market Forecast 2010-2029, In: Airbus (Ed.), Toulouse.

Althen, W., Graumann, M. and Niedermeyer, M. ,2001. Alternative Wettbewerbsstrategien von Fluggesellschaften in der Luftfrachtbranche. Zeitschrift fur betriebswirtschaftliche Forschung (ZfbF) 53 Jg.: 420-441.

Anily, S., Bramel, J., 1999. Approximation algorithms for the capacitated traveling salesman problem with pickups and deliveries. *Naval Research Logestics* 46(6), 654-670.

Anily, S., Hassin, R., 1992. The swapping problem. *Networks* 22(4), 419-433.

Azadian, F., Murat, A.E., Chinnam, R.B., 2012. Dynamic routing of time-sensitive air cargo using real-time information. *Transportation Research Part E: Logistics and Transportation Review* 48(1), 355-372.

Azaron, A., Kianfar, F., 2003. Dynamic shortest path in stochastic dynamic networks: Ship routing problem. *European Journal of Opertions Research* 144(1), 138-156.

Bander, J.L., White, C.C., 2002. A Heuristic Search Approach for a Nonstationary Stochastic Shortest Path Problem with Terminal Cost. *Transportation Science* 36(2), 218-230.

Basar, G., Bhat, C., 2004. A parameterized consideration set model for airport choice: an application to the San Francisco Bay Area. *Transportation Research Part B: Methodological* 38(10), 889-904.

Berbeglia, G., Cordeau, J.-F., Gribkovskaia, I., Laporte, G., 2007. Static pickup and delivery problems: a classification scheme and survey. TOP 15(1), 1-31.

Berbeglia, G., Cordeau, J.-F., Laporte, G., 2010. Dynamic pickup and delivery problems. *European Journal of Operations Research* 202(1), 8-15.

Bertsekas, D.P., 2005. Dynamic programming and suboptimal control: A survey from ADP to MPC. *European Journal of Control* 11(4-5), 310-334.

Bertsekas, D.P., Tsitsiklis, J.N., 1991. An Analysis of Stochastic Shortest Path Problems. *Mathematics of Operations Research* 16(3), 580-595.

Boeing, 2008. World air cargo forecast 2008-2009.

Boeing, 2010. World air cargo forecast 2010-2011, In: Company, T.B. (Ed.), Seattle, WA.

Bookbinder, J.H., Matuk, T.A., 2009. Logistics and Transportation in Global Supply Chains: Review, Critique, and Prospects. *Informs Tutorial in Operations Research*:182-211.

Bowen, J., Leinbach, T., 2004. Market Concentration In The Air Freight Forwarding Industry. *Tijdschrift voor Economische en Sociale Geografie* 95 (2):174-188.

Bureau of Transportation Statistics, U.S. Dept. of Transportation, Last Accessed March 2010, http://www.transtats.bts.gov/

Chalasani, P., Motwani, R., 1999. Approximating capacitated routing and delivery problems. *SIAM Journal on Computing* 28(6), 2133.

Chatterji, G.B., Sridhar., B., 2005. National airspace System Delay Estimation Using Weather Weighted Traffic Counts, *AIAA Guidance, Navigation, and Control Conf. and Exh.*, SF, California.

Chayanupatkul, A., Hall, R., Epstein, D., 2004. Freight routing and containerization in a package network that accounts for sortation constraints and costs. METRANS Transportation Center, U. of Southern California.

Chew, E.P., Huang, H.C., Johnson, E.L., Nemhauser, G.L., Sokol, J.S., Leong, C.H., 2006. Short-Term Booking Of Air Cargo Space. *European Journal of Operational Research* 174 (3):1979-1990.

Conejo, A., Castillo, E., Minguez, R., Garcia-Bertrand, R., 2006. Decomposition techniques in mathematical programming engineering and science applications. Springer, Berlin; Heidelberg; New York.

Cordeau, J.-F., Laporte, G., 2007. The dial-a-ride problem: models and algorithms. *Annals of Operations Research* 153(1), 29-46.

Crainic, T.G., Gendreau, M., Potvin, J.Y., 2009. Intelligent freight-transportation systems: Assessment and the contribution of operations research. *Transportation Research Part C: Emerging Technologies* 17 (6):541-557.

Desrochers, M., Laporte, G., 1991. Improvements and extensions to the Miller-Tucker-Zemlin subtour elimination constraints. *Operations Research Letters* 10(1), 27-36.

Diana, M., Dessouky, M.M., 2004. A new regret insertion heuristic for solving large-scale dial-a-ride problems with time windows. *Transportation Research Part B: Methodological* 38(6), 539-557.

Doganis, R., Flying Off Course: The Economics of International Airlines, 3rd edition. London: Routledge, 2002.

Fisher, M.L., 2004. The Lagrangian Relaxation Method for Solving Integer Programming Problems. *Management Science* 50(12), 1861-1871.

Frederickson, G., 1978. Approximation Algorithms for Some Routing Problems. *SIAM Journal on Computing* 7(2), 178.

Fu, L., 2001. An adaptive routing algorithm for in-vehicle route guidance systems with real-time information. *Transportation Research Part B: Methodological* 35(8), 749-765.

Fu, L., Rilett, L.R., 1998. Expected shortest paths in dynamic and stochastic traffic networks. *Transportation Research Part B: Methodological* 32(7), 499-516.

Gao, S., Chabini, I., 2006. Optimal routing policy problems in stochastic time-dependent networks. *Transportation Research Part B: Methodological* 40(2), 93-122.

Geoffrion, A.M., 1974. Lagrangean relaxation for integer programming, In: Balinski, M.L. (Ed.), Approaches to Integer Programming. Springer Berlin Heidelberg, pp. 82-114.

Golob, T.F., Regan, A.C., 2000. Freight industry attitudes towards policies to reduce congestion. Transportation Research Part E: Logistics and Transportation Review 36 (1):55-77.

Green, K., Whitten, D., Inman, A., 2008. The impact of logistics performance on organizational performance in a supply chain context. *Supply Chain Management* 13 (4):317-327.

Hall, R.W., 1986. The Fastest Path through a Network with Random Time-Dependent Travel Times. *Transportation Science* 20(3), 182-188.

Hall, R.W., 2002. Alternative Access and Locations for Air Cargo. METRANS Transportation Center, University of Southern California.

Hansen, M., Bolic, T., 2001. Delay and Flight Time Normalization Procedures for Major Airports: LAX Case Study, *NEXTOR Research Report*., Berkeley, CA.

Held, M., Wolfe, P., Crowder, H.P., 1974. Validation of subgradient optimization. *Mathematical Programming* 6(1), 62-88.

Hellermann, R., 2006. *Capacity options for revenue management theory and applications in the air cargo industry*. Springer, Berlin; Heidelberg; New York.

Hernández-Pérez, H., Salazar-González, J.-J., 2004a. A branch-and-cut algorithm for a traveling salesman problem with pickup and delivery. *Discrete Applied Mathematics* 145(1), 126-139.

Hernández-Pérez, H., Salazar-González, J.-J., 2004b. Heuristics for the One-Commodity Pickup-and-Delivery Traveling Salesman Problem. *Transportation Science* 38(2), 245-255.

Hernández-Pérez, H., Salazar-González, J.-J., 2007. The one-commodity pickup-and-delivery traveling salesman problem: Inequalities and algorithms. *Networks* 50(4), 258-272.

Hernández-Pérez, H., Salazar-González, J.-J., 2009. The multi-commodity one-to-one pickup-and-delivery traveling salesman problem. *European Journal of Operations Research* 196(3), 987-995.

Jarrah, A.I.Z., Yu, G., Krishnamurthy, N., Rakshit, A., 1993. A Decision Support Framework for Airline Flight Cancellations and Delays. *Transportation Science* 27(3), 266-280.

Kasarda, J.D., Stephen, J.A., Mori, M., 2006. The Impact of Air Cargo on the Global Economy. Air Cargo Forum, The International Air Cargo Association, Calgary, Canada.

Kim , S., Lewis, M.E., White, C.C., 2005a. Optimal vehicle routing with real-time traffic information. *Intelligent Transportation Systems, IEEE Transactions on Intelligent Transportation Systems* 6(2), 178-188.

Kim , S., Lewis, M.E., White, C.C., 2005b. State Space Reduction for Non-stationary Stochastic Shortest Path Problems with Real-Time Traffic Information. *IEEE Transactions on Intelligent Transportation Systems* 6(3), 273 - 284.

Klier, T.H., Rubenstein, J.M., 2008. *Who really made your car? : restructuring and geographic change in the auto industry*. W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.

Kohl, N., Madsen, O.B.G., 1997. An Optimization Algorithm for the Vehicle Routing Problem with Time Windows Based on Lagrangian Relaxation. *Operations Research* 45(3), 395-406.

Laporte, G., 1992. The vehicle routing problem: An overview of exact and approximate algorithms. *European Journal of Operation Research* 59(3), 345-358.

Laporte, G., 2009. Fifty Years of Vehicle Routing. *Transportation Science* 43(4), 408-416.

Lim, C., Sherali, H., 2006. Convergence and Computational Analyses for Some Variable Target Value and Subgradient Deflection Methods. *Computational Optimization and Applications* 34(3), 409-428.

Long, D., Wingrove, E., Lee, D., Gribko, J., Hemm, R., Kostiuk, P., 1999. A Method for Evaluating Air Carrier Operational Strategies and Forecasting Air Traffic with Flight Delay. LMI, McLean, VA.

Loo, B.P.Y., 2008. Passengers' Airport Choice within Multi-Airport Regions (MARs): Some Insights from a Stated Preference Survey at the Hong Kong International Airport. *Journal of Transportation Geography* 16(2), 117-125.

Manuj, I., Mentzer, J., 2008. Global Supply Chain Risk Management. *Journal of Business Logistics* 29 (1):133-155.

Margreta , M., Ford , C., Dipo , M.A., 2009. U.S. Freight on the Move: Highlights from the 2007 Commodity Flow Survey Preliminary Data. US Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics

Miguel Andres, F., 2007. Analysis of the efficiency of urban commercial vehicle tours: Data collection, methodology, and policy implications. *Transportation Research Part B: Methodological* 41(9), 1014-1032.

Miller-Hooks, E.D., Mahmassani, H.S., 1998. Least possible time paths in stochastic, time-varying networks. *Computers & Operations Research* 25(12), 1107-1125.

Miller-Hooks, E.D., Mahmassani, H.S., 2000. Least Expected Time Paths in Stochastic, Time-Varying Transportation Networks. *Transportation Science* 34(2), 198-215.

Mueller, E., Chatterji, G., 2002. Analysis of aircraft arrival and departure delay characteristics, *In Proceedings of the AIAA ATIO Conference*, Los Angeles, CA, October 1-3, 2002.

Nsakanda, A.L., Turcotte, M., Diaby, M., 2004. Air cargo operations evaluation and analysis through simulation, *Proceedings of the 36th conference on Winter simulation*. Washington, D.C., 1790-1798.

Opasanon, S., Miller-Hooks, E., 2001. Least Expected Time Hyperpaths in Stochastic, Time-Varying Multimodal Networks. *TRB: Journal of the Transportation Research Board* 1771(1), 89-96.

Parragh, S., Doerner, K., Hartl, R., 2008a. A survey on pickup and delivery problems (Part I: Transportation between customers and depot). *Journal für Betriebswirtschaft* 58(1), 21-51.

Parragh, S., Doerner, K., Hartl, R., 2008b. A survey on pickup and delivery problems (Part II: Transportation between pickup and delivery locations). *Journal für Betriebswirtschaft* 58(2), 81-117.

Psaraftis, H.N., Tsitsiklis, J.N., 1993. Dynamic Shortest Paths in Acyclic Networks with Markovian Arc Costs. *Operations Research* 41(1), 91-101.

Razzaque, M.A., Sheng, C.C., 1998. Outsourcing of logistics functions: a literature survey. *International Journal of Physical Distribution & Logistics Management* 28 (2):89-107.

Rupp, N.G., Holmes, G.M., 2006. An Investigation into the Determinants of Flight Cancellations. *Economica* 73(292), 749-783.

Schrank, D., Lomax T., 2011. The 2011 Urban Mobility Report. Annual report, *Texas Transportation Institute*. The Texas A&M University System.

Solomon, M.M., 1987. Algorithms for the Vehicle Routing and Scheduling Problems with Time Window Constraints. *Operations Research* 35(2), 254-265.

Tang, C.-H., Yan, S., Chen, Y.-H., 2008. An Integrated Model and Solution Algorithms for Passenger, Cargo, and Combi flight scheduling. *Transportation Research Part E*, 44(6), 1004-1024.

Thomas, B.W., White III, C.C., 2007. The dynamic shortest path problem with anticipation. *European Journal Operational Research* 176(2), 836-854.

Tien, S., Ball, M., Subramanian, B., 2008. Constructing a Passenger Trip Delay Metric: An Aggregate-level Approach, *NEXTOR Technical Report*, 1-7.

Toth, P., Vigo, D., 2001. The vehicle routing problem. Society for Industrial and Applied Mathematics, Philadelphia.

Tu, Y., Ball, M.O., Jank, W.S., 2008. Estimating Flight Departure Delay Distributions—A Statistical approach with long-term trend and Short-Term Pattern. *Journal of the American Statistical Association* 103(481), 112-125.

US Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics 2006. Freight in America. Washington, DC.

US Department of Transportation, 2007. Transportation Statistics Annual Report 2007. Washington, DC.

US Department of Transportation, 2009. Pocket Guide to Transportation Washington, DC.

Waller, S.T., Ziliaskopoulos, A.K., 2002. On the online shortest path problem with limited arc cost dependencies. *Networks* 40(4), 216-227.

Wang, D., Sherry, L., 2007. Trend Analysis of Airline Passenger Trip Delays, In the Proceedings of the 86th Transportation Research Board Annual Meeting, Washington D.C.

Yan, S., Chen, S.-C., Chen, C.-H., 2006. Air Cargo Fleet Routing and Timetable Setting with Multiple On-Time Demands. *Transportation Research Part E* 42(5), 409-430.

Zhai, Q., Guan, X., Cui, J., 2002. Unit commitment with identical units successive subproblem solving method based on Lagrangian relaxation. Power Systems, IEEE Transactions on 17(4), 1250-1257.

Zhao, X., Luh, P.B., Wang, J., 1999. Surrogate Gradient Algorithm for Lagrangian Relaxation. *Journal of Optimization Theory and Applications* 100(3), 699-712.

Ziliaskopoulos, A., Wardell, W., 2000. An intermodal optimum path algorithm for multimodal networks with dynamic arc travel times and switching delays. *European Journal of Operations Research* 125(3), 486-502.

ABSTRACT

AN INTEGRATED FRAMEWORK FOR FREIGHT FORWARDERS: EXPLOITATION OF DYNAMIC INFORMATION FOR MULTIMODAL TRANSPORTATION

by

FARSHID AZADIAN

August 2012

Advisor: Dr. Alper E. Murat

Co-Advisor: Dr. Ratna Babu Chinnam

Major: Industrial Engineering

Degree: Doctor of Philosophy

Advent of real-time information broadcasting technologies, growth in demand for air-cargo, and increased congestion and variability on air-road network, are the main forces compelling today's air-freight forwarders to improve their operational decision-making to be more competitive and responsive to needs of customers. This research studies the air-cargo transportation on both road (short-haul) and air (long haul) network from the perspective of a mid-size freight forwarder.

We develop a routing algorithm for congestion avoidance on air-network based on historical data and introduce an innovative approach to incorporate real-time information to enable dynamic routing of cargo on a stochastic air-network. In the road network, we introduce a new class of pickup and delivery problems to carry out the customer load pickups, fleet

management, cargo-to-flight assignments, and airport deliveries in a multiple airport region under alternative access airport policy.

The main contributions of this research to the air-cargo literature are the study of the value of real-time information and introduction of the concept of dynamic air-cargo routing. In addition, this is the first study that provides an operational framework to implement the alternative access airport policy. This research also contributes to operations research and logistics literature by introducing a new class of pickup and deliveries with time-sensitive and pair-dependent cost structure. It also contributes an innovative algorithm based on successive subproblem solving for Lagrangian decomposed mixed integer programming that shows to be efficient in obtaining near optimal solutions in reasonable time.

The performances of the algorithms presented in this research are tested through experimental and real-world case studies. The results demonstrate that dynamic routing with real-time information can dramatically improve delivery reliability and reduce expected cost on the air-network. Moreover, they confirm that alternative access airport policy can greatly enhance a forwarder's options and reduce the operational and service costs while improving the service levels.

AUTOBIOGRAPHICAL STATEMENT

Farshid Azadian received his PhD degree in Industrial Engineering from Wayne State University (Detroit, Michigan) in 2012. He holds B.Sc. in Textile Engineering and M.Sc. in Engineering Management from Amirkabir University (Tehran, Iran). He has worked in the textile industry as an engineering consultant for years and worked closely with the Iran National Institute of Standards as an expert and representative of the Association of Textile Industries of Iran. He also ran his own business that designed and produced special textile machinery.

He has published in Transportation Research Part E: Logistics and Transportation Review and presented at numerous conferences including INFORMS, METRANS, IIE Annual Meetings in addition to several regional transportation conferences.

His research interests are in planning and modeling of freight transportation logistics systems for global supply chains. He studied in particular the application of advanced operations research methodologies for freight routing under Intelligent Transportation Systems, complex pickup and delivery problems, and multiple airport region operational planning and scheduling.