Dynamic routing of time-sensitive air cargo using real-time information

Farshid Azadian, Alper E. Murat*, Ratna Babu Chinnam

Department of Industrial and Systems Engineering, Wayne State University, 4815 Fourth St., Detroit, MI 48201, USA

ARTICLE INFO

Article history:
Received 21 May 2010
Received in revised form 16 May 2011
Accepted 24 May 2011

Keywords:
Air cargo
Transportation
Dynamic routing
Flight departure delay
Freight forwarder

ABSTRACT

The route planning of time-sensitive air-cargo is becoming more important with the growing air-network congestion and delays. We consider a freight forwarder's routing of a time-sensitive air-cargo in the presence of real-time and historical information regarding flight availability, departure delays and travel times. A departure delay estimation model is developed to account for real-time information inaccuracy. A novel Markov decision model is formulated and solved with online backward induction. Through synthetic experiments and case studies, we demonstrate that dynamic routing with real-time information can improve delivery reliability and reduce expected cost.

© 2011 Elsevier Ltd. All rights reserved.

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 US 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 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 $60–100 K/h in a medium-sized finished vehicle assembly plant.³ In response, the OEMs often resort to expedited shipping by either chartering a freighter or a cargo helicopter for time-definite delivery which can cost $100 Ks 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 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 affordable 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).

² For instance, the shipping rate for an LD2 container with dimensions (61.5 × 60.4 × 64) inches and weight 1228 lbs from Cleveland to Seattle on 22 March 2010 with UPS is $49.9K–$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.
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. 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 Ziliaskopoulous (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 Ziliaskopoulous (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). Ziliaskopoulous 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 Ziliaskopoulous 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.

3. Dynamic air cargo routing

Let \( G = (N,A) \) 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 \( A_i \subseteq A \) denote the set of flights between airports \( n' \) to \( n'' \) where \( l = (n', n'') \), then \( i \in A_i \) 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.

The flight arcs have three parameters affecting the flight selection decisions which are illustrated through the time sequence depiction in Fig. 1. First parameter is the stochastic travel time of arc \( i (\tau_i) \), 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 \( (\theta_i) \). Node arrival prior to \( \theta_i \) 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 \( \theta_i \). Alternatively, \( \theta_i \) 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 \( (\delta_i) \) corresponding to an uncontrollable waiting time at the origin node of an arc \( (\text{flight } i) \) 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 \( \delta_i \). Accordingly, if the flight has not departed past the scheduled departure time, the actual departure time depends on \( \delta_i \), 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 \( (\delta_i) \) 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 depa-

\[ \text{DepDelayMinutes} = \text{DepartureTime} - \text{ScheduledDepartureTime} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]

\[ \text{TaxiOut} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoAcceptance} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]

\[ \text{TaxiOut} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoAcceptance} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]

\[ \text{TaxiOut} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoAcceptance} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]

\[ \text{TaxiOut} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoAcceptance} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]

\[ \text{TaxiOut} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoAcceptance} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]

\[ \text{TaxiOut} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoAcceptance} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]

\[ \text{TaxiOut} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoAcceptance} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]

\[ \text{TaxiOut} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoAcceptance} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]

\[ \text{TaxiOut} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoAcceptance} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]

\[ \text{TaxiOut} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoAcceptance} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]

\[ \text{TaxiOut} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoAcceptance} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]

\[ \text{TaxiOut} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoArrivalToAirport} \]

\[ \text{AirCargoAcceptance} \]

\[ \text{ScheduledGateClosing} \]

\[ \text{ActualGateClosing} \]

\[ \text{DepartureDelay} \]

\[ \text{TravelTime} \]

\[ \text{AirTime} \]

\[ \text{TaxiIn} \]
tures 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 $-3.8, -3.8, -4.0, -5.4, -4.0, -4.4, \text{ and } -3.7 \text{ min}$ which are negligible compared to the average late departure delays of $31.9, 32.4, 33.2, 48.5, 26.2, 29.8, \text{ and } 32.7 \text{ min}$ 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 min for shipments under 100 lbs and 1–2 h for larger shipments depending on the carrier and airport. The cut-off times for transfers can range between 30 min 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 (Nakanda 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. Nakanda et al. (2004) report on terminal cut-off time of 45 min 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 landing 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. 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 ($\delta$) as excessive delays. 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 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.

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 $z_i$ denote the “on-time” departure probability of flight $i$, i.e. $\psi(\delta_i = 0) = z_i$ and $\delta_i$ 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 ($\xi$) such that after $\xi$ the flight is considered as unavailable. Provided that $\xi$ is chosen sufficiently large, any flight that is delayed longer than $\xi$ 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 $\delta_i$ is set to 0. Accordingly, we have $(\delta_i) = 1 - z_i$ for $0 < \delta_i < \epsilon$ and $\psi(\delta_i) = 0$ for $\delta_i < 0, \delta_i \geq 0$.

At any given time $t$, the decision maker has access to real-time information on the departure delay $\delta_i(t)$ 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

\footnote{The US 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 US 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/acp-ipc-menu-eng.html).}

\footnote{Delays longer than a threshold typically lead to cancellation or other recovery methods, rather than delays subsequent flights (AhmadBeygi et al., 2008).}

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. $P(\delta_i = \hat{\delta}_i|\hat{\delta}_i)$ corresponds to the case of perfect information. However, once the real-time announcement ($\hat{\delta}_i$) on the departure delay ($\delta_i$) 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. $P(\delta_i = \hat{\delta}_i|\hat{\delta}_i) = 0$ 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 $P(\delta_i = \hat{\delta}_i|\hat{\delta}_i)$ 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, $\hat{\delta}_i \in (v^i, v^{i+1})$ for $r = 1, \ldots, m$, where $v^i$ and $v^{i+1}$ define the upper and lower bound on the departure delay for interval $r$, respectively. Note that $v^i \leq v^i_r$ for $r_1 \leq r_2$ and $v^i_0 = 0$ and $v^m = \zeta$.

The delay intervals $(v^i, v^{i+1})$ 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. $P(v^i \leq \delta_i \leq v^{i+1}, v^i \leq \delta_i \leq v^{i+1})$. 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 $P_r = P(v^r \leq \delta_i \leq v^{r+1})$. Given that the announced delay information is tail conditionally accurate, we have $P(\delta_i = 0)$ for $\delta_i \leq v^i_r$ and $\delta_i \leq \delta_i$ where $\delta_i \in (v^r, v^{r+1})$. 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 $\hat{\delta}_i \in (v^i, v^{i+1})$ for a given $1 \leq r \leq m$ as,

$$P(\delta_i | v^r \leq \delta_i \leq v^{r+1}) = \sum_{r' \geq r} P(\delta_i | v^r \leq \delta_i \leq v^{r+1}) P(v^r \leq \delta_i \leq v^{r+1} | v^r \leq \delta_i \leq v^{r+1})$$

For $r' > r$, the last term in (1) corresponds to the case $r' = r$. From the Bayes’ rule, we have,

$$P(v^r \leq \delta_i \leq v^{r+1} | v^r \leq \delta_i \leq v^{r+1}) = \frac{P(v^r \leq \delta_i \leq v^{r+1}, v^r \leq \delta_i \leq v^{r+1})}{P(v^r \leq \delta_i \leq v^{r+1})}$$

for $r' > r$, (2)

$$P(\delta_i | v^r \leq \delta_i \leq v^{r+1}) = \frac{P(\delta_i \leq \delta_i \leq v^{r+1}, v^r \leq \delta_i \leq v^{r+1})}{P(v^r \leq \delta_i \leq v^{r+1})}$$

for $r' < r$, (3)

We calculate $P(\delta_i | v^r \leq \delta_i \leq v^{r+1})$ in (1) for $r' > r$ and $\delta_i \in (v^r, v^{r+1})$ as,

$$P(\delta_i | v^r \leq \delta_i \leq v^{r+1}) = \frac{P(\delta_i, v^r \leq \delta_i \leq v^{r+1})}{P(v^r \leq \delta_i \leq v^{r+1})} = \frac{\psi(\delta_i)}{\psi(v^{r+1}) - \psi(v^r)}$$

(4)

For $\delta_i(v^r, v^{r+1})$, we have $P(\delta_i | v^r \leq \delta_i \leq v^{r+1}) = 0$.

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

$$P(\delta_i | \hat{\delta}_i \leq \delta_i \leq v^{i+1}) = \frac{\psi(\hat{\delta}_i)}{\psi(v^{i+1}) - \psi(\hat{\delta}_i)}$$

(5)

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

Hence, the conditional delay distribution is as follows:

$$P(\delta_i | v^r \leq \delta_i \leq v^{r+1}) = \psi(\delta_i) \left( \frac{P(\hat{\delta}_i \leq \delta_i \leq v^{i+1} | v^r \leq \delta_i \leq v^{r+1})}{\psi(v^{i+1}) - \psi(\hat{\delta}_i)} + \sum_{r' < r} \frac{P(v^r \leq \delta_i \leq v^{r+1} | v^r \leq \delta_i \leq v^{r+1})}{\psi(v^{r+1}) - \psi(v^r)} \right)$$

(6)

For $m = 1$ (e.g. $v^1 = 0$ and $v^2 = \zeta$) 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 $v^r \leq \delta_i \leq v^{r+1}$ as:

$$q_i(\delta_i | \hat{\delta}_i) = P(\hat{\delta}_i = t - \delta_i | v^r \leq \delta_i \leq v^{r+1})$$

(7)

Note that we suppress $t$ for $q_i(\hat{\delta}_i, \delta_i)$ 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 time $t$ as:
\[ Q_i(t, \hat{\delta}_i) = \sum_{t' = t}^{\delta_i} q_i(t' - \hat{\theta}_i, \hat{\delta}_i). \]

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

### 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 \( p = (i_1, i_2, \ldots, i_k) \), \( k = 1, 2, \ldots, K \) is defined as sequence of flights such that \( i_k \in A_{i_{k-1}} \) where \( i_k \equiv (n^o_k, n^d_k) \) and \( n^o_k \equiv n^o_d \). Note that \( n^o_k \equiv n^o_d \) and \( n^d_k \equiv n^d_d \). Let \( p_5 \) 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 \( \hat{\theta}_i \leq t \) 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 \( t_k \in [1, \ldots, T] \). 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 \( \Omega_k = (n_k, t_k, \hat{\Lambda}_k, z_k) \). 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., \( \hat{\Delta}_k = \{\hat{\delta}_i \mid \forall i \in A \} \), and static route indicator, i.e., \( z_k = 1 \) if the flights are selected from the static route until stage \( k \) and \( z_k = 0 \) 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 \( \Omega_k \) as the set of all orderings of all the available flights scheduled to depart from airport \( n_k \) denoted with \( P(n_k, t_k) \). Denote 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 \( \pi = (\pi_1, \pi_2, \ldots, \pi_k) \) be the policy of the routing and is composed of policies for each of the \( K - 1 \) decision stages. For a given state \( \Omega_k \), the policy \( \pi - k(\Omega_k) \) is a deterministic Markov policy that chooses an ordering of flights departing from node \( n_k \), i.e., \( \pi_k(\Omega_k) \in P(n_k, t_k) \). Therefore, the expected total cost for a given policy vector \( \pi = (\pi_1, \pi_2, \ldots, \pi_k) \) is as follows:

\[
F(n_k, t_1, \hat{\Lambda}_1, 1|\pi) = E_{\hat{\delta}_1} \left\{ g_{k-1}(\Omega_{k-1}) + \sum_{k=1}^{K} g_k(\Omega_k, \pi_k(\Omega_k), \hat{\Delta}_k) \right\},
\]

where \( (n_k, t_1, \hat{\Lambda}_1, 1) \) is the starting state of the system and the \( \Delta_k \) is the actualized departure delay vector at stage \( k \). The single stage cost \( g_k(\Omega_k, \pi_k(\Omega_k), \hat{\Delta}_k) \) is cost of the flight ordering selected given the actualized departure delay \( \Delta_k \). 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 \( F(n_k, t_1, \hat{\Lambda}_1, 1|\pi) \) over the policy vector \( \pi = (\pi_1, \pi_2, \ldots, \pi_k) \) as follows:

\[
F'(n_k, t_1, \hat{\Lambda}_1, 1) = \min_{\pi = (\pi_1, \pi_2, \ldots, \pi_k)} F(n_k, t_1, \hat{\Lambda}_1, 1|\pi).
\]

The corresponding optimal policy is then,

\[
\pi^* = \min_{\pi = (\pi_1, \pi_2, \ldots, \pi_k)} F(n_k, t_1, \hat{\Lambda}_1, 1|\pi).
\]

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

\[
F'(\Omega_k) = \min_{\pi_k} E_{\hat{\delta}_k} \left\{ g_k(\Omega_k, \pi_k(\Omega_k), \hat{\Delta}_k) + F'(\Omega_{k+1}) \right\} \quad \forall k = 1, \ldots, K.
\]

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 out-bound flight from \( n_k \) can be in either available or unavailable state at \( t_k \). Let \( \gamma_i \) denote the steady-state probability that flight \( i \) is not cancelled or diverted and has also sufficient capacity. Then the probability that the \( i^{th} \) flight is available at \( t_k \) and can depart with the air-cargo, \( P(i_k, \hat{\delta}_i) \), is calculated as follows,

\[ \gamma_i \]
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 \( P_i(t_k, \hat{\delta}_i) = 1 - P_i(t_k, \delta_i) \). Note that Eq. (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 \( \Omega_k \), \( C_i(\Omega_k) \) 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),

\[
C_i(\Omega_k) = \begin{cases} 
0 & \text{if } i = p_i(k) \text{ and } z_k^i = 1, \quad \text{(Case I)}, \\
L_i & \text{if } z_k^i = 0, \quad \text{(Case II)}, \\
H + L_i & \text{if } i \neq p_i(k) \text{ and } z_k^i = 1, \quad \text{(Case III)},
\end{cases}
\]

where \( p_i(k) \) 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 \( L_i < L_s \), i.e., for case (II).

\( C_i(\Omega_k) \) 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).

Therefore, the online backward induction approach uses stationary departure delay information \( \hat{\delta}_k \) is calculated as,

\[
f_i\left( n_k, t_k, \hat{\delta}_k, z_k^i \right) = C_i(\Omega_k) + \sum_{\tau_i} \sum_{z_k^i} q_i(\hat{\delta}_i, \delta_i) P(\tau_i | \delta_i) F^h(n_{k-1}, \delta_i + \hat{\delta}_i + \tau_i, \hat{\delta}_k+1, z_k^{i+1}), \quad (15)
\]

where \( C_i \) stage \( k \) cost of choosing flight \( i, \hat{\delta}_i \) is the stochastic departure delay satisfying \( \hat{\delta}_i \leq \delta_i \leq \zeta_i \), where \( \delta_i + \hat{\delta}_i + \tau_i \) is the stochastic arrival time to \( n_{k-1} \). \( q_i(\hat{\delta}_i, \delta_i) \) 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 \( z_k^{i+1} = 1 \) if \( z_k^i = 1 \) and \( i = p_i(k) \) and \( z_k^i = 0 \) otherwise.

Let \( \pi_k \in P(n_k, t_k) \) 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 \( f_{ji} \left( n_k, t_k, \Delta_k, z_k^j \right) < f_{ij} \left( n_k, t_k, \Delta_k, z_k^i \right) \), 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 \( \pi_k \) as \( P_{ij}(t_k, \hat{\delta}_j) \left[ \prod_{(j < j') \, \pi_{k}(t_k, \hat{\delta}_{j'})} P_{ij'}(t_k, \hat{\delta}_{j'}) \right] \), which considers that all higher ranked flights are unavailable. Then, the expected cost-to-go of the flight ordering \( \pi_k \) at time \( t_k \) with departure delay information \( \hat{\delta}_k \) from node \( n_k \) is,

\[
F\left( n_k, t_k, \hat{\delta}_k, z_k^i | \pi_k \right) = M \prod_{i \in \pi_k} P_i(t_k, \hat{\delta}_i) + \sum_{(j \in \pi_k)} P_{ij}(t_k, \hat{\delta}_j) \left[ \prod_{(j < j') \, \pi_{k}(t_k, \hat{\delta}_{j'})} P_{ij'}(t_k, \hat{\delta}_{j'}) \right] f_{ji} \left( n_k, t_k, \hat{\delta}_k, z_k^j \right).
\]

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

\[
\sum_{(j \in \pi_k)} P_{ij}(t_k, \hat{\delta}_j) \left[ \prod_{(j < j') \, \pi_{k}(t_k, \hat{\delta}_{j'})} P_{ij'}(t_k, \hat{\delta}_{j'}) \right] C_i(\Omega_k) + M \prod_{i \in \pi_k} P_i(t_k, \hat{\delta}_i) + \sum_{(j \in \pi_k)} P_{ij}(t_k, \hat{\delta}_j) \left[ \prod_{(j < j') \, \pi_{k}(t_k, \hat{\delta}_{j'})} P_{ij'}(t_k, \hat{\delta}_{j'}) \right] f_{ji} \left( n_k, t_k, \hat{\delta}_k, z_k^j \right) \forall k = 1, \ldots, K, \quad (17)
\]

The backward induction approach is often used to solve \( F(\Omega_k) \) for an optimal policy \( \pi^* \) offline, i.e. before the trip starts. However, the size of the state space is \( O(2^{|N|T/\delta t}|M^k|) \) 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 h discretized with 5 min time intervals). Whereas we have \( 2^{|N|T/\delta t}|M^k| = 4320 \) states for \( m = 1 \), the size of the state-space grows to \( 2^{|N|T/\delta t}|M^k| \approx 4 \times 10^{27} \) 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 \( \hat{\delta}_k \) 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 committed at the time of node arrival \( t_k \), the only departure delay information available and used for this decision is \( \hat{\delta}_k \). Therefore, the online backward induction approach uses stationary departure delay information \( \hat{\delta}_k \) 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 \( \hat{\delta}_{k+1} \).
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 Fig. 2 together with the problem parameters. The parameters for every flight are the probability of on-time departure (\( \alpha \)), the expected departure delay (\( \mu \)) if the flight is delayed, the scheduled departure time (\( \theta \)), and the mean (\( \tau \)) and standard deviation (\( \sigma \)) 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 Fig. 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. \( S_1, S_2, S_3, S_4, S_4, S_6 \), 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.

Fig. 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 \( \rho \) as dynamic policy's improvement over static policy as a percentage of total possible under perfect information.

\[
\rho = 100 \times \frac{F^*_s(\Omega_1) - F^*(\Omega_1)}{F^*_p(\Omega_1) - F^*_y},
\]

where \( F^*_s(\Omega_1) \), \( F^*(\Omega_1) \) and \( F^*_y \) 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 \( \mathcal{N}_1, \mathcal{N}_2 \), we also sample flight travel time \( \tau \). 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.

Fig. 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 \( \mathcal{N}_4 \), Fig. 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 \( S_{3-6} \) is the availability of flight \( 8 \), in essence, provides the dynamic

![Fig. 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 (\( \mathcal{N}_4 \)).](image-url)
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.

Fig. 4a indicates that the rate of improvement with increased accuracy is diminishing for \( {\mathcal{V}}_0 \) to \( {\mathcal{V}}_4 \). 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. \( {\mathcal{V}}_0 \) versus \( {\mathcal{V}}_1 \) and \( {\mathcal{V}}_2 \). The effect of delay distribution is illustrated in Fig. 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 \( C \).

Another important performance measure for the shippers and freight forwarders is the delivery reliability, i.e., the percentage of shipments arriving on time. Fig. 5 shows the conditional expected tardiness for \( {\mathcal{V}}_0, {\mathcal{V}}_1, {\mathcal{V}}_2, \) and \( {\mathcal{V}}_4 \) at different levels of announced delay accuracy and delivery due dates. Fig. 5a and b illustrate that increasing information accuracy improves tardiness performance. Case for \( {\mathcal{V}}_3 \) is similar to \( {\mathcal{V}}_0 \), but the difference in static and dynamic policy tardiness is more remarkable. In the case of \( {\mathcal{V}}_4 \), the conditional expected tardiness of two policies with no real-time information is similar and insensitive to the delay distribution (Fig. 5c). Further, with the increased level of announced delay accuracy, the effect of the delay distribution on the conditional expected tardiness diminishes.

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.

5.1. Estimation of flight departure delay and travel time

The flight \( i \) is delayed with probability \( (1 - \pi_i) \) and the corresponding departure delay \( (\delta_i) \) has non-negative and continuous probability density \( \phi(\delta_i) \) and cumulative density \( \Phi(\delta_i) \). 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, weather, and 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 (\( \pi_i \)) and the cancellation and diversion percentages (\( \gamma_i \)). We considered those flights with delays in excess of \( \zeta = 90 \) min 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. Fig. 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). Fig. 7 illustrates the frequency plots of the data in two clusters of LGA-ORD flights.

Using the same database, we also estimated the travel time distributions conditional on the departure delay. Fig. 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.
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.

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). 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 problem parameters as shown in Table 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.

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 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 \( M = 1400 \) min. Fig. 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.

Fig. 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 pol-

---

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.
icy’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. Fig. 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 (Fig. 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 min, the dynamic policy can attain deliveries as early as at 792 min.

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.

5.1.2. Case Study II: La Guardia Airport (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 presents problem parameters and distributions.

Fig. 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
Fig. 8. Travel time distributions for different announcement accuracy levels (LGA-SEA case study).

Fig. 9. LGA to SEA case study, flight path frequency (a) and improvement (b).

(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. Fig. 10b illustrates that the expected performance improvement of dynamic routing policy over static policy. Clearly, even for limited information accuracy, dynamic can realize 20% perfor-
mance 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.

### Table 2

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

<table>
<thead>
<tr>
<th>Flight label</th>
<th>Carrier</th>
<th>Flight num.</th>
<th>From</th>
<th>To</th>
<th>Sch. dep.</th>
<th>Departure delay</th>
<th>Duration (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean (min)</td>
<td>On-time (%)</td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>DL</td>
<td>19</td>
<td>LGA</td>
<td>ATL</td>
<td>6:00</td>
<td>42</td>
<td>45</td>
</tr>
<tr>
<td>2</td>
<td>DL</td>
<td>533</td>
<td>LGA</td>
<td>ATL</td>
<td>7:00</td>
<td>36</td>
<td>77</td>
</tr>
<tr>
<td>3</td>
<td>DL</td>
<td>423</td>
<td>LGA</td>
<td>ATL</td>
<td>8:00</td>
<td>14</td>
<td>83</td>
</tr>
<tr>
<td>4</td>
<td>NW</td>
<td>541</td>
<td>LGA</td>
<td>DTW</td>
<td>6:00</td>
<td>6</td>
<td>43</td>
</tr>
<tr>
<td>5</td>
<td>NW</td>
<td>533</td>
<td>LGA</td>
<td>DTW</td>
<td>7:25</td>
<td>21</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>NW</td>
<td>172</td>
<td>LGA</td>
<td>MEM</td>
<td>6:30</td>
<td>14</td>
<td>84</td>
</tr>
<tr>
<td>7</td>
<td>NW</td>
<td>509</td>
<td>LGA</td>
<td>MEM</td>
<td>11:00</td>
<td>22</td>
<td>77</td>
</tr>
<tr>
<td>8</td>
<td>NW</td>
<td>1424</td>
<td>LGA</td>
<td>MSP</td>
<td>6:00</td>
<td>28</td>
<td>51</td>
</tr>
<tr>
<td>9</td>
<td>NW</td>
<td>525</td>
<td>LGA</td>
<td>MSP</td>
<td>8:15</td>
<td>26</td>
<td>53</td>
</tr>
<tr>
<td>10</td>
<td>US</td>
<td>1455</td>
<td>LGA</td>
<td>CLT</td>
<td>6:30</td>
<td>13</td>
<td>93</td>
</tr>
<tr>
<td>11</td>
<td>US</td>
<td>1021</td>
<td>LGA</td>
<td>CLT</td>
<td>8:10</td>
<td>18</td>
<td>93</td>
</tr>
<tr>
<td>12</td>
<td>DL</td>
<td>1915</td>
<td>ATL</td>
<td>DWF</td>
<td>8:40</td>
<td>15</td>
<td>62</td>
</tr>
<tr>
<td>13</td>
<td>DL</td>
<td>1917</td>
<td>ATL</td>
<td>DWF</td>
<td>9:45</td>
<td>8</td>
<td>41</td>
</tr>
<tr>
<td>14</td>
<td>DL</td>
<td>1919</td>
<td>ATL</td>
<td>DWF</td>
<td>11:00</td>
<td>16</td>
<td>60</td>
</tr>
<tr>
<td>15</td>
<td>DL</td>
<td>1941</td>
<td>ATL</td>
<td>DWF</td>
<td>12:25</td>
<td>28</td>
<td>68</td>
</tr>
<tr>
<td>16</td>
<td>NW</td>
<td>1193</td>
<td>DTW</td>
<td>DWF</td>
<td>9:31</td>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td>17</td>
<td>NW</td>
<td>407</td>
<td>DTW</td>
<td>DWF</td>
<td>12:40</td>
<td>22</td>
<td>58</td>
</tr>
<tr>
<td>18</td>
<td>NW</td>
<td>1148</td>
<td>MEM</td>
<td>DWF</td>
<td>15:20</td>
<td>33</td>
<td>69</td>
</tr>
<tr>
<td>19</td>
<td>NW</td>
<td>1125</td>
<td>MSP</td>
<td>DWF</td>
<td>10:15</td>
<td>7</td>
<td>81</td>
</tr>
<tr>
<td>20</td>
<td>NW</td>
<td>1167</td>
<td>MSP</td>
<td>DWF</td>
<td>15:45</td>
<td>14</td>
<td>41</td>
</tr>
<tr>
<td>21</td>
<td>US</td>
<td>1772</td>
<td>CLT</td>
<td>DWF</td>
<td>9:25</td>
<td>13</td>
<td>47</td>
</tr>
<tr>
<td>22</td>
<td>US</td>
<td>1101</td>
<td>CLT</td>
<td>DWF</td>
<td>11:24</td>
<td>29</td>
<td>33</td>
</tr>
</tbody>
</table>

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

![Flight path distribution (a) and improvement of dynamic policy over static policy (b).](image)

Fig. 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. Fig. 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 min. 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 min. is only 13% for dynamic policy with $m = 5$.

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.

There are multiple extensions possible of this study. First extension is the consideration of diminishing cargo capacity on flights as we approach departure time, hence, forcing the freight forwarder to make advanced flight commitments on certain segments. Second 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 avenue of research is to study the effect of process efficiency (e.g. loading and unloading times) and carrier flexibility (e.g. late cut-off times) on the dynamic routing performance.

Acknowledgments

We are grateful to the two anonymous reviewers and the editor whose valuable suggestions and comments improved the technical content and presentation clarity of this manuscript. This work was supported by funding grant DTRT06-G-0039 from the US Department of Transportation through the University Transportation Center at University of Toledo (UT-UTC).

References

Hall, R.W., 1986. The fastest path through a network with random time-dependent travel times. Transportation Science 20 (3), 182–188.