Schedule Planning and Endogeneity of Travelers' Decisions in Congested Large-Scale Transportation Networks

A Thesis
Submitted to the Faculty
in partial fulfillment of the requirements for the
degree of

Doctor of Philosophy

by

Keji Wei

Thayer School of Engineering
Dartmouth College
Hanover, New Hampshire

Feb 2020

Examining Committee:

Chairman _____
Vikrant Vaze

Member _____
Alexandre Jacquillat

Member _____
Robert A. Shumsky

Member _____
Amro M. Farid

_____
F. Jon Kull, Ph.D.
Dean of Guarini School of Graduate and Advanced Studies

ABSTRACT

The core task in passenger transportation systems planning can be described as one of matching the network-wide demand with adequate capacity by deploying resources over the corresponding parts of the network. A significant challenge inherent in this task is the uncertainty in demand and supply in transportation systems. Demand uncertainty manifests itself in the form of the passenger choice behavior. Supply uncertainty presents itself as delays and disruptions in resource availability. This thesis tackles a set of three related research questions within passenger transportation systems while explicitly incorporating the inherent uncertainty in both demand and supply during the planning phases of air and urban transportation systems.

Flight delay propagation results in enormous additional operating costs for the airlines, passengers and the aviation system as a whole. The first part of this thesis is focused on proposing, optimizing and validating a methodological framework for estimating the extent of crew-propagated delays and disruptions. We identify the factors that influence the extent of crew-propagated delays, and incorporate them into a robust crew scheduling model. We then develop a fast heuristic approach for solving the inverse of this robust crew scheduling problem to generate crew schedules that are similar to real-world crew scheduling samples. Along with various other findings, our results show that airlines avoid up to 80% of crew-related delays through advanced planning methods.

The second part of this thesis introduces an original integrated optimization approach to comprehensive flight timetabling and fleet assignment under endogenous passenger choice. The resulting optimization model is formulated as a large-scale mixed-integer linear program. We propose an original multi-phase solution approach, which effectively combines several heuristics, to optimize the network-wide timetable of a major airline within a realistic computational budget. Using case study data from Alaska Airlines, computational results suggest that the combination of this model formulation and solution approach can result in significant profit improvements, as compared to the most advanced existing ap-

proaches to flight timetabling. Additional computational experiments based on several extensions also demonstrate the benefits of this modeling and computational framework to support various types of strategic airline decision-making in the context of frequency planning, revenue management, and post-merger integration.

With the soaring popularity of ride-hailing, the interdependence between transit ridership, ride-hailing ridership and urban congestion motivates the following question: can public transit and ride-hailing co-exist and thrive in a way that enhances the urban transportation ecosystem as a whole? To answer this question, in the third part of this thesis, we develop a mathematical and computational framework that optimizes transit schedules, while explicitly accounting for their impacts on road congestion and passengers' mode choice between transit and ride-hailing. The problem is formulated as a mixed-integer nonlinear program, and solved using a bi-level decomposition algorithm. Based on computational case study experiments in New York City, our optimized transit schedules consistently lead to 0.4-3% system-wide cost reduction. This amounts to rush hour savings of millions of dollars per day, while simultaneously reducing the costs to passengers and to transportation service providers. These benefits are driven by a better alignment of available transportation options with passengers' preferences—by re-distributing public transit resources where it provides the strongest societal benefits. These results are robust to underlying assumptions about passenger demand, transit level of service, the dynamics of ride-hailing operations, and transit fare structures. Ultimately, by explicitly accounting for ride-hailing competition, passengers' preferences and traffic congestion, transit agencies can develop schedules that lower costs for passengers, operators and the system as a whole—a rare win-win-win outcome.

Acknowledgements

The dissertation is the end of my enjoyable and meaningful PhD journey at Dartmouth College.

First and foremost, I would like to thank my advisor Vikrant Vaze. I still remember my first time giving an academic presentation in public. It was he who helped me revise the slides and gave me feedback on the rehearsal. I still remember my first time writing an academic journal paper. It was he who helped me improve the writing word-by-word. Vikrant has given me the source of support which has gradually brought me to the world of operations research and transportation. Vikrant, your passion for research, your unselfish guidance and your attention to rigor have made a deep impact on my study and life. I am very fortunate to have had you as my mentor who has helped me become a better person.

I would also like to thank my other advisor, Alexandre Jacquillat. I am greatly indebted to his mentoring. His continuous guidance and advice have helped me improve a lot. Alex, I still remember the countless times when Vikrant, you and I have discussed research by phone in the past five years. Those moments are among the most cherished memories in my life.

I would also like to thank my other interdisciplinary doctoral committee members. Robert Shumsky's business perspective has sparked a lot of new insights and inspired me, especially in the research in Section 4 of this dissertation, and Amro Farid has shared with me his valuable advice about how to write a high-quality thesis proposal.

I was very fortunate to have been on this PhD journey alongside Reed Harder. Reed, it was fun going to conferences with you annually and sharing our research experiences with each other many times in the office. Many other friends also contributed a lot in my research journey including, but not limited to, Yadong Li, Navid Rashedi, Chiwei Yan and Tianli Zhou.

I am greatly indebted to Thayer School of Engineering, which has been a home to me all this time. Thanks to Ligin Solamen, Phyo Aung Kyaw, Chen Li, David Baltrusaitis,

Xiaomin Han, Dennis Wu and Yipeng Shi. It has been my privilege watching Superbowls with you guys every year to witness New England Patriots keep on winning and winning. It was enjoyable to have hotpot dinners with you guys in Hanover, a world of ice and snow. It was amazing to share personal experiences with you guys including various topics like race, religion, life choices and career. Ultimately, those amazing times hanging out with you all made my life in Dartmouth colorful.

Finally, I am deeply grateful to my parents, Hengyi Wei and Xiaowen Sun, who have always supported me unconditionally on the other side of the Pacific Ocean. I wouldn't be here today without their constant love and support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Passenger transportation systems are an integral part of our lives and an important pillar of the global economy. The degree of availability, quality, reliability and affordability of passenger transportation shapes our personal, professional and social experiences. A large proportion of intercity passenger travel is related to the daily commute between home and work, which underscores the value of adequate planning of urban transportation systems for short-distance travel. On the other hand, air travel is the dominant mode for long-distance intercity travel especially in the United States, which highlights the need for efficient air transportation systems planning. This thesis is focused on three related challenges associated with air and urban public transportation systems planning.

Within air transportation, the airline schedule planning process primarily determines the performance of the air transportation system. This process consists of a variety of decision-making steps including schedule design, fleet assignment, aircraft routing and crew scheduling, which are typically carried out sequentially. Chapters 2 and 3 of this thesis are focused on the airline planning process, with particular emphasis on schedule design (also known as *timetabling*), fleet assignment and crew scheduling steps.

Efficiently operated urban public transit is the key to a well-performing urban transportation system. Public transit planning is typically performed in a sequence of steps

including network design, line planning, and timetabling [77]. More recently, the ride-hailing services are also playing an increasingly important role in urban transportation systems. Chapter 4 of this thesis is focused on urban public transit planning with a particular emphasis on timetabling, while explicitly accounting for the interactions with ride-hailing services and traffic congestion.

In this chapter, we first discuss the uncertainty in scheduled transportation systems from both demand and supply sides. Then, in Section 1.3, we present the main contributions of thesis. Finally, Section 1.4 outlines the structure of the rest of the thesis.

1.1 Demand Uncertainty: Passenger Choice Behavior

Demand uncertainty manifests itself in the form of the passengers' demand patterns and choice behavior. The number of passengers interested in traveling and the fraction of them interested in choosing a particular travel option are both dependent on a variety of attributes of the available travel alternatives as well as on each passenger's own personal characteristics. However, the relationship between these attributes of the available travel alternatives and the passengers' choice decisions is highly complex, nonlinear and fraught with uncertainty. Psychologists and behavioral economists have striven to understand and quantify these relationships for several decades, with considerable success. Yet, examples of successful explicit incorporation of these quantitative relationships into transportation systems optimization models are rare and computationally intractable. This is a challenge that we attempt to overcome in this thesis.

For example, an airline's profitability is heavily driven by its ability to attract passengers under competition from other airlines. Many attributes of an airline's schedule offerings, like flight departure times, flight arrival times, assigned aircraft types, number of connections in the offered passenger itineraries, and ticket prices, significantly affect the passenger demand for the airline. As a result, airlines have a strong motivation to explicitly

incorporate these passenger demand dynamics within their timetabling and fleet assignment approaches. Another example, from an urban transportation standpoint, is that of a public transit system trying to redesign its schedule in the face of competition from ride-hailing services. In the recent years, ride-hailing services have attracted a lot of passengers away from public transit leading to a reduction in public transit demand. Compared to traditional public transit services, modern app-based ride-hailing operators (such as Uber) offer several distinct advantages such as convenient door-to-door service, cashless transactions, competitive pricing and access to a broad pool of drivers. Hence, there is a need for public transit systems to incorporate these new and evolving passenger demand dynamics into their planning and scheduling practices to reinvent and sustain these systems of vital societal importance. We attempt to do just that in this thesis.

1.2 Supply Uncertainty: Flight Disruptions and Urban Traffic Congestion

Supply uncertainty presents itself as delays and disruptions to resource availability. Growing congestion and delays throughout the passenger transportation systems not only imposes large costs on the transportation service operators, transportation infrastructure managers, and passengers, but also are a serious threat to our overall quality of life. Transportation delays, disruptions and congestion are estimated to cost the world economy hundreds billions of dollars annually.

In the aviation domain, flight delays and disruptions result in annual economic losses worth tens of billions of dollars. These additional costs include direct costs to airlines and passengers, as well as indirect costs like lost demand and reduced labor productivity. The US Federal Aviation Administration (FAA) estimated the annual costs of flight delays in 2017 to be $\$26.6$ billion [122]. These delays and disruptions are caused by a variety of factors including airport congestion, mechanical problems, baggage delays, delays due to

late arriving aircraft, delays due to late arriving crew, security delays, etc. The Bureau of Transportation Statistics categorizes delays into various categories related to these delay causes [30]. However, these public datasets lack a dedicated category for delays due to the propagation of upstream crew delays and disruptions. There exists a considerable body of literature and rich public datasets on the travel and delay patterns of aircraft and passengers. The equivalent travel and delay patterns for airline crew, however, are lacking. There is a lack of publicly available information about not only the crew-propagated delays and disruptions (CPDD), but also the crew itineraries themselves. The estimates of crew itineraries or even a validated methodology to come up with such estimates is not available. We address these gaps in this thesis.

In the ground transportation domain, a report from the Center for Economics and Business Research (CEBR) and INRIX (a leading international provider of real-time traffic information) estimated the road traffic congestion cost in US to be $124 billion for year 2013 and that number would be up to $185 billion in 2030 [71]. Recently, many reports have claimed that the ride-hailing services are not just accelerating the decline of public transit in the US, but also contributing significantly to the worsening urban road congestion. This thesis focuses on developing models and methods to alleviate a part of these concerns by better scheduling of transit systems.

1.3 Thesis Contributions

This research makes the following three sets of contributions.

1.3.1 Chapter 2: Modeling Crew Itineraries and Delays in the National Air Transportation System

This chapter makes the following four main contributions. First, we propose a comprehensive airline crew pairing optimization model formulation that minimizes the combination

of planned crew costs and the various features that make the crew schedules vulnerable to propagation of delays and disruptions. Second, we solve this model to near optimality by combining known ideas such as branch-and-bound and delayed column generation as well as a sequence of new heuristic ideas developed by us. The sizes of the networks in the problem instances solved by us far exceed those solved in past studies on robust or recoverable crew pairing optimization. Third, we embed this crew pairing generation problem within an upper-level calibration framework wherein a parameterized crew pairing optimization problem is solved repeatedly by varying the parameters until the resulting crew pairings are similar to those used by the airlines. This upper-level calibration problem represents the inverse of the crew pairing generation problem. We employ a local-search heuristic for solving the upper-level calibration problem. Our algorithm is motivated by that of Schaefer et al. [111] and borrows some features from theirs. Ours, however, is the first study to formulate and solve this inverse crew pairing generation problem. Finally, we generate and validate crew pairing solutions that are similar to those used by the airlines in the real world in terms of their potential for the crew-propagated delays and disruptions (CPDD). The out-of-sample testing results demonstrate the accuracy and stability of our modeling framework and algorithms. An important finding is that the ratio between the planned crew cost and approximate delay costs is found to be stable across airlines and aircraft types.

1.3.2 Chapter 3: Airline Timetable Development and Fleet Assignment Incorporating Passenger Choice

This chapter develops and applies an original modeling and algorithmic approach that integrates airlines' comprehensive timetable development and fleet assignment decisions under endogenous passenger choice. Our model formulation leverages the concept of a *sales-based linear programming model* from the airline revenue management literature [59]. Our approach contrasts with the existing literature in three major ways. First, our timetabling model does not involve only incremental timetabling decisions using an existing timetable

as a starting point. Instead, it addresses the comprehensive timetabling problem starting from a clean slate. Second, it captures the endogeneity of passengers' booking decisions through a discrete-choice model, unlike most existing incremental approaches that assume the passenger demand to be invariant with marginal timetabling adjustments. Third, it develops a suite of computational algorithms that enable solving the integrated model for problem instances of realistic sizes and deriving practical insights from computational experiments. Specifically, this chapter makes the following four contributions.

1. It incorporates a sales-based linear programming framework to accurately capture the itinerary-level demand substitution effects into a comprehensive timetable development and fleet assignment optimization model. The resulting model is formulated as a mixed-integer linear program (MILP). To our knowledge, ours is the first research study to capture an airline's comprehensive timetable development problem under passenger choice.

2. We design an effective multi-phase solution approach to solve this model for large-scale problem instances. We demonstrate that, by narrowing down the flight's departure time range step-by-step, a high quality solution could be obtained within reasonable computational run-times—even though the model is intractable with commercial solvers. Additionally, we develop several variable-fixing and symmetry-inducing accelerated heuristics that are shown to obtain an even better solution. We embed these approaches within our multi-phase MILP framework.

3. We present a detailed comparison of our integrated comprehensive approach with the various incremental timetable development approaches found in the literature and in practice, using a major U.S. legacy airline carrier's network. We validate that our modeling and algorithmic framework can yield significant profit improvements for major airlines.

4. We perform additional case studies under modeling, computational and practical ex-

tensions to provide several practical insights and demonstrate the benefits of this modeling and computational framework. These extensions provide computational tools for strategic decision-making in the contexts of frequency planning, revenue management, and post-merger integration.

1.3.3 Chapter 4: Transit Planning Optimization under Ride-hailing Competition and Traffic Congestion

This chapter makes the following three main contributions. First, it develops a mathematical model that optimizes public transit schedules (including frequency and timetabling on each line) while capturing the alternatives provided by ride-hailing, passengers' travel mode choices, vehicle flows in road networks, and traffic congestion. On the demand side, we consider a discrete choice model where travelers choose a transportation option based on prices, travel times and other attributes. We leverage the Generalized Attraction Model (GAM) from Gallego et al. [59] to integrate this discrete choice model into our optimization model. On the supply side, we leverage a traffic equilibrium model that estimates vehicle flows and resulting congestion on road networks, as a function of public transit schedules and travelers' mode choices. The problem is formulated as a mixed-integer nonlinear program (MINLP). To our knowledge, this is the first study that optimizes public transit schedules while capturing interdependencies with ride-hailing and traffic congestion. Second, it designs an original bi-level algorithm that yields high-quality solutions to this MINLP in reasonable computational times. It includes two interconnected procedures. An outer loop optimizes public transit schedules, given travelers' mode choices and estimates of traffic congestion on road networks. Given a transit schedule, an inner loop leverages a user-equilibrium traffic flow model to replicate the paths of ride-hailing vehicles and update traffic congestion levels—by fitting a high-resolution polynomial curve to data-driven estimates of travel times. Through these two procedures, we iteratively update, until convergence, public transit schedules, travelers' mode choices, vehicle flows on road

networks, and traffic congestion. Third, it performs case study experiments in New York City to evaluate the tractability and performance of our modeling and solution framework. Computationally, our bi-level algorithm converges in reasonable runtimes, spanning 8–15 hours. Practically, our optimized transit schedules lead to $0.4-3\%$ system-wide cost reduction, as compared to the actual schedule. This amounts to rush hour savings of millions of dollars per day, while simultaneously reducing costs to passengers and transportation service providers. These benefits are driven by a better alignment of available transportation options with passengers' preferences and ride-hailing services. These results are robust to underlying assumptions about passenger demand, transit level-of-service requirements, the dynamics of ride-hailing operations, and transit fare structures. Ultimately, by explicitly accounting for ride-hailing competition, passengers' preferences and traffic congestion, transit agencies can develop schedules that lower costs for passengers, operators and the system as a whole—a rare win-win-win outcome.

1.4 Organization of the Thesis

The remainder of this thesis consists of four more chapters. Chapter 2 focuses on identifying the nature, extent and causes of propagation of delays and disruptions through crew connections. It particularly tackles the supply-side uncertainty within the context of air transportation systems planning. Chapter 3 focuses on generating optimal airline timetables and fleet assignments by incorporating passenger choice to improve airline's profitability. It particularly tackles the demand-side uncertainty within the context of air transportation systems planning. Chapter 4 focuses on optimizing transit timetables by accounting for passenger choice, ride-hailing competition, and traffic congestion. It tackles the demand- as well as supply-side uncertainty within the context of urban transportation systems planning. Finally, Chapter 5 concludes the thesis and provides directions for future research.

Chapter 2

Modeling Crew Itineraries and Delays in the National Air Transportation System

2.1 Motivation

Flight delays and disruptions cost tens of billions of dollars annually to the world economy. The total cost of flight delays in the U.S. in 2007 was especially large, estimated to be approximately 31.2 billion [13]. Even though the data from the Bureau of Transportation Statistics (BTS) shows that the level of flight delays has somewhat reduced since its 2007 peak, delay costs still represent considerable amount of resource wastage for the U.S. National Air Transportation System, resulting in profit reductions or losses to the airlines and additional inconvenience to the passengers. Over the first half of this decade, that is, from January 1st 2011 to December 31st 2015, only 79.21% of the domestic flights in the U.S. had a delay of 15 minutes or less [28]. During the same period, 1.68% of the U.S. domestic flights were canceled.

Determining the causes of flight delays and disruptions has been a topic of considerable interest among researchers and policymakers: see Allan et al. [7] for an early example. The U.S. Department of Transportation (DOT) classifies flight delays into five main categories.

They are Air Carrier Delays, Late Arriving Aircraft Delays, National Aviation System Delays, Extreme Weather Delays, and Security Delays. Between Jan 1st 2011 and Dec 31st 2015, Air Carrier Delays (defined as those within the control of the airline, such as those due to maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.) accounted for nearly 32% of all flight delays, while another 34% were attributed to Late Arriving Aircraft Delays (those due to late arrival of the previous flights using the same aircraft) and about 31% were classified as the National Aviation System Delays (defined as those due to non-extreme weather conditions, airport operations, heavy traffic volume, air traffic control, etc). However, this public data lacks a dedicated category for flight delays due to the propagation of upstream crew delays and disruptions. Delayed or disrupted flights may generate delays and disruptions to subsequent flights because the crew for those flights is delayed, out of position, or unable to operate the scheduled flights without violating government regulations or collective bargaining agreements (CBAs). Presently, these delays (henceforth called as the Crew-Propagated Delays and Disruptions in this chapter or CPDD for short) are considered to be a subset of the rather broad category of Air Carrier Delays. Accurate estimation of the CPDD is critical not only as a step toward fully understanding the aviation system performance, but also for informing government policy and air carrier decisions related to airline crew scheduling. In this chapter, we develop an approach to estimate the crew-propagated delays and disruptions, which are similar in concept to the aircraft-related delays and disruptions reported by the DOT's Late Arriving Aircraft Delays category.

There is yet another motivation for conducting this research. Public data sources lack information about not only the CPDD, but also the crew itineraries themselves. Many of the past studies related to flight delay propagation assume knowledge of aircraft connections, crew connections, and passenger connections. There is a large amount of literature focusing on airline recovery optimization (see Petersen et al. [106], for a recent example, and Barnhart et al. [20] for a detailed review) which uses aircraft, crew and passenger

schedules as inputs to their computational case studies. There is also a growing body of literature on various strategies for airport congestion mitigation (e.g., [75, 141, 132, 154]), which assumes the knowledge of aircraft, crew and passenger connections. A report by the Office of Inspector General (2010) acknowledges that despite the significance of delay propagation effects, their extent and nature is not well quantified or understood, partly due to the limitations of the existing data. Several past studies have focused on analyzing the extent of flight delay propagation using a variety of approaches such as analytical queuing models [108], delay propagation trees [4], and statistical models. These studies have also highlighted the importance of the availability of aircraft, crew and passenger connections data.

However, only the aircraft connections information is available publicly through the BTS (Bureau of Transportation Statistics) website. A prior study by Barnhart et al. [20] used a statistical approach to estimate passenger itineraries, passenger delays and disruptions. But similar estimates of crew itineraries or even a validated methodology to come up with such estimates is not available. Apart from aiding in future research studies such as those mentioned above, such estimates of crew itineraries would also be beneficial for assessing the full impact of any delay mitigation strategy being considered by airlines and/or government. In this chapter, similar to the work on passenger itinerary and delays estimation by Barnhart et al. [20], we develop a crew itinerary estimation methodology to generate a database of estimated crew itineraries that will enable accurate estimation of the crew-propagated delays and disruptions (CPDD) consistent with their real-world values.

Note that we *do not* attempt to develop an approach to generate crew itineraries that are identical to the real-world airline crew itineraries in *every possible aspect.* Such objective would not only be extremely difficult to attain, but also likely cause overfitting issues based on the limited confidential data samples which we use for the estimation purposes. Instead, we develop a robust process to generate crew itineraries that are similar to the real-world airline crew itineraries in their potential for causing the crew-propagated delays and dis-

ruptions. There are possibly other, non-delay related, aspects of crew itineraries that could be of relevance for other purposes. However, the focus of this chapter is to ensure that our process is accurate and stable in terms of the CPDD estimation.

The CPDD for an airline during a particular time period are strongly dependent on the chosen set of crew pairings. Crew pairing is defined as a sequence of flights covered by a crew member that follows a number of regulations, and can be considered to be the smallest self-contained unit comprising crew itineraries. Thus, estimating crew pairings is at the core of the challenge of estimating the CPDD. A crew pairing problem is the one of generating a set of crew pairings that covers all available flights legs. Crew pairing decisions have a significant effect on airlines' planned and operational costs. Operational costs include delay and disruption costs due to irregular operations, and are amplified by their propagation through crew connections. Thus, understanding the extent, causes and impacts of propagation of delays and disruptions is essential for developing methods to reduce them and to improve the overall aviation system performance. Previous research studies on crew pairing generation have focused on minimizing the planned crew costs, and sometimes a subset of the various components of operational costs, but none of them have focused on quantifying and understanding the CPDD in the real-world airline networks. Therefore, there is a need to develop approaches and models that produce crew schedules that are similar to the real-world crew schedules in terms of the CPDD.

Finally, we note that the present research project was originally motivated and funded by an aircraft delay modeling limitation faced by the U.S. Federal Aviation Administration's (FAA) Office of Performance Analysis. The FAA analyzes and forecasts, on a monthly basis, aircraft delays at the nation's major airports, with the objective of identifying airports with significant potential for delays months in advance, so that appropriate actions may be taken to prevent or mitigate such delays. A discrete events simulation platform developed by the FAA for this purpose models aircraft-based delay propagation using public data on aircraft itineraries [28], but does not account for crew-based propagation

effects due to the lack of crew itinerary data. This is believed to contribute to an underestimation of the propagated delays, and served as the motivation for conducting the research work presented in this chapter.

2.1.1 Crew-Pairing Optimization Problem

A crew pairing consists of a sequence of duties, where a duty is defined as the set of tasks to be performed by a crew member during a given day. Duties are connected by rest periods. Each duty is made up of a set of consecutive flights with some gaps between them. These gaps are called sit times. A pairing should begin and end at a crew base which is usually the domicile of a crew member. Both pairings and duties are subject to various regulations and contractual restrictions. Typically, these include the following.

- The total flying time within a duty cannot exceed an upper bound. There is also an upper bound on the total elapsed time within a duty.
- There is a lower bound on the sit time which guarantees that the crew has enough time to connect between two consecutive flights within a duty.
- The rest time between duties should be greater than or equal to a minimum rest time which ensures that the crew is sufficiently rested between duties.
- There is typically an upper limit on the number of duties within a pairing.

In addition to these rules, even when ignoring the operational cost considerations, crew pairings also have a highly non-linear pay structure. Note that the crew pay is commonly expressed in the units of hours of crew flying, which we will use throughout this chapter. For a typical North American airline, the planned cost of a pairing p is the maximum of two terms: sum of the costs C_d of all its duties and a fixed fraction (ζ) of the total time away from base ($TAFB_p$). Thus the planned cost of a pairing p (measured in the units of

hours of crew flying) is given by

$$c_p = \max(\sum_d C_d, \zeta * TAFB_p) \tag{2.1}$$

For each duty d, the planned cost (C_d) is the maximum of three terms, a minimum guaranteed pay (δ) per duty, flying hour (fly_d) of the duty, and a fixed fraction (ϵ) of the duty elapsed time ($elpased_d$). Parameters δ, ϵ, ζ may vary across different carriers. Thus the planned cost of duty d (measured in the units of hours of crew flying) can be written as

$$c_d = \max(\delta, fly_d, \epsilon * elapsed_d) \tag{2.2}$$

The objective of the deterministic crew pairing problem is to minimize the planned crew cost and is usually modeled as a set partitioning problem [15]. We denote the set of flights by F and the set of pairings by P. a_{lp} is 1 if pairing p contains flight i and 0 otherwise. x_p is a binary decision variable which equals 1 if pairing p is chosen in the crew pairing solution, and 0 otherwise. Then, the crew pairing problem can be formulated (ignoring crew deadheads) as

$$\begin{aligned} \text{Min} \quad & \sum_{p \in P} (c_p * x_p) \\ \text{subject to:} \quad & \sum_{p \in P} (a_{ip} x_p) = 1, \forall i \in F, & (2.3) \\ & x_p \in (0,1), \forall p \in P, & (2.4) \end{aligned}$$

2.1.2 Literature Review

As mentioned earlier, from an application perspective, our work is motivated by the work of Barnhart et al. [20]. Using one year of flight delay data and a one-quarter sample of confidential passenger booking data from an airline, they estimated passenger itinerary flows and developed insights into the factors that affect the performance of the U.S. national

air transportation system from a passenger perspective. They developed a methodology to model historical travel and delays for passengers. From a methodological perspective however, our work is fundamentally different from theirs. While they used a statistical approach to estimate passenger itineraries, estimation of crew itineraries is considerably more complicated because of the complex rules governing what constitutes a legal crew itinerary. Thus, a statistical estimation approach is unsuitable for our task. Also, while the number of possible passenger itineraries per day can be in the thousands for a large airline, the number of legal crew itineraries is usually larger by several orders of magnitude, often making it very difficult or impossible to even enumerate all of them exhaustively.

Estimating delay propagation through crew connections is also much more complex than estimating the same through aircraft connections due to the more complex nature of crew work regulations than aircraft maintenance regulations. Several past studies on crew pairing optimization have tried to identify and capture one or more dimensions of a crew schedule that affect the extent of propagation. Broadly, these past studies can be divided into three main categories. First category of studies aims to incorporate one or more features that affect the ease of recovering the crew schedules after a disruption. Here, crew schedule recovery refers to the set of reactive measures available to an airline to bring its crew schedule back on track after a disruption and it typically includes alternatives such as delayed flight departures, crew swaps, reserve crews, flight cancellations, etc. Second category of studies aims to generate crew pairings that are difficult to get disrupted and/or have a low disruption cost. Studies in the first and second categories deal exclusively with crew schedules without capturing the relationship between crew-based and aircraft-based propagation. The third category attempts to capture this interdependence.

Studies in the first category usually focus only on one or two specific factors that improve recovery potential. Shebalov and Klabjan [116] maximize the number of move-up crews, wherein a move-up crew for a flight is a crew that is not actually assigned to that flight but can be feasibly and legally assigned to it. On the other hand, Gao et al. [60]

extend the fleet purity idea proposed by Smith and Johnson [125] to crew base purity. The crew base purity idea restricts the number of crew bases allowed to serve each airport in order to increase the opportunities to find a move-up crew in crew recovery. Shebalov and Klabjan [116] as well as Gao et al. [60] capture the potential for crew swaps, which is an important dimension of crew recovery process, but do not explicitly capture the extent of delay propagation through crew connections.

Studies in the second category apply a variety of robust planning approaches to airline crew scheduling. Lu and Gzara [90] developed a non-linear robust optimization model for balancing planned and operational crew costs for a simplified crew pay structure focused solely on the Time Away From Base (TAFB). Yen and Birge [152] develop a two-stage stochastic programming model that implements a simplified recovery model for the second stage. Schaefer et al. [111] adjust the cost of each crew pairing to include a combination of planned costs and a linear approximation of delay costs. The delay cost approximation function is fine-tuned based on a discrete events simulation software named SimAir [110]. The delay cost is assumed to be a function of four attributes, namely, 1) the sit time between consecutive flights within a duty, 2) the rest time between consecutive duties, 3) the total flying time in a duty, and 4) the total elapsed time in a duty. The rationale behind these choices is that the potential for the propagation of delays and disruptions is greater when the crew's sit and rest times are too short and when the per-duty flying and elapsed times are too long. Yen and Birge [152] as well as Schaefer et al. [111] account for the differences in the delay propagation potential of different crew pairings, but neither captures recovery actions such as crew swaps.

Because delays can propagate due to both late arriving/unavailable aircraft and late arriving/unavailable crew, there are many interdependencies between the effects of aircraft schedules and crew schedules on the propagation of delays and disruptions through the overall flight network. The aircraft scheduling and crew scheduling stages of the airline schedule planning process are conventionally solved in a sequential manner. However,

recognizing the interdependencies between the two stages, in terms of the planned costs and delay propagation potential, some recent studies (such as Dunbar et al. [51] and Weide et al. [147]) have developed integrated robust optimization models. Weide et al. [147] attempt to increase the buffer in crew connection times when the crew changes aircraft. Dunbar et al. [51] solve an integrated fleet assignment, aircraft routing and crew pairing problem with a weighted average objective function that incorporates robustness solely as measured by the number of aircraft changes between successive flights in a crew itinerary. Mercier et al. [99] also incorporate aircraft changes by the crew as a measure of robustness in their integrated aircraft routing and crew scheduling model. Tam et al. [134] describe and evaluate a bi-criteria optimization approach to balance the planned crew costs and a single robustness measure which penalizes crew connections with aircraft changes and small crew sit times. Studies in this category usually emphasize aircraft changes and crew sit times but do not focus much on crew recovery potential, crew rest times, duty flying times, or duty elapsed times.

In summary, past research studies in airline crew scheduling have identified the various features of airline crew schedules that affect the CPDD. However, no prior study has combined these different features into a single optimization model. Additionally, while some past studies, such as Yen and Birge [152], have attempted to incorporate the actual delay costs into the crew pairing optimization models, these models have been highly simplified due to computational tractability issues. Finally, and most importantly, all aforementioned studies have focused, implicitly or explicitly, on finding an "optimal" crew schedule with respect to a known optimization formulation. The problem that we solve in this chapter can be thought of as the inverse of this problem. Given an actual crew pairing sample, our goal is to reverse engineer the process used, and the problem solved, by the airlines to generate the crew pairings that the airlines actually used. This will enable us to generate similar crew pairings for other airlines, and/or other aircraft families, and/or other time periods than the ones for which the crew pairing data sample is available. It is common knowledge that the

major airlines typically use advanced optimization solvers to generate crew pairings. Furthermore, in addition to minimizing planned costs, most airlines are known to directly or indirectly attempt to reduce delay and disruption costs as well. However, the exact models and algorithms used by a particular airline for crew pairing optimization are proprietary. Therefore, in this chapter we reverse engineer airlines' crew pairing generation process with the objective of generating pairings that are similar to the actual airline-generated crew pairings in terms of their CPDD potential.

The rest of this chapter is structured as follows. Section 2.2 describes our overall modeling approach and problem formulation. Section 2.3 describes the solution approach, including the exact algorithms and heuristic ideas developed by us to solve this challenging problem. Section 2.4 describes our computational case studies in terms of data and preprocessing, and presents evidence of the computational tractability of our approach. Section 2.5 describes the calibration and validation results obtained from our series of computational experiments. Section 2.6 further validates our results in terms of the CPDDs and also describes how to use our results for estimating the CPDDs for any given network. Finally, Section 2.7 discusses the main conclusions and the directions for future research.

2.2 Modeling Framework

Our objective is to generate crew itineraries that are similar to the real-world crew itineraries as measured by the extent of the CPDD. Therefore, we first need to develop an appropriate similarity metric for comparing two crew pairing solutions with each other, for any given flight network. Defining similarity directly based on the actual costs of propagated delays and disruptions is problematic for multiple reasons. Propagated delays and disruptions depend on not only the crew schedules but also the underlying root (i.e., non-propagated) delays and disruptions, as well as the operational recovery actions used by the airline. Hence these are difficult to model accurately. Moreover, while planning the crew sched-

ule, the airline itself is unaware of the exact set of root delays and disruptions that it will face on a given day of operations. For these reasons, accurate calculation of the CPDD costs is impossible. Instead, we measure the similarity between crew pairing solutions in terms of their *CPDD potential.* As explained in Section 2.1.2, the CPDD potential is a function of various features of a crew schedule. In Section 2.2.1, we classify these features into four categories and select six representative features for inclusion in our model. Then in Section 2.2.2, we provide the mathematical formulation for the crew pairing optimization problem used as the basis of our calibration framework. Finally, in Section 2.2.3, we give a mathematical formulation for our calibration problem of minimizing the distance (i.e., maximizing the similarity) between the estimated and actual crew pairing solutions in terms of their CPDD potential as quantified by these six features.

2.2.1 Representative Features

In the absence of sufficient schedule buffers and recovery opportunities, delays and disruptions propagate to downstream flights, leading to additional operating costs. Therefore, besides the planned crew costs, airlines often consider some of these buffers and/or recovery opportunities during crew scheduling to reduce these extra operational costs. There are a variety of mechanisms through which delays and disruptions affect downstream flights. When the sit time buffer (defined as the scheduled sit time minus the minimum required sit time) or the rest time buffer (defined as the scheduled rest time minus the minimum required rest time) between two consecutive flights is less than the arrival delay of the first flight, delay propagates to the second flight unless some recovery action, such as a crew swap, is able to prevent it. Thus, the sit time buffers, the rest time buffers, and the crew recovery potential affect the crew-propagated delays and disruptions. However, if these two flights are scheduled to be operated by the same aircraft, then this delay to the second flight will be unavoidable because of aircraft based propagation, irrespective of whether the crew is on time. Note that, as per the DOT classification, delay propagation in such

situations is classified as aircraft based propagation, causing the CPDDs to be counted as zero to avoid double counting. Thus, whether or not the crew travels with the same aircraft affects the CPDDs. Finally, if flight delays result in violation of any of the crew duty regulations and/or CBA rules, such as the total flying time in a duty or the total elapsed time in a duty, then the later flight becomes inoperable by its scheduled crew, resulting in either a flight cancellation or some crew recovery action. Thus, the available buffers (defined as the maximum allowable value minus the scheduled value) in the total flying time or the total elapsed time in a duty also affect the CPDDs. This discussion motivates our classification of features affecting the CPDDs as well as our choice of the representative features.

We divide the features affecting the CPDD into four categories: Aircraft Change, Push-Back, Crew Legality, and Crew Swaps. This categorization highlights the variety of ways in which delays and disruptions can propagate through crew connections, and it facilitates any future revisions or extensions of the feature-set based on the methodologies we have developed.

- **Aircraft Change**

  We first motivate this category with an example. Consider two flights, Flight 1 and Flight 2, scheduled to be operated consecutively by the same crew within the same duty. If Flight 1 and Flight 2 are scheduled to be operated by the same aircraft, then irrespective of whether Flight 1's is delayed or not, by the time the aircraft is ready to operate Flight 2, the crew will typically be ready as well. Thus, there will be either no delay propagation or there will be some delay propagation attributed to the late arriving aircraft. However, no crew-propagated delay or disruption will occur. On the other hand, if Flight 1 and Flight 2 are scheduled to be operated by different aircraft, then to avoid delay propagation from Flight 1 to Flight 2, the crew on Flight 1 will need to exit that aircraft, reach the aircraft scheduled to operate Flight 2 and get ready to start operating it before the scheduled departure time of Flight 2. In this scenario, delay might propagate through the crew connection. Therefore, if the crew

needs to change aircraft between consecutive flights within a duty, there is a potential for the CPDD to occur. Hence whether or not the crew stays with the aircraft is an important factor affecting the CPDD. Therefore the number of times a crew switches aircraft within a duty is included as one of the representative features in our model.

- **Push-Back**

  When a flight's arrival is delayed, and the same crew within the same duty is scheduled to operate a subsequent flight, which is not scheduled to be operated by the same aircraft, a simple policy is to delay the subsequent flight until its scheduled crew is ready to operate it, regardless of how severe the delay is. We call this as the push-back strategy [110]. Similarly, when the arrival of the last flight in a crew duty (which is not the last duty in the crew pairing) is delayed, push-back strategy may be used to delay the departure of the first flight in the crew's next duty regardless of how severe the delay is and irrespective of whether or not the same aircraft is scheduled to operate the two flights. Note that, under the push-back strategy, delay propagates through the crew connection when the buffer in the crew sit time or the crew rest time exceeds the arrival delay of the first flight. Thus, the crew sit time buffer (between flights not scheduled to be operated by the same aircraft) and the crew rest time buffer are important factors affecting the CPDD potential, and hence both are used as representative features in our model.

- **Crew Legality**

  When developing crew schedules, airlines must adhere to FAA crew safety regulations and CBAs regarding the maximum flying time in a duty and the maximum elapsed time in a duty. For example, if FAA regulations limit a pilot to a maximum of 8 hours of flying time during a duty, and if the scheduled flying time is exactly 8 hours or just under 8 hours, then even a small delay to one of the earlier flights could cause the actual flying time to exceed 8 hours. This would disallow the pilot

to operate the last flight in the duty until the completion of a rest period, either leading to a flight schedule disruption such as cancellation or large delay, or triggering a crew recovery action such as a crew swap or the use of reserve crews. Note that, under this scenario, the CPDD occur even when the crew connection time buffer is large and/or the crew is not scheduled to change aircraft between flights. A similar argument holds when the scheduled elapsed time in a duty is equal to or just under the maximum allowable duty elapsed time. Thus, the buffer in the flying time and the elapsed time in a crew duty are important factors affecting the CPDD potential, and hence both of these are used in our model as representative features.

- **Crew Swaps**

  As mentioned in Section 2.1.2, crew schedule recovery actions include alternatives such as delayed flight departures, crew swaps, reserve crews, flight cancellations, etc. While delaying flight departures is the default alternative, under significant disruption events, it can result in very large and expensive delays. A flight cancellation cannot be done in isolation; typically it leads to cancellation of one or more other flights scheduled to be operated by the same aircraft and requires extensive amounts of passenger rebooking. Use of reserve crews is constrained by their availability and is typically an expensive strategy as well. A crew swap involves assigning a late arriving crew to operate a flight with a later departure time than its originally scheduled flight and instead using a different crew to operate the earlier flight. To allow a crew swap, the two swapped pairings must be from the same crew base, must end on the same day, and either crew must be qualified (in terms of equipment, route and airport certifications) to operate the subsequent flights in both pairings [116].

  Compared to other crew recovery actions such as cancellations or reserve crews, swaps are typically less expensive, and therefore airlines find it beneficial to increase the crew-swapping opportunities. Gao et al. [60] introduced the concept of crew base

purity to restrict the number of crew bases serving each airport. They found that improving the crew base purity can significantly increase crew-swapping opportunities and thus reduce the cost of crew recovery. They describe the idea of using an adjacency graph to quantify the extent of crew swapping potential. In an adjacency graph, airports are represented by nodes and the existence of an arc implies that there is at least one flight connecting the two nodes. For a specific airline's network, distance between two airports in an adjacency graph is defined as the minimum number of arcs that need to be traversed to go from one airport to the other. Crews serving airports that are more distant from the crew base, lead to fewer crew swapping opportunities and thus lower recovery potential. In our model, the number of times a crew visits an airport which is at a distance of 2 or more from its base is used as a feature representative of the crew swapping potential and hence representative of the CPDD potential. Other features indicating the crew recovery potential, such as, the number of reserve crew members available at various airports, could also be potentially included as representative features. However, we did not include them because of the lack of data on the availability of reserve crews in our dataset.

2.2.2 Robust Crew Pairing Formulation

The six representative features identified in Section 2.2.1 were integrated into a mathematical model that generates crew pairings that are similar to the real-world airline crew pairings. Our model formulation is motivated by the work of Schaefer et al. [111], which used a penalty method for quantifying and maximizing the robustness of a crew schedule. They optimize the total expected operational cost of a crew pairing solution, which is defined as the sum of the planned cost and a linear function of four attributes of each crew pairing serving as proxy measures of its robustness. They assume that the aircraft are always available and hence no delay propagates through the aircraft connections. Also, the recovery method is assumed to be push-back only. Finally, they assume that the opera-

tional cost of a crew pairing solution is the sum of the operational costs of the individual chosen pairings, and that interaction between crew pairings does not have an effect on the operational costs. We retain this last assumption, but partially relax the first and second assumption as follows. Similar to the four attributes chosen by Schaefer et al. [111], we also include, in our representative features set, the scheduled sit time when crew changes planes, the scheduled rest time between duties, flying time in a duty, and elapsed time in a duty. Additionally, we also include as one of our representative features, the number of times a crew changes aircraft between successive flights within a duty. This provides a partial proxy for the additional delay that may result from the late arriving aircraft. Similarly, we also include the crew base purity, as measured by the number of times the crew arrives at an airport whose distance from the base is 2 or greater in the adjacency graph. We define these as the instances of violation of the crew base purity. Crew base purity provides a proxy for the crew recovery potential through crew swaps, as described in Section 2.2.1. Thus, we used the following six features:

Feature 1: Scheduled sit time when the crew changes aircraft

Feature 2: Scheduled rest time between duties

Feature 3: Flying time in a duty

Feature 4: Elapsed time in a duty

Feature 5: Number of crew base purity violations

Feature 6: Number of aircraft changes by the crew within a duty

Our method of incorporating these features into the crew pairing optimization model is an extension of the penalty method developed by Schaefer et al. [111]. For For any pairing p, let c_p be its planned cost, and f_p be the penalty cost as a function of feature i. Then the

total cost (c_p) of pairing p (measured in the units of hours of crew flying) is defined as

$$\bar{c_p} = c_p + \sum_{i=1}^{6} f_p(i) \tag{2.5}$$

For Features 3 and 4, as their scheduled value approaches the largest acceptable value, the potential for the crew-propagated delays and disruptions (CPDD) increases. For instance, the FAA requires that a crew must rest if it has already flown for 8 hours in a duty. As the scheduled flying time in a duty increases, the chances of this pairing becoming illegal during operation increase because of increased likelihood of violation of this rule. Similarly, for Features 1 and 2, as their scheduled value approaches the smallest acceptable value, the CPDD potential increases. For Feature i, let δ_i denote the relevant bound, that is, lower bound for Features 1 and 2 and upper bound for Features 3 and 4. For example, for Feature 2, δ_2 is the minimum rest time as allowed by the FAA regulations and the CBAs. For δ_2 of 10 hours, rest periods shorter than 10 hours in length are not permitted. Let $Count(i, p)$ be the number of times that Feature i occurs in pairing p, and let $V_{i,p}^{j}$ be the value of the j^{th} occurrence of Feature i in pairing p. For instance, if pairing p has three duties with elapsed times of lengths $10, 12, 5$ hours respectively, then $Count(4, p) = 3, V_{4,P}^{2} = 12$, and $V_{4,P}^{3} = 5$. We use parameters α_i to represent the maximum penalty, and β_i to represent the slope in Feature i's penalty function. So, for the first four features, the function $f_p(i)$ is defined as

$$f_p(i) = \sum_{j=1}^{Count(i,p)} max(\alpha_i - \beta_i |V_{i,p}^{j} - \delta_i|, 0), \forall i \in \{1, 2, 3, 4\}, \tag{2.6}$$

The form of function $f_p(i)$ described by Equation (2.6) is similar to that used by Schaefer et al. [111]. It assumes that $f_p(i)$ is additive across the effects of all occurrences of feature i in pairing p. Also, it assumes that, within a range, the effect of the value of the feature in each occurrence is linear and increases as the value of the feature gets increasingly closer to the relevant bound δ_i. At the bound, the effect has the maximum value α_i,

because this leaves zero buffer in case of any prior delays or disruptions, and hence creates the maximum CPDD potential. Farthest away from the bound (i.e., at a distance of $\frac{\alpha_i}{\beta_i}$), the the effect is zero. This is because large enough buffers almost fully eliminate any CPDD potential.

For defining the penalty function for Feature 5, we observe that if most airports are directly connected to the crew base, the airline has a greater potential for recovery by finding a move-up crew. Also, as for Feature 6, we observe that if most crews stay with the aircraft, then most of the delay propagation would be attributed to late arriving aircraft, rather than being counted as part of the CPDD. So we penalize the number of occurrences of crew changing aircraft and the number of occurrences of crew base purity violations. Let parameters γ_i, $i \in \{5, 6\}$ denote the penalty weights for Features 5 and 6. With $Count(i, p)$ defined the same way as that for Features 1 through 4, the function $f_p(i)$ for Features 5 and 6 is defined as

$$f_p(i) = \gamma_i * Count(i, p), \forall i \in \{5, 6\}, \tag{2.7}$$

Note that this expression is simpler than that for Features 1 through 4 because the number of aircraft changes and the number of crew base purity violations directly have an effect on the CPDD potential, as against the effects of Features 1 through 4 which depend on the difference between the feature value and the relevant bound. This results in a crew pairing optimization model given by

$$\text{Min} \quad \sum_{p \in P} (c_p + \sum_{i=1}^{6} f_p(i)) x_p \tag{2.8}$$

$$\text{subject to:} \quad \sum_{p \in P} (a_{ip} x_p) = 1, \forall i \in F, \tag{2.9}$$

$$x_p \in (0, 1), \forall p \in P, \tag{2.10}$$

2.2.3 Calibration Framework

There are several ways of conceptualizing our calibration problem. Given the optimization model (2.8) to (2.10), we could consider the calibration problem as one of estimating the parameters $\alpha_i, i \in \{1,2,3,4\}, \beta_i, i \in \{1,2,3,4\}$ and $\gamma_i, i \in \{5,6\}$. Thus it is an inverse optimization problem. While an inverse linear optimization problem has been shown to be another linear optimization problem [5], and hence is easy to solve, similar results do not exist for an inverse integer optimization problem (IIOP). Recently, Lamperski and Schaefer [84] developed an approach to formulate the IIOP as an integer optimization problem with exponentially larger size. Others have proposed heuristic approaches for solving variants of the IIOP (see Duan and Wang [50]; and Wang [143]; for recent examples). However, these are computationally intensive and deal with only small-sized problems. A crew pairing optimization problem, on the other hand, typically consists of millions of (or more) variables, and is typically solved using complex, resource-intensive algorithms such as branch-and-price [17]. Therefore, solving an inverse version of such a problem is extremely challenging for realistic problem sizes and no existing study has addressed this challenge successfully.

Alternatively, the calibration problem could also be considered a type of supervised machine learning problem where the goal is to generate crew pairing solutions similar to those in the labeled training data by learning the parameters $\alpha_i, i \in \{1,2,3,4\}, \beta_i, i \in \{1,2,3,4\}$ and $\gamma_i, i \in \{5,6\}$. This labeled training data, represented by a set of crew pairings, is in the form of a set of sequences of flights. This is in a non-standard structure for supervised machine learning, and the mechanism through which the parameters affect the labels is also very complicated. Thus, none of the typical supervised learning approaches, such as support vector machines or neural networks, to name a few, are directly applicable.

This discussion suggests that our calibration problem has several unique attributes, and is computationally much more expensive compared with what existing methods have been shown to solve. Therefore, we propose a new mathematical framework and a solution heuristic for solving this calibration problem. First, in this section we describe the frame-

work and the relevant mathematical notation. Then, in Section 2.3, we describe the solution heuristic.

Let us denote the set of parameters by $PARAMS$. Thus, $PARAMS = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \beta_3, \beta_4, \gamma_5, \gamma_6\}$. Let $\hat{x}(PARAMS)$ be the crew pairing solution generated by solving the optimization model (2.7) to (2.10) for a given set of $PARAMS$ values, and let x be the real-world airline's scheduled crew pairing solution in our sample data. So, for each set of parameters, we have

$$\hat{x}(PARAMS)) \in \underset{X_p \in \{0,1\}, \forall p \in P}{\operatorname{argmin}} \{\sum_{p \in P}(c_p + \sum_{i=1}^{6} f_p(i))x_p : \sum_{p \in P}(a_{ip}x_p) = 1, i \in F\} \quad (2.11)$$

Also, let $F^{\hat{X}}(i) = \sum_{p \in P}(f_p(i)\hat{x}_p)$ and $F^{\hat{X}}(i) = \sum_{p \in P}(f_p(i)x_p)$ be the values of the i=i^{th} components of the penalty functions corresponding to the crew pairing solutions $\hat{x}$ and x respectively. Then the calibration problem is formulated as follows:

$$\text{Minimize} \quad \sum_{i=1}^{6} |F^{\hat{x}}(i) - F^{x}(i)| \quad (2.12)$$

Subject to:

$$f_p(i) = \sum_{j=1}^{Count(i,p)} max(\alpha_i - \beta_i |V_{i,p}^{j} - \delta_i|, 0), \ \forall i \in 1, 2, 3, 4 \quad (2.13)$$

$$f_p(i) = \gamma_i * Count(i, p), \ \forall i \in 5, 6 \quad (2.14)$$

$$\hat{x} \in \underset{x_p \in 0,1 \ \forall p \in P}{\operatorname{argmin}} (\sum_{p \in P}(c_p + \sum_{i=1}^{6} f_p(i))x_p : \sum_{p \in P}(a_{ip}x_p) = 1, i \in F) \quad (2.15)$$

Note that this formulation minimizes the L1 norm of the difference between $F^{\hat{x}}(i)$ and $F^{x}(i)$. Alternatively, we could consider minimizing other norms (such as L2 norm) as well. Our computational experiments with L1 and L2 norms showed that these two alternative formulations did not lead to any significant changes in our results.

2.3 Solution Approach

To generate crew pairings that are similar to those scheduled by the airline, we need to solve the calibration optimization problem represented by (2.12)-(2.15). The similarity or closeness between the two crew pairing solutions provides a measure of the success of the calibration process. However, in order to truly assess the stability of this approach, we need to perform out-of-sample testing. As described in Section 2.4, we use one set of sample data to calibrate the parameters and then use the same calibrated values with another set of sample data (from a different airline, and/or different aircraft family, and/or different time period) to assess the stability of our approach. But before that, we need to develop a heuristic to solve this very difficult problem represented by (2.12)-(2.15). Note that the right hand side of constraint (2.15), in itself, is a very challenging problem for large network sizes. It is a type of robust crew pairing optimization problem, and no prior study in the literature has solved such problems of size as large as those of the networks used in this chapter. Therefore, we develop and implement new heuristic approaches to solve the inverse of this already very difficult problem. In this section, we describe the solution approach. Then, in Section 2.4, we present our computational results.

We begin this section by describing, in Section 2.3.1, our overall heuristic for solving the calibration problem. This involves repeatedly solving instances of the model (2.8) to (2.10). Section 2.3.2 summarizes the solution approach for model (2.8) to (2.10), which itself includes repeatedly solving instances of the LP (linear programming) relaxation of this integer optimization problem. The solution to the LP relaxation of model (2.8) to (2.10) involves repeatedly solving instances of a sub-problem called the pricing problem. The process for solving this pricing problem is described in Section 2.3.3.

2.3.1 Local Search Heuristic for the Calibration Problem

We use a local search method for solving the optimization problem given by (2.12)-(2.15). It starts with all parameters in the set PARAMS initialized to 0. It then varies the parameters corresponding to each feature, one feature at a time, using a simple grid search to identify any opportunities for improving the objective function (2.12). The algorithm terminates when no better solution can be found in an iteration and returns the current best solution. This detailed procedure can be found in the Appendix A.2. Note that these grid-searches require us to examine various combinations of $PARAMS$ values to calculate the $\sum_{i=1}^{6} |F^{\hat{x}}(i) - F^{x}(i)|$ value. Examining each combination of $PARAMS$ values requires solving the robust crew pairing optimization problem given by (2.8) to (2.10). Next, we discuss the process for solving this problem.

2.3.2 Crew-Pairing Solution Approach

Typically, the deterministic crew pairing optimization problem is solved by techniques such as branch-and-price [17], which combine ideas from the branch-and-bound algorithm for solving integer optimization problems with the delayed column generation ideas for solving large-scale linear optimization problems. The reader is referred to Kasirzadeh et al. [78] or a detailed review of the state-of-the-art techniques in this area. Unlike previous studies, our goal is not just to solve the robust crew pairing problem once. Instead, its solution constitutes a sub-problem within our overall calibration optimization process described in Section 2.3.1. The overall calibration algorithm requires solving hundreds of these individual crew pairing optimization problems. Therefore, our computational performance requirements are far more stringent than those of most prior studies in the literature. We cannot afford to wait for several hours to solve the crew pairing optimization problem. Instead of using column generation at each node of the branch-and-bound tree, which is very time consuming, we use a heuristic strategy to solve this problem. As explained in Section 2.4, this strategy helps us in obtaining solutions that are within a small optimality gap. This

strategy can be summarized as follows. It refers to two other algorithms, *Algorithm A and Algorithm B*, which are described in Appendix A.1.

Heuristic Solution Strategy for the Robust Crew-Pairing Optimization Problem

Step 1: Form the Restricted Master Problem (RMP) by including only a small subset of columns and relaxing the integrality constraints.

Step 2: Solve the RMP to find a set of dual variable values.

Step 3: Using the dual variables from Step 2, solve the pricing problem with Algorithm B to identify if one or more variables have negative reduced costs. If so, add all variables with negative reduced costs to RMP's column pool and go back to Step 2; else go to Step 4.

Step 4: Using the dual variables from Step 2, solve the pricing problem with Algorithm A to identify if one or more variables have negative reduced costs. If so, add all variables with negative reduced costs to RMP's column pool and go back to Step 2; else go to Step 5.

Step 5: Fix the largest fractional variable to 1 and check if an integer solution is obtained. If not, go back to Step 2; else stop.

This algorithm was developed after experimenting with various alternative heuristic ideas, and each step was chosen carefully based on the computational performance with and without it. Our computational experiments revealed that Step 5 helps by improving the computational performance substantially while increasing the optimality gap by very little or nothing. Also, we found that decomposing the pricing problem's solution process into two steps, i.e., using Algorithm B in Step 3 and Algorithm A in Step 4, was a vital part of the computational speedup that we achieved. Without this, we would not have been able to finish all our experiments in reasonable amounts of time to accomplish this research project. More details about this two-step approach are provided in Section 2.3.3.

2.3.3 Solution to the Pricing Problem

Researchers have proposed and implemented a variety of methods for solving the pricing problem. *Multi-Label Shortest Path* (MLSP) is a commonly used pricing algorithm in the crew pairing context [48]. Unlike the deterministic version, our robust crew pairing problem involves a more complicated objective function including the planned costs and six types of penalty costs. Furthermore, as explained in Section 2.4, our network size is the largest among all existing research studies addressing any variety of the robust crew pairing problem. Therefore, we cannot directly use an existing method to solve the problem to near optimality in a limited time. Therefore, we develop a new two-step approach to solve the pricing problem to optimality. This approach is presented in Appendix A.1.

2.4 Case Study

In this section, we apply the models presented in Section 2 and the solution methods presented in Section 2.3 (and the Appendices) to four networks from two airlines across multiple time periods. We use confidential airline data containing crew scheduling samples acquired from these two airlines to calibrate and validate our parameterized crew pairing models. The data sources and data preprocessing steps are described in Section 2.4.1, while the computational performance of our models is highlighted in Section 2.4.2. Section 2.5 presents the detailed calibration and validation results which demonstrate the accuracy and stability of our approach.

2.4.1 Data Source and Data Preprocessing

We acquired crew schedule samples from one major regional carrier (RC) and one major network legacy carrier (NLC) in the U.S. The RC has a homogenous fleet consisting of only one fleet family and the data available to us spanned two full months, namely, March and April 2014. The NLC's operations consisted of several different fleet families. However,

for our computational experiments we chose only the three largest networks, namely, those operated by A320, B737 and B757 aircraft types, because the others were much smaller in size. The NLC's crew scheduling data sample spanned one full year, from August 2013 to July 2014.

Aside from this confidential data, we also used the Airline On-Time Performance (AOTP) database from the BTS website (Bureau of Transportation Statistics [28]) which contains on-time arrival data for domestic flights by all major U.S. carriers. Most importantly for our purposes, AOTP provides tail number for each flight, which is a key piece of information useful to track aircraft rotations in real-world airline schedules. Since our data is obtained from two separate sources, some data preprocessing steps, including data cleaning and merging, needed to be performed before using the data for model calibration and validation. We refer the reader to the Appendix A.3 for the detailed preprocessing steps.

We obtained the true values of the planned crew cost parameters, such as $\delta, \epsilon and \zeta$, as well as the values of the lower limits on the crew sit times and the crew rest times, and the upper limits on the maximum duty flying time and the maximum duty elapsed time from both the airlines represented in our data samples. Finally, note that all our data is related to the cockpit (and not the cabin) crew schedules, which are more stringent in their regulations and hence are expected to be responsible for a majority of the crew-propagated delays and disruptions. So our analysis is restricted to the cockpit crew schedules only, and hence deals with a large part, but not all, of the crew-propagated delays and disruptions. Note that this is a limitation of the available data and not of our methodology which would be valid if we were to perform a similar analysis with the cabin crew scheduling data.

2.4.2 Computational Experiments

The CPLEX 12.5 solver with its default settings is used to solve all the linear and integer optimization problems. An 8-thread / 4-core Intel i7-X5600CPU with 8GB RAM and Windows 7 Professional as the operating system was used for all computational experiments.

Table 2.1: Computational performance of our heuristic

Network Size (Flights)	Pricing Approach	Root LP Bound	Our Integer Solution	Gap	Solution Time (hours)
102	Algorithm A Only	398.36	398.37	0.025%	< 0.1
	Algorithm B + Algorithm A	398.36	398.37	0.025%	< 0.1
3300	Algorithm A Only	12760.43	12832.12	0.56%	2
	Algorithm B + Algorithm A	12760.43	12832.12	0.56%	2

For demonstrating the computational performance of our crew pairing optimization approach, we consider two networks: one with 102 flights and the other with 3300 flights. We first solve the root node LP relaxation to optimality to get a lower bound on the optimal objective function value. This is listed in the third column of Table 2.1. Then using the method described in Section 2.3.2, we obtain a feasible, but not necessarily optimal, solution of the integer optimization problem. Its objective function value is listed in the fourth column. Fifth column gives the gap between the values in the third and fourth columns by dividing the difference between the two by the value in the third column. Note that this gap gives an upper bound on the true optimality gap of our heuristically obtained solution. The last column gives the total runtime for obtaining the solution in the fourth column.

The second column lists the solution approach. We first list the performance of our overall heuristic using the exact SPPRC method, (as described in Appendix A.1), i.e., without using Algorithm B. We also list the performance of our modified method, i.e., when using both Algorithm A and Algorithm B. Across all cases in Table 1, the gap was at most 0.56%. For the small network, Algorithm B does not help in speeding up because Algorithm A alone is sufficient to solve it within a few minutes. However, in case of the large network with 3300 flights, the combined use of Algorithm A and Algorithm B, as described in Section 2.3.2, significantly reduces the overall computational time from 10 hours to 2 hours. Similar improvements were observed in all our large network instances. This demonstrates the value of using our modified two-step pricing approach.

2.5 Calibration and Validation Results

2.5.1 Calibration Results

Table 2.2 lists the estimated parameters resulting from our calibration process as described in Section 2.4.1, while Table 2.3 lists the penalty function values corresponding to each feature.

Table 2.2: Parameter results

Feature Type	Parameter	RC	NLC-A320	NLC-B737	NLC-B757
Type 1	α_1	1	0.3	0.3	0.8
	β_1	1.5	0.65	1	3
Type 2	α_2	0.5	0	0	1.5
	β_2	0.15	0	0	1.1
Type 3	α_3	0.4	1.1	3	1
	β_3	1.4	0.4	1.25	1
Type 4	α_4	2	1.65	2	3.8
	β_4	1.5	0.5	0.7	1.3
Type 5	γ_5	0	0.025	0.08	0
Type 6	γ_6	0.05	0.07	0.4	0

Table 2.3: Penalty function values

Cost Type		Planned	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	%age
RC	Airline Sample	4590.40	105.87	42.05	0	3.15	0	33.90	3.87%
	Without Calibration	4143.38	293.72	46.94	4.08	93.60	0	43.34	10.41%
	With Calibration	4244.99	91.23	40.66	0.41	8.44	0	34.80	**3.97%**
NLC-320	Airline Sample	4800.45	5.26	0	15.91	19.39	3.42	28.07	1.48%
	Without Calibration	4515.51	13.05	0	62.24	148.29	5.30	37.45	5.57%
	With Calibration	4539.24	5.27	0	9.48	12.53	3.38	28.00	**1.28%**
NLC-B737	Airline Sample	7448.00	0.77	0	53.74	32.75	9.44	182.4	3.61%
	Without Calibration	6696.80	9.35	0	300.36	280.52	12.56	280.00	11.65%
	With Calibration	6773.94	2.05	0	15.77	12.21	10.88	175.20	**3.09%**
NLC-B757	Airline Sample	976.18	0.44	1.08	0.47	3.52	0	0	0.56%
	Without Calibration	925.08	0.41	2.17	2.28	22.25	0	0	2.85%
	With Calibration	927.09	0	0	0.77	6.26	0	0	**0.75%**

Tables 2.2 and 2.3 present results using four distinct networks, namely, regional carrier's complete network (RC) excluding the flights that were filtered out in pre-processing, and the network legacy carrier's networks using the A320 fleet family (NLC-A320), the

B737 fleet family (NLC-B737), and the B757 fleet family (NLC-B757). The numbers of flights in the RC, NLC-A320, NLC-B737 and NLC-B757 networks were 2432, 1200, 1840, and 147, respectively. All experiments were conducted over a seven day time horizon and the maximum number of duties allowed in a single crew pairing was four in all cases.

In addition to the penalty function values corresponding to each feature, Table 2.3 lists the planned cost values for comparison purposes. In the last column, the total penalty cost as a percentage of the planned cost is listed. For each network, there are three rows. All three rows provide the components of the objective function evaluated using the calibrated parameter values listed in Table 2.2. The first row uses the actual airline-provided crew pairings. The second row uses the crew pairings obtained by solving the crew pairing optimization problem by setting all parameters to 0. Finally, the third row uses the crew pairings obtained by solving the crew pairing optimization problem by setting all parameters to their calibrated values (listed in Table 2.2). Note that the calibration algorithm does not explicitly attempt to match the planned cost values, because our aim is to match the CPDD potential alone. Yet, for all four networks, the planned costs of the crew pairing solution generated by our approach are found to be closer to the actual airline-provided crew schedules with calibration than without calibration. Across networks and cost types, the cost values with calibration were found to be closer (in most cases significantly closer), than the cost values without calibration, to the airline-provided crew pairing solutions in 22 out of the 23 network-cost type combinations in Table 2.3. Note that we have excluded the network-cost type combinations where all three values are zeros, which happens in 5 instances. In Section 2.5.2, we provide a metric for an easy comparison of this degree of closeness in the form of a percentage error measure.

Tables 2.2 and 2.3 exhibit several differences between the four networks. Some of these differences reflect the differences in the crew pay and crew legality rules. For the NLC-B737 and NLC-A320 networks, crew pay does not depend on the time away from base (i.e., parameter ζ in (1) equals 0). As a result, there is no tradeoff associated with the

length of the rest period. For networks with nonzero ζ values, having shorter rest periods can cause additional delay propagation while having longer rest periods can add to the planned crew costs. Absent this tradeoff, for the NLC-A320 and NLC-B737 networks, the optimization simply sets the rest period lengths such that the Type 2 penalty function value is zero irrespective of the values of α_2 and β_2. For the RC network, we find that irrespective of the values of the Type 5 parameters α_5 and β_5, Type 5 penalty function value equals zero. Recall that Type 5 penalty function penalizes the number of crew-base purity violations. Because of the simple hub-and-spoke structure of the RC network, most crew travel from a hub to a spoke and back, and there isn't much opportunity for changing the Type 5 penalty cost by varying γ_5. Therefore, for the RC network, γ_5 value remains 0 even after calibration. Finally the NLC-B757 network is the smallest among the four, due to which crews usually don't have too many alternatives other than staying with the aircraft and the crews do not end up going more than a distance of 1 unit away from the crew base in the adjacency graph. This simplified structure of the network explains why both Type 5 and Type 6 parameters and the corresponding penalty function values are set to 0 for the NLC-B757 network.

Although these four flight networks vary in size, and although the absolute CPDD level cannot be directly compared across the four networks, the values in the last column of Table 2.3 range between 0.75% and 4% across all four networks, for the airline-provided crew schedules and also for the solution generated by our calibrated model. These numbers are much higher for the crew-schedules generated using the uncalibrated model. These results demonstrate that our approach generates crew schedules whose balance between planned and operational costs is similar to that of the actual crew-scheduling solution used by the airlines. Previous studies involving robust crew pairing optimization, such as Yen and Birge [152], have emphasized the importance of finding the right tradeoff between the planned and operational costs. They test effects of different penalty parameters to control this tradeoff, but do not provide explicit insights into the right tradeoff values. Our results,

for the first time, allow us to get a measure of the perceived balance between the planned costs and the penalty costs as reflected by the airlines' actual crew scheduling practices. Table 2.3 suggests that the right balance between the planned and penalty costs across the four networks is in a relatively narrow range of 0.75% to 4% and is thus quite stable across airlines and aircraft families.

Unlike previous studies, such as Schaefer et al. [111], Shebalov and Klabjan [116], Gao et al. [60] which focus on minimizing a subset of the factors affecting the CPDD, we use a more comprehensive approach by including a wider variety of factors. Unlike Yen and Birge [152] who consider the total expected cost of future disruptions, our approach can give a separate ratio between the penalty cost corresponding to each feature of the operational cost and the planned cost. By allowing penalty costs of each component to be assessed separately, we get a clearer understanding of the relative importance of each component as perceived by the airlines.

Table 2.3 provides some preliminary evidence of the effectiveness and accuracy of our calibration framework. However, there are several shortcomings of using the in-sample penalty costs to assess the similarity of our solutions to the airline-provided crew schedules. First, this in-sample comparison has an inherent bias because we are using the same data samples to calibrate and to test the accuracy. We address this concern in Section 2.5.2 by presenting results of computational experiments where the parameter calibration is performed using one dataset and then other datasets are used to assess the out-of-sample accuracy of our approach. Second, we are measuring the closeness of the two solutions using penalty functions, which themselves depend on the calibrated parameter values. To make our comparisons more meaningful, we compare the distributions of the actual feature values in Section 2.5.3. Finally, all methods used by us for evaluating the accuracy of our approach depend on the features that we deem to be good proxies for the CPDD. While many of these were chosen and are well-supported by previous research studies, they are not likely to be precise measures of the CPDD. Therefore, a true test of the performance

of our approach can only be conducted by comparing the actual CPDD. This concern is addressed in Section 2.6.1.

2.5.2 Out-of-Sample Validation Results

This section demonstrates the accuracy and stability of our results through out-of-sample validation. First, in Tables 2.4 through 2.7, we present the results where the calibration and validation datasets belong to two different time periods for the same airline and for the same fleet family. For the RC network, we choose March 2014 as the calibration set and April 2014 as the validation set. For the three NLC networks, we select the first week of one month from each quarter to represent flight schedules through a full year. Specifically, we use January 2014 data for calibration and perform validation using datasets from April 2014, July 2014 and October 2013. Additionally, February 2014 dataset is also used to perform validation for a scenario where the calibration and validation datasets are not too far apart in time from each other. The intent of this validation is to test the validity of using parameters calibrated using one time period to predict crew schedules for another time period for the same airline and the same aircraft family. If the results are found to be stable across time periods, then this allows us to use crew scheduling data samples from one period to estimate crew schedules for other periods and thus reduces our data requirements if we were to estimate the CPDD across long periods of time.

Let $C_i, i \in \{1, ..., 6\}$ be the Type i penalty cost associated with the crew schedule generated by our approach, and $C_i^{Airline}$ be the Type i penalty cost associated with the corresponding airline-provided crew schedule. Then we define the Absolute Percentage Error (APE) as $\frac{|C_i - c_i^{Airline}|}{\sum_i^N c_i^{Airline}}$, where N is the total number of components of the penalty cost function, i.e., the total number of robustness features. Note that this is not a commonly used method of error representation, but is chosen because it offers certain advantages for our problem setting. First, the choice of denominator $(\sum_i^N c_i^{Airline})$ in the APE expression guarantees that we do not have issues related to division by zero. Contrast this choice of

denominator with a more standard $C_i^{Airline}$ term as the denominator, which would have led to division-by-zero issues in many cases, such as the Type 3 error for RC network, as presented in Table 2.3. Second, since we use the same denominator for all features, it is easy and meaningful to compare the percentage errors across the different features.

Tables 2.4 through 2.7 list the APEs for each feature and also the average and maximum values across features. The columns titled "Before" and "After" list the errors for crew pairing solutions generated using uncalibrated and calibrated parameters respectively. Note that as described in Section 2.5.1, in all cases the penalty function evaluation is performed using the calibrated parameters. Looking at the results presented in Tables 2.4 through 2.7, several observations can be made. Errors are substantially lower in most cases after calibration than before. The improvement is especially clear when looking at the average or maximum values of the APEs across types. Average and maximum APEs are reduced substantially by the calibration process and a reduction is observed across all calibration and validation datasets. In many cases the reduction is by one or more orders of magnitudes. The APEs are slightly lower in the out-of-sample validation datasets compared with the in-sample calibration datasets, especially for the RC network. However, the out-of-sample APEs are consistently reduced by the calibration process demonstrating the stability and effectiveness of our approach. Moreover, seasonality is not found to play a significant role in terms of the errors. The out-of-sample validation errors did not worsen and stayed stable as the time between the calibration and validation datasets increased from 1 month (for February) to 6 months (for July). The consistently lower error values with the calibrated parameters, as measured individually, using averages, or using maximum values, indicate that our approach produces crew pairings that are stable across time periods of up to several months.

Tables 2.8 through 2.11 present the cross-validation results where the validation is performed on a dataset which belongs to a different airline, a different fleet family, and in some cases, a different time period compared with the calibration dataset. This constitutes

Table 2.4: Cross-time period validation results: RC

Data	Calibration (Mar)		Validation (Apr)	
	Before	After	Before	After
Type 1	101.6%	7.9%	58.9%	28.8%
Type 2	2.6%	0.8%	1.9%	0.3%
Type 3	2.2%	0.2%	1.6%	0.1%
Type 4	48.9%	2.9%	26.2%	0.2%
Type 5	0	0	0	0
Type 6	5.1%	0.5%	2.9%	0.4%
Average	26.7%	2.0%	15.3%	5.0%
Maximum	101.6%	7.9%	58.9%	28.8%

Table 2.5: Cross-time period validation results: NLC-A320

Data	Calibration (Jan)		Validation (Feb)		Validation (Apr)		Validation (Jul)		Validation (Oct)	
	Before	After	Before	After	Before	After	Before	After	Before	After
Type 1	10.8%	0.0%	12.0%	3.1%	9.3%	2.5%	11.3%	2.6%	21.1%	5.5%
Type 2	0	0	0%	0%	0	0	0%	0%	0%	0%
Type 3	64.3%	8.9%	59.8%	4.0%	39.2 %	12.9%	40.8%	8.6%	70.9%	11.5%
Type 4	178.9	9.5%	228.9%	8.2%	140.3%	31.0%	134.6%	23.1%	299.2%	0.3%
Type 5	2.6%	0.1%	1.9%	0.7%	3.0%	1.6%	3.2%	2.6%	4.4%	3.3%
Type 6	13.0%	0.1%	12.0%	1.5%	9.4%	1.9%	8.2%	3.7%	18.9%	5.5%
Average	**44.9**%	**3.1**%	52.4%	2.9%	33.5%	8.3%	33.0%	6.8%	69.1%	4.4%
Maximum	178.9%	9.5%	228.9%	8.18%	140.3%	31.0%	134.6%	23.1%	299.2%	11.5%

Table 2.6: Cross-time period validation results: NLC-B737

Data	Calibration (Jan)		Validation (Feb)		Validation (Apr)		Validation (Jul)		Validation (Oct)	
	Before	After	Before	After	Before	After	Before	After	Before	After
Type 1	3.1%	0.5%	3.9%	0.9%	2.0%	0.3%	2.6%	0.9%	3.1%	0.7%
Type 2	0	0	0%	0%	0	0	0%	0%	0%	0%
Type 3	88.4%	13.6%	76.3%	7.8%	48.9 %	13.4%	68.1%	12.8%	93.8%	6.9%
Type 4	88.8%	7.4%	86.9%	5.2%	60.7%	10.3%	70.9%	10.5%	100.2%	8.2%
Type 5	1.1%	0.5%	1.1%	1.0%	2.0%	0.6%	2.3%	0.7%	0.4%	1.1%
Type 6	35.0%	2.6%	30.2%	14.1%	26.3%	7.4%	28.0%	6.8%	35.3%	18.0%
Average	**36.1**%	**3.1**%	33.1%	4.8%	23.3%	5.3%	28.7%	5.3%	38.8%	5.8%
Maximum	88.8%	13.6%	86.9%	14.1%	60.7%	13.4%	70.9%	12.8%	100.2%	18.0%

Table 2.7: Cross-time period validation results: NLC-B757

Data	Calibration (Jan)		Validation (Feb)		Validation (Apr)		Validation (Jul)		Validation (Oct)	
	Before	After	Before	After	Before	After	Before	After	Before	After
Type 1	0.5%	8.0%	0.0%	2.9%	0.0%	0.0%	3.7%	0.8%	3.8%	3.8%
Type 2	19.8%	19.6%	0.0%	0.0%	161.6 %	0.0%	16.0%	0.8%	105.5%	10.2%
Type 3	32.8%	5.4%	14.1%	3.9%	54.7 %	0.0%	17.4%	19.8%	20.3%	9.6%
Type 4	339.9%	49.7%	185.1%	92.0%	1080.8%	25.1%	309.9%	79.1%	321.7%	39.4%
Type 5	0	0	0%	0%	0	0	0%	0%	0%	0%
Type 6	0	0	0%	0%	0	0	0%	0%	0%	0%
Average	**65.5**%	**13.8**%	33.2%	16.5%	216.2%	4.2%	57.8%	16.8%	75.2%	10.5%
Maximum	339.9%	49.7%	185.1%	92.0%	1080.8%	25.1%	209.9%	79.1%	321.7%	39.4%

an important test of our calibration approach because realistically we cannot expect crew scheduling samples to be available for all combinations of airlines, fleet types and time periods. Instead, if we are able to access a small crew scheduling sample from one airline, for one fleet family, and for one time period, it is desirable to use that sample to calibrate parameters of a model that can then be used to generate crew schedules for other airline types, other fleet types and/or other time periods. Tables 2.8 through 2.11 present these cross-validation results where the four chosen combinations of networks and time periods are RC March 2014, NLC-B737 January 2014, NLC-B757 January 2014, and NLC-A320 January 2014. In each table, we use the parameter sets obtained by calibration over the networks listed in the top row.

Each table represents results of validation using a single network and time period combination specified in the table caption. The intent of this validation is to test whether our parameters calibrated for one combination of airline, fleet family and time period still perform well for other combinations of airlines, fleet families and time periods.

Table 2.8: Validation across airline, fleet family and time period for the RC network for Mar 2014

Data	RC (Calibration))		NLC-A320		NLC-B737		NLC-B757	
	Before	After	Before	After	Before	After	Before	After
Type 1	101.6%	7.9%	21.4%	6.7%	6.1%	0.6%	43.2%	1.1%
Type 2	2.6%	0.8%	22.0%	16.5%	0	0	0.8%	7.5%
Type 3	2.2%	0.2%	28.8%	11.6%	31.1%	16.8%	18.5%	1.8%
Type 4	48.9%	2.9%	50.8%	21.1%	26.4%	7.4%	149.5%	51.2%
Type 5	0%	0%	0.3%	1.1%	0.4%	2.3%	0	0
Type 6	5.1%	0.5%	6.5%	1.4%	15.7%	7.7%	0	0
Average	**26.7**%	**2.1**%	**21.6**%	**9.7**%	**13.3**%	**5.8**%	**35.3**%	**10.3**
Maximum	101.6%	7.9%	50.8%	21.1%	31.1%	16.8%	149.5%	51.2

Tables 2.8 through 2.11 show that the average and maximum errors (APEs) after calibration are much smaller when compared to those before calibration for all combinations of the calibration and validation datasets. However, when compared with the calibration errors, the validation errors are typically larger. This is especially obvious in Table 2.8 where the calibration is performed using the RC network for March 2014 and the validation is

Table 2.9: Validation across airline, fleet Family and time period for the NLC-A320 network for Jan 2014

Data	NLC-A320 (Calibration)		RC		NLC-B737		NLC-B757	
	Before	After	Before	After	Before	After	Before	After
Type 1	10.8%	0.0%	36.9%	38.2%	2.4%	0.2%	34.1%	4.8%
Type 2	0	0	22.0%	16.5%	0	0	6.6%	6.3%
Type 3	64.3%	8.9%	1.1%	0.0%	52.6%	5.3%	29.7%	0.5%
Type 4	178.9%	9.5%	85.9%	4.0%	73.6%	3.0%	895.2%	52.4%
Type 5	2.6%	0.1%	0	0	2.9%	1.7%	0	0
Type 6	13.0%	0.1%	8.2%	0.6%	25.8%	16.6%	0	0
Average	**44.9**%	**3.1**%	**25.7**%	**9.9**%	**26.2**%	**4.5**%	**160.9**%	**10.7**
Maximum	178.9%	9.5%	85.9%	38.2%	73.6%	16.6%	895.2%	52.4

Table 2.10: Validation across airline, fleet family and time period for the NLC-B737 network for Jan 2014

Data	NLC-B757 (Calibration)		RC		NLC-A320		NLC-B757	
	Before	After	Before	After	Before	After	Before	After
Type 1	3.1%	0.5%	60.2%	28.9%	17.1%	7.4%	27.6%	2.9%
Type 2	0	0	7.2%	22.6%	0	0	38.6%	0.3%
Type 3	88.4%	13.6%	2.6%	0.3%	120.1%	13.4%	37.0%	3.3%
Type 4	88.8%	7.4%	118.7%	0.2%	242.9%	20.6%	685.7%	80.6%
Type 5	1.1%	0.5%	0	0	3.5%	3.6%	0	0
Type 6	35.0%	2.6%	12.1%	5.8%	44.0%	33.8%	0	0
Average	**36.1**%	**4.1**%	**33.5**%	**9.6**%	**71.3**%	**13.1**%	**131.5**%	**14.5**
Maximum	88.8%	13.6%	118.7%	28.9%	242.9%	33.8%	685.7%	80.6

Table 2.11: Validation across airline, fleet family and time period for the NLC-B757 network for Jan 2014

Data	NLC-B737 (Calibration)		RC		NLC-A320		NLC-B737	
	Before	After	Before	After	Before	After	Before	After
Type 1	0.5%	8.0%	21.4%	25.0%	2.5%	2.6%	0	0
Type 2	19.8%	19.6%	50.4%	19.6%	0	0	0	0
Type 3	32.8%	5.4%	3.1%	1.4%	69.4%	6.1%	70.4%	6.4%
Type 4	339.9%	49.7%	87.0%	33.5%	101.5%	16.5%	48.3%	0.1%
Type 5	0%	0%	0	0	2.3%	1.8%	3.2%	5.2%
Type 6	0	0	6.3%	0.9%	5.7%	2.5%	13.9%	4.0%
Average	**65.5**%	**13.8**%	**28.0**%	**13.4**%	**30.2**%	**4.9**%	**22.6**%	**2.6**%
Maximum	339.9%	49.7%	87.0%	33.5%	101.5%	16.5%	70.4%	6.4%

performed using the three NLC networks for January 2014. This seems to suggest that the three NLC networks are more "simila" to each other in terms of their calibrated parameters than the similarity between NLC and RC networks. This is not surprising given that the RC network exhibits many differences in the network structure, schedules and flight durations when compared with the three NLC networks. Moreover, Tables 2.9 and 2.10 together sug-

gest that the parameters for the NLC-A320 and NLC-B737 networks are especially similar to each other as reflected by their low cross-validation errors. This phenomenon can also be explained by the fact that A320 and B737 aircraft families are similar to each other. They are both single aisle, twin-engine aircraft with similar seating capacity and range capabilities causing their flight networks to also look similar to each other.

Thus, Tables 2.8 through 2.11 provide several interesting insights. First, they demonstrate that the out-of-sample validation errors are considerably lower using the calibrated than the uncalibrated parameters even when the calibration was performed using a crew-scheduling sample from a different airline type and/or fleet family. However, we also note that the error reduction by using the calibrated rather than the uncalibrated parameters is greater when the calibration and validation datasets are more similar, in terms of airline type and fleet family. This suggests that, on one hand, when estimating crew schedules for a given flight network, it is advisable to use a parameter set that has been calibrated using a flight network that shares as many of its attributes as possible. On the other hand, using any set of calibrated parameters is still likely to be considerably better than using uncalibrated parameters. Even if the calibrated parameters are from a different time period, different airline, and/or different fleet family, they improve the accuracy considerably compared with the uncalibrated parameters, that is, compared with solving the deterministic crew-scheduling problem. Thus, while it is advisable and beneficial to have a wide variety of airline crew schedule samples, our calibration approach enhances the degree of similarity of the generated crew-pairing solution with the actual pairing solution used by the airline even when crew sampling data are relatively scarce.

2.5.3 Validating Crew Pairing Distributions

In this section, we perform additional validation of our results by directly comparing the distributions of the features that affect the crew-propagated delays and disruptions (CPDD) for our results against the distributions of those features for the airline-provided crew pair-

ing solutions. Note that, we do not expect our solution to result in precisely the same crew pairings as those scheduled by the airline. Instead, our goal is to ensure that the distributions of the features affecting the CPDD are similar between our solution and the airline-provided solution so that the two crew pairing solutions possess similar CPDD potential. This validation approach is similar in spirit to that used by Barnhart et al. [20] to compare the distributions of features of passenger itinerary flows.

We consider the following distributions for validation purposes.

1. Distribution of the flying time in a duty.
2. Distribution of the elapsed time in a duty.
3. Distribution of the scheduled sit times.
4. Distribution of the scheduled rest times.

These correspond to Features 1 through 4 described in Section 2.2.1. The last two features are not included in this type of validation to avoid redundancy. Features 5 and 6 are simply the counts of occurrences of crew base purity violations and aircraft changes within a crew duty, and hence are fully represented by the penalty function comparisons in Section 2.5.2.

The Chi-squire statistic and the Kolmogorov-Smirnov statistic are two commonly used metrics for comparing two distributions to each other. The lower the values of these statistics, the more similar are the two distributions. Table 2.12 compares the distributions of these four features. For the RC network, the calibration is performed using the March 2014 dataset and the validation is performed using the April 2014 dataset while for the three NLC networks, the calibration is performed using the January 2014 dataset and the validation is performed using the February 2014 dataset. Note that we do not present the rest time distributions for the NLC-A320 and NLC-B737 networks for the reasons mentioned in Section 2.5.1. These results in Table 2.12 further reinforce our conclusion that the calibrated models generate crew-pairing solutions that are very similar to those provided by the

airline in terms of the distributions of the CPDD potential, when tested on both in-sample and out-of-sample datasets. In almost all cases, the calibrated parameters yield a better fit to real-world distributions compared with the uncalibrated ones and in many cases the improvement is large.

Table 2.12: Validating distributions of crew pairing solution features

Dataset	Feature	Chi-Square				Kolmogorov-Smirnov			
		Before		After		Before		After	
		Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
RC, Calibration (March 2014)	Flying Time	48.86	0	0.15	0.9852	0.67	0.03	0.33	>0.2
	Elapsed Time	94.19	0	36.22	0	0.40	>0.2	0.20	>0.2
	Sit Time	89.67	0	106.84	0	0.50	0.18	0.50	0.18
	Rest Time	6.95	0.0735	1.27	0.7363	0.33	>0.2	0.33	>0.2
RC, Validation (April 2014)	Flying Time	40.55	0	7.26	0.0641	0.67	0.03	0.33	>0.2
	Elapsed Time	83.15	0	3.58	0.3105	0.40	>0.2	0.20	>0.2
	Sit Time	157.16	0	7.64	0.89	0.75	<0.01	0.50	0.12
	Rest Time	6.63	0.0847	17.74	0.0005	0.33	>0.2	0.33	>0.2
NLC-A320, Calibration (January 2014)	Flying Time	73.47	0	1.25	0.741	0.75	<0.01	0.25	>0.2
	Elapsed Time	171.22	0	1.32	0.7244	0.75	<0.01	0.25	>0.2
	Sit Time	42.13	0	13.3	0.0099	0.40	>0.2	0.20	>0.2
NLC-A320, Validation (February 2014)	Flying Time	60.36	0	8.95	0.03	0.75	<0.01	0.25	>0.2
	Elapsed Time	160.34	0	10.75	0.0132	0.50	0.17	0.25	>0.2
	Sit Time	55.16	0	31.90	0	0.40	>0.2	0.20	>0.2
NLC-B737, Calibration (January 2014)	Flying Time	281.96	0	49.77	0	0.67	0.03	0.33	>0.2
	Elapsed Time	284.30	0	21.63	0.0001	0.60	0.07	0.40	>0.2
	Sit Time	81.07	0	27.98	0	0.75	<0.01	0.50	0.12
NLC-B737, Validation (February 2014)	Flying Time	113.51	0	20.09	0.0002	0.50	0.17	0.17	>0.2
	Elapsed Time	262.42	0	11.15	0.0109	0.60	0.07	0.20	>0.2
	Sit Time	54.66	0	33.15	0	0.25	>0.2	0.25	>0.2
NLC-B757, Calibration (January 2014)	Flying Time	3.45	0.3273	0	1	0.25	>0.2	0	>0.2
	Elapsed Time	4.85	0.1831	3.40	0.334	0.40	>0.2	0.20	>0.2
	Sit Time	3.67	0.4525	0.37	0.9849	0.50	0.12	0.25	>0.2
	Rest Time	3.92	0.2702	9.23	0.0264	0.33	>0.2	0.33	>0.2
NLC-B757, Validation (February 2014)	Flying Time	2.38	0.4974	2.31	0.5106	0.25	>0.2	0.25	>0.2
	Elapsed Time	3.32	0.3449	2.23	0.5261	0.20	>0.2	0.20	>0.2
	Sit Time	0.22	0.9944	2.25	0.6899	0.25	>0.2	0.25	>0.2
	Rest Time	7.10	0.0688	2.98	0.3947	0.33	>0.2	0.33	>0.2

2.6 Estimation and Validation of the Crew-Propagated Delays and Disruptions

All calibration and validation results presented in Section 2.5 were based on comparisons of the crew pairing solutions in terms of the values of the penalty function components and distributions of the features that are representative of the crew-propagated delays and disruptions (CPDD). In this section, we focus directly on the CPDD. In Section 2.6.1, we

provide additional validation of our results in terms of the CPDD. Then, in Section 2.6.2 we present the process for estimating the CPDD for any given airline, fleet family and time period, thus highlighting how the crew pairing solutions generated by our research can be used to estimate the CPDD for any given network.

2.6.1 Validation Based on the Crew-Propagated Delays and Disruptions

As argued in Section 2.2, the crew-propagated delays and disruptions (CPDD) depend not only on the crew schedules but also on the root delays and on the operational recovery actions used by the airline. As researchers, typically we are not aware of the exact set of recovery strategies and parameters that were used by the airlines because such information often tends to be confidential. Moreover, neither the researchers nor the airlines are aware of the exact levels of future root delays when planning the crew schedules for those future time periods. These limitations make it difficult to accurately calculate and compare the CPDD values. In this section, we use an existing crew-recovery software tool, named SimAir, for simulating airline operational recovery and for estimating the CPDD. For a detailed description of SimAir, the reader is referred to Rosenberger et al. [110]. Some past studies in robust crew scheduling, such as Schaefer et al. [111], have used SimAir to evaluate the operational performance of the crew schedules.

While SimAir is one of the most sophisticated existing tools available to researchers for simulating airline operations, its recovery approach does not include crew swaps. Thus, we expect SimAir to overestimate the total operational costs to some extent. Moreover, SimAir requires distributions of various components of root delays as inputs, but they are not easy to ascertain accurately. SimAir divides each flight into a variety of phases such as gate departure, taxi-out, take-off, en-route, touch-down, and taxi-in, and models delays in each phase in detail using separate delay probability distributions. For a given flight, estimating these delay distributions empirically requires making several assumptions and

analyzing a lot of historical delay data. As such, this is beyond the scope of our current validation process. Therefore, in order to provide an example of how to validate our crew pairing solutions based on the CPDD, we make assumptions about the delays in different flight phases. For the sake of simplicity, we divide all flights into two categories, namely long-haul and short-haul, and assume the delay distributions to be the same for all flights within a category. To keep the presentation simple, in this section we will focus only on the NLC-B757 network, which is the smallest across the four case study networks described in Section 2.4.

Table 2.13 provides a summary of the root delay distributions for all individual flight phases that we used as inputs to SimAir. If the scheduled block time is greater than 2 hours, then we categorize the flight as a long-haul flight, and otherwise we categorize it as a short-haul flight. The actual flying time equals the scheduled block time of the flight minus a random variable with positive expected value. Therefore, the mean of the Gaussian distributed random variable used for modeling the flying time delay is negative, as indicated in the last two rows of Table 2.13. Values detailed in Table 2.13 are the same as those suggested as default values in the SimAir User's Manual.

Table 2.13: Summary of the SimAir delay distributions

Event	Distribution	Mean (Standard Deviation)	Unit
Departure Gate Delay	Gaussian Distribution	1 (3)	Minutes
Take-off Service Rate	Gaussian Distribution	20 (2)	Take offs per hour
Touch-Down Service Rate	Gaussian Distribution	20 (2)	Touch downs per hour
Taxi-Out Duration	Constant	5	Minutes
Taxi-In Duration	Constant	5	Minutes
Flying Time Delay (Long-Haul)	Gaussian Distribution	-20 (5.948)	Minutes
Flying Time Delay (Short-Haul)	Gaussian Distribution	-20 (5.665)	Minutes

Using SimAir, we test three crew pairing solutions for the NLC-B757 network, namely, the one actually used by the airline, the one generated by solving the deterministic crew pairing optimization problem (i.e., by setting all parameters to 0), and the one generated by our approach using the calibrated parameters. For each of these three crew pairing solutions, we conduct 2000 simulation runs. We test the stability of our simulation results by

calculating the 95% confidence interval for 15-minute flight on-time performance for the three crew pairing solutions. They are found to be [88.14%, 88.44%], [86.13%, 86.43%], and [87.73%, 88.05%], respectively, for the airline sample, the pre-calibration crew pairing solution, and the post-calibration crew pairing solution. These narrow intervals not only demonstrate the stability of our results but also show that the on-time performance for the airline's actual crew pairing solution is much closer to that obtained by our approach after calibration than before calibration. Table 2.14 provides full details of SimAir result comparisons across the three crew pairing solutions. First row of Table 2.14 lists the number of simulation runs while the remaining nine rows list various simulation result summary statistics. Interestingly, we find that the airline sample simulation results are closer to the results after calibration than the results before calibration in terms of each of these nine criteria, and in many cases the similarity improvement due to calibration is substantial. This further validates the effectiveness of our approach.

Table 2.14: SimAir-based validation results

	Airline Sample	Before Calibration	After Calibration
Number of Simulation Runs	2000	2000	2000
15-Minute On-time Performance	88.29%	86.28%	87.89%
Number of Reserve Crew Calls	0	1.99	0.05
Number of Crew Deadheads	0	0.98	0.05
Flight Legs Ferried	3.02%	3.46%	2.95%
Flights Delayed ≤ 0 min	29.41%	28.49%	29.23%
Flights Delayed $(0, 15]$ min	58.88%	57.79%	58.66%
Flights Delayed $(15, 45]$ min	3.47%	3.37%	3.52%
Flights Delayed > 45 min	3.47%	3.37%	3.52%
Canceled Flights	7.28%	9.58%	7.76%

2.6.2 Estimating the Crew-Propagated Delays and Disruptions

In this chapter, we developed an approach to generate crew pairing solutions that are similar to the actual crew pairing solutions used by the airlines in the real world, in terms of their potential for the crew-propagated delays and disruptions (CPDD). As mentioned in Section 2.1, this work has at least three main types of applications. First, it is the first step toward

estimating the extent to which delays and disruptions propagate through crew connections. Second, it allows us to assess and compare the effectiveness of various operational recovery strategies used by the airlines. Finally, it allows us to evaluate and compare the full impact of various candidate strategies for congestion and delay mitigation that are being considered by the airlines, airports, air traffic control system, and the government. In this section, we briefly describe how each of these objectives can be achieved using our results.

Table 2.2 provides four different sets of parameters representing four different airline networks. The robust crew pairing optimization model (2.8)-(2.10) can be solved for each of these four sets of parameters to come up with an estimated crew schedule for any given airline network of interest. As our results in Section 2.5 indicate, when picking the right set of parameters, it is advisable to choose a set that corresponds to a network which is the most similar (in terms of the airline type, fleet type and time period) to the network of interest. However, no matter which parameter set is picked, using calibrated parameters gives a far better fit than solving the deterministic crew pairing model in all cases. Alternatively, the calibration approach described in Sections 2.3 and 2.4 could be used to generate more suitable parameters in case a better-matching airline crew schedule sample is available.

Once the crew schedules are estimated, they can be used to estimate the CPDD. Note that, for accurately estimating delays in a historical dataset, some knowledge or assumption regarding the recovery strategies used by the airlines is necessary. For a given set of root delays and for a given operational recovery strategy, our crew schedules can be used to estimate the historical CPDD values in a relatively straightforward manner. On the other hand, for evaluating and comparing different operational recovery strategies and different congestion mitigation strategies, a root delay simulator such as SimAir can be used in combination with our estimated crew schedules to calculate the full extent of delay reduction achievable by these strategies. Note that, in all cases, total propagated delays and disruptions should be measured by accounting for the propagation through aircraft connections as well as crew connections. However, the aircraft connections are publicly available and

hence are not a bottleneck in this overall process.

2.7 Summary

In this chapter, for the first time, the inverse of the robust crew-pairing-generation problem was presented, formulated, and solved to gain insights into the extent of the robustness of real-world airline crew-scheduling practices. The problem was formulated as one of learning the parameters of the robust optimization objective function using real-world airline crew-scheduling samples. A heuristic solution approach was developed and implemented. It involved solving the forward problem (the robust crew-pairing problem) repeatedly to minimize a similarity measure between the solution of the robust crew-pairing problem and the actual airline crew schedule samples by identifying the optimal set of objective function parameters. The forward problem minimizes the sum of the planned cost and the penalty costs, which penalize the crew pairings for six different features that make them vulnerable to the propagation of delays and disruptions. A sequence of exact methods and heuristic ideas was used to solve this robust crew-pairing problem to near optimality. This allowed the overall parameter calibration problem to be solved in a reasonable amount of time.

Several new insights into the airline crew-pairing generation process were obtained. First, compared with the crew pairings obtained by solving the deterministic crew-pairing problem, the calibrated parameters led to crew pairings that are considerably closer to the actual airline crew schedules in all our experiments. In most cases, the accuracy improvement was substantial. This suggests that airlines do take into account robustness or the potential for propagation of delays and disruptions when creating their crew schedules. Furthermore, we found that the crew pairings calibrated using four different airline networks performed similarly to each other, and much better than the deterministic crew-pairing solutions, in terms of their closeness to the actual crew schedules, even when the calibration

and evaluation were not conducted on the same network. This suggests that the calibrated parameters are relatively stable. Thus, even in cases where the data available for model training are for a network somewhat dissimilar to the one of interest, it is better to use the calibrated parameters than the uncalibrated ones. However, for maximizing estimation accuracy, whenever possible, it advisable to use parameters calibrated with a network that is as similar to the one of interest as possible, in terms of airline type, fleet type, and time period. Finally, this chapter presented, for the first time in the literature, a measure of the trade-off as perceived by the airlines between the crew salary costs and the costs of the crew-propagated delays and disruptions as reflected by the calibrated robust crew pairing objective functions. Across the four networks, the ratio of the penalty costs (representing the costs of the CPDDs) and the crew salary costs was found to lie between 0.5% and 4%. Note that this is inferred based on the crew pairings used by the airlines and not based on the actual costs of these delays and disruptions.

In addition to these insights, as described in Subsection 2.6.2, this research makes the estimated crew pairing solutions available for further research and analysis. We have made this entire model calibration code as well as the resulting calibrated crew-pairing solutions publicly available for future research. These estimated crew-pairing solutions are useful to gauge the extent of delays and disruptions that propagate across the airline networks through crew connections. While these crew-pairing estimates do give a starting point to estimate the CPDDs, the important next step toward accurately estimating historical delay propagation is to develop an understanding of the crew recovery strategies used by the airlines in the real world. Once we have access to a historical sample of actual crew recovery actions, a framework similar to the one developed in this chapter could be used to learn the airline crew recovery optimization process as well. This will be the next step in our research project.

Chapter 3

Airline Timetable Development and Fleet Assignment Incorporating Passenger Choice

3.1 Introduction

Airline profitability is affected by a variety of complex factors, some of which are under an airline's control, including the set of Origin-Destination (O-D) markets being served, the frequency of the flights offered on each segment, the times at which these flights are scheduled, and the fares which are offered. Maintaining consistent profitability levels requires coordination across an airline's departments including network planning, marketing, pricing, revenue management and operations. Airlines have traditionally been at the forefront of innovation in data-driven decision-making and developing automated decision-support tools. Many of the world's leading airlines have been using large-scale optimization models for decades. Yet, completely automated decision-making has been a rarity for some of the most critical decision-making steps taken by airlines across the world. In fact, some of the most critical decisions are still handled on an *ad hoc* basis with no or limited support

from data-driven tools. This may result in missed revenue opportunities and imbalances between the passengers' preferences and the offered traveling options. Timetable development is one such critical decision-making step for which analytical optimization methods remain limited.

The airline planning process consists of a variety of decision-making steps including schedule design, fleet assignment, aircraft routing and crew scheduling, which are typically carried out sequentially. Schedule design itself contains two sub-steps: frequency planning and timetable development (or *timetabling* for short). Frequency planning involves the choice of the number of flights operated daily on each nonstop segment in the airline's network. Timetabling involves determining the scheduled departure and arrival times for each of these flights. The next step, fleet assignment, determines the fleet type to be used for each flight, with the objective of matching the number of available seats with passenger demand. While larger aircraft may lead to higher operating costs, smaller aircraft may result in revenue losses because some passengers may get "spilled" if the number of available seats is smaller than passenger demand. Next, aircraft routing involves generating feasible routes for each individual aircraft while ensuring that maintenance requirements are satisfied. Last, crew scheduling includes crew pairing, i.e., combining flights into sequences starting and ending at a crew base, and crew rostering, i.e., creating monthly schedules from these flight sequences and assigning them to individual crew members. All these decisions, taken together, determine the airline's service offerings and daily operations, and can therefore have a considerable impact on its profitability. While some of them (e.g., fleet assignment) have been the subject of considerable research, others (e.g., timetabling) lack systematic decision-making tools.

Airline timetabling decisions are significantly complicated by the endogeneity of fleet assignment decisions and passenger booking decisions. From a passenger's perspective, important determinants of the attractiveness of each itinerary include the departure and arrival times of the flights in that itinerary, as well as connection times in case of a connecting

itinerary, all of which are defined by flight timetables. For example, itineraries departing during early mornings and late afternoons are typically more attractive than other itineraries to many passengers. Additionally, flight timetables that generate more passenger connection opportunities lead to more flying options for passengers. These are especially useful in low-demand markets which do not support nonstop flights. An obvious goal of timetabling is to schedule flights at times that are most attractive to passengers (also known as peak times). However, a naive implementation of such strategy could violate constraints related to fleet availability and aircraft operations. Also, as mentioned earlier, different fleet types vary in seating capacity and can lead to considerable variations in profit even for the same timetable. In summary, in order to generate feasible and profitable timetables, it is essential to integrate fleet assignment decisions and the dynamics of passengers' booking decisions across alternative itineraries into the airlines' timetabling decision-making process.

Most existing research in airline timetabling considers only *incremental* changes to existing timetables. The commonly used approaches start with a feasible timetable (e.g., from historical operations), and aim to find marginal improvements by retiming of some flight legs and/or using a combination of mandatory and optional flight legs. There is very little research on *comprehensive* (as against incremental) timetabling methods where an optimal timetable is developed from scratch rather than relying on modifications to an existing timetable. This is, in part, due to the mathematical and computational challenges associated with developing and solving integrated timetabling and fleet assignment models. Mathematically, this requires the integration of highly nonlinear and non-convex passenger choice models into the large-scale mixed-integer optimization frameworks necessary to solve timetabling and fleet assignment problems faced by major airlines. Computationally, this requires new algorithms for solving the resulting optimization models within reasonable time horizons to demonstrate the benefits of the comprehensive timetabling approach over incremental approaches that already exist in the literature and in practice. The purpose of the research presented in this chapter is to address these two interrelated challenges.

This chapter introduces an original integrated optimization approach to comprehensive timetabling and fleet assignment under endogenous passenger choice. While our modeling approach can capture a variety of discrete-choice models of passengers' booking decisions, we use here the Generalized Attraction Model (GAM) recently proposed by Gallego et al. [59]. This encompasses, as special cases, many well-known discrete-choice models, such as the multinomial logit model (MNL). We use a linearization technique to integrate the GAM into our optimization model of flight timetabling and fleet assignment while retaining a mixed-integer linear programming structure.

The resulting model, despite being a mixed-integer linear (rather than nonlinear) optimization model, is still highly intractable by commercial solvers due to its extremely large size. We thus develop an original multi-phase solution approach, along with several heuristics, to optimize the network-wide timetable of a major airline carrier within a realistic computational budget. From a practical standpoint, our modeling and computational framework provides decision support for airlines interested in generating new timetables, evaluating or enhancing their existing timetables, or analyzing various business strategies such as considerations of mergers and acquisitions. Furthermore, we demonstrate the value of integrating timetabling decisions with other planning steps, such as frequency planning and revenue management, to improve the airline's overall profitability.

Note that this is a particularly exciting time in the airline industry to develop a *tractable* approach to comprehensive timetabling optimization. In particular, we are aware that at least one of the world's ten largest airline carriers is working closely with a prominent airline optimization software vendor to develop their timetables from scratch. Our discussions with that vendor and multiple large airline carriers have indicated a strong interest in this topic, and have suggested that solution tractability remains the primary challenge in such endeavors. Nevertheless, we are not aware of any study in the existing scientific literature that addresses this challenge explicitly.

3.1.1 Literature Review

In this subsection, we first provide an overview of the fleet assignment model (FAM). We then discuss existing studies that integrate passenger flows and incremental timetabling decisions into the FAM, and limitations thereof. We conclude with a brief review of the recently studied discrete-choice models of passengers' booking decisions that we will integrate into our optimization model.

The airline fleet assignment problem has been a canonical problem in transportation science since the early work of Abara [1] and Hane et al. [67]. The initial model formulations and computational solutions reported in these studies showed the potential for significant profit improvements through the use of optimization models and automated decision support to match aircraft fleet types with passenger demand under a variety of operating constraints. These two studies used a flight leg-based passenger demand assumption. This was then improved by integrating an itinerary-based demand model to capture the effects of hub-and-spoke network operations on optimal fleet assignment decisions [18, 19]. Computationally, FAM involves large-scale optimization problems and, thus faces considerable challenges in solving them to optimality (or near optimality) in reasonable time frames. These challenges are particularly severe under the itinerary-based demand assumption, which increases the model requirements significantly as compared to the flight leg-based demand assumption. Therefore, extensive FAM research has also focused on developing efficient solution approaches based on various techniques, including Benders decomposition (e.g., Jacobs et al. [73]) or very large-scale neighborhood search (e.g., Ahuja et al. [6]).

One of the main challenges in FAM involves capturing the endogeneity of passenger choice, i.e., the impact of airline fleeting decisions on passenger flows across the network of flights Barnhart and Vaze [16]. Two important considerations when modeling such endogeneity are spill (i.e., the revenue lost if the assigned fleet type is unable to satisfy passenger demand) and recapture (i.e., the part of the spilled revenue that is re-captured by

redirecting passengers to other flights of the same airline). These were first addressed by Kniker [81] and Barnhart et al. [18], who combined a Passenger Mix Model (PMM) into the FAM to approximate spill revenue in an itinerary-based modeling framework. PMM was formulated as a large-scale network flow problem and solved using column and row generation techniques. Subsequent work extended these ideas to optimize fleet assignment adjustments in response to improved demand forecasts that become available closer to the day of departure, known as demand-driven re-fleeting [118]. This was integrated into the FAM by Sherali and Zhu [117] in a two-stage setting where the first stage only performs assignment decisions at a higher "fleet family" level, while the second stage adjusts them in response to demand realizations.

Thus, important advances have been made toward capturing passengers' booking decisions and resulting passenger flows into airlines' fleet assignment decisions. However, none of the aforementioned studies considers flight timetabling explicitly. Before proceeding further, we note that airline timetabling decisions, airline fleet assignment decisions and passenger booking decisions are closely interrelated. For example, consider a nonstop segment between Airport A and Airport B with a block time (defined as the difference between the departure time and arrival time) of 2 hours, with two timetabling alternatives:

i. Two flights per day from A to B: 9 am – 11 am (i.e., departure at 9 am and arrival at 11 am) and 3 pm – 5 pm; and two flights per day from B to A: 12 pm – 2 pm and 6 pm – 8 pm.

ii. Two flights per day from A to B: 9 am – 11 am and 6 pm – 8 pm; and two flights per day from B to A: 9 am – 11 am and 6 pm – 8 pm.

Although the two alternatives have the same number of flights per day, they might lead to very different passenger flows, revenues and operating costs in the FAM. For instance, Alternative (ii) might generate higher demand due to the alignment of its flight offerings with peak morning and evening hours. On the other hand, Alternative (i) can be operated

with a single aircraft (assuming a minimum aircraft turnaround time of 1 hour or less), which might lead to lower network-wide costs. Moreover, the profit contribution of these two alternatives also depends on the fleet assignment decisions; for example, Alternative (ii) might not yield much higher revenue than Alternative (i) if the available number of seats on each flight is small. Therefore, any timetabling optimization model needs to integrate fleet assignment decisions and passenger booking decisions into it.

Given the considerable computational requirements of solving the FAM alone, augmenting it to integrate with timetabling decisions and passenger booking decisions results in extremely complex optimization models. For this reason, researchers have mostly focused on incremental timetabling approaches, often by only allowing small deviations from an existing schedule. For example, Lohatepanont and Barnhart [89] consider sets of mandatory and optional flight legs, along with spill and recapture models. The only timetabling flexibility in this model is the possibility of eliminating a subset of the optional flight legs. This approach has been improved by considering stochastic passenger demands [151] and by involving aircraft and passenger delay costs [107]. Other studies have added other decisions into this framework, such as flight block times [76], the re-timing of flight legs under congestion [126], aircraft routing [121], frequency planning and multi-modal competition [32], etc. Most recently, Abdelghany et al. [2] developed a timetabling model for revenue maximization that considers passenger demand shifts among airline competitors. This was formulated using an incremental approach that allowed each flight's departure time to be optimized within a specific time window. Moreover, this formulation included a non-convex objective function, which makes the model impractical to solve in most real-world instances. More closely related to our approach, Wang et al. [142] incorporated passenger choice into a mixed-integer linear program for fleet assignment, where passenger spill and recapture are based on the "attractiveness" of itineraries. They also suggested the possibility of extending their formulation to capture *incremental* timetabling decisions. This contrasts with our goal of addressing the *comprehensive* timetabling problem. More-

over, their computational results suggest that the consideration of passenger choice makes the resulting model extremely hard to solve as compared to the traditional itinerary-based FAM. No results were reported for even the incremental flight timetabling problem.

Turning to the passengers' booking decisions, a separate recent stream of literature has focused on passenger choice modeling in the airline industry [61, 74]. The most commonly used model is the multinomial logit model (MNL), which assumes that each passenger's itinerary choice depends on the utilities derived from that itinerary and from all the alternative itineraries in a given choice set. More advanced approaches such as the mixed multinomial logit model [95] and the generalized extreme value model [25] have been proposed to capture passenger behaviors more accurately. Empirically, Coldren et al. [42] and Koppelman et al. [82] have applied discrete-choice models using data from United Airlines and Boeing, respectively, to identify the drivers of passengers' booking decisions across competing itineraries. Most recently, Lurkin et al. [92] extended the MNL to account for price endogeneity by using an approach based on instrumental variables. We use the model estimated in this study in our computational experiments.

In summary, determining the "optimal airline timetable" remains an open question. All the relevant previous studies are based on incremental improvements to an existing timetable. Recently, empirical discrete-choice model specifications of airline passenger choice behavior have become available, but have not been integrated into airline fleeting or timetabling decisions. Therefore, new methodologies are required to

i. formulate a comprehensive (as opposed to incremental) airline timetabling model

ii. integrate a model of passenger choice that reflects the endogeneity of passengers' booking decisions.

iii. ensure the tractability of the integrated model.

The rest of the chapter is organized as follows. Section 3.2 presents our integrated comprehensive timetabling and fleet assignment model. Section 3.3 presents the multi-phase

solution approach and our acceleration heuristics. Section 3.4 lists our computational results using different test networks with varying sizes to highlight the benefits of our solution approach. Section 3.5 compares the results from our model with those obtained using a variety of incremental approaches. Section 3.6 presents the results of three different extensions of our modeling and algorithmic framework. We summarize the main findings of the chapter and discuss future research opportunities in Section 3.7.

3.2 Integrated Timetable Development and Fleet-Assignment Model

Our overall modeling architecture is shown in Figure 3.1. Its main element is our *integrated timetable development and fleet assignment model*, formulated as a mixed-integer linear program. It takes the perspective of a single airline (the "host airline", henceforth). The model's inputs are the host airline's available fleet and operating costs, on the supply side, and the attractiveness of the host airline's itineraries and its competitors' itineraries, on the demand side. These inputs are obtained from a variety of databases as well as aircraft operating cost models [131] and itinerary choice models [91]. The model is solved using a multi-phase solution approach along with two heuristics: a *Fleet-Fixing* Approach and a *Symmetry-Inducing* Approach. The model then provides the comprehensive flight timetables and fleet assignment solutions.

Before proceeding further, two important observations regarding the scope of this research are noteworthy. First, the research focuses on optimizing the timetable development and fleet assignment decisions from the perspective of *a single airline*, and assumes that the decisions of all other airlines are fixed. In practice, one could argue that any scheduling change from one airline may trigger potential responses from the other airlines. However, given the complexity of the problem, we ignore these competitive dynamics in this research, as commonly done in the fleet assignment and schedule design literature. Sec-

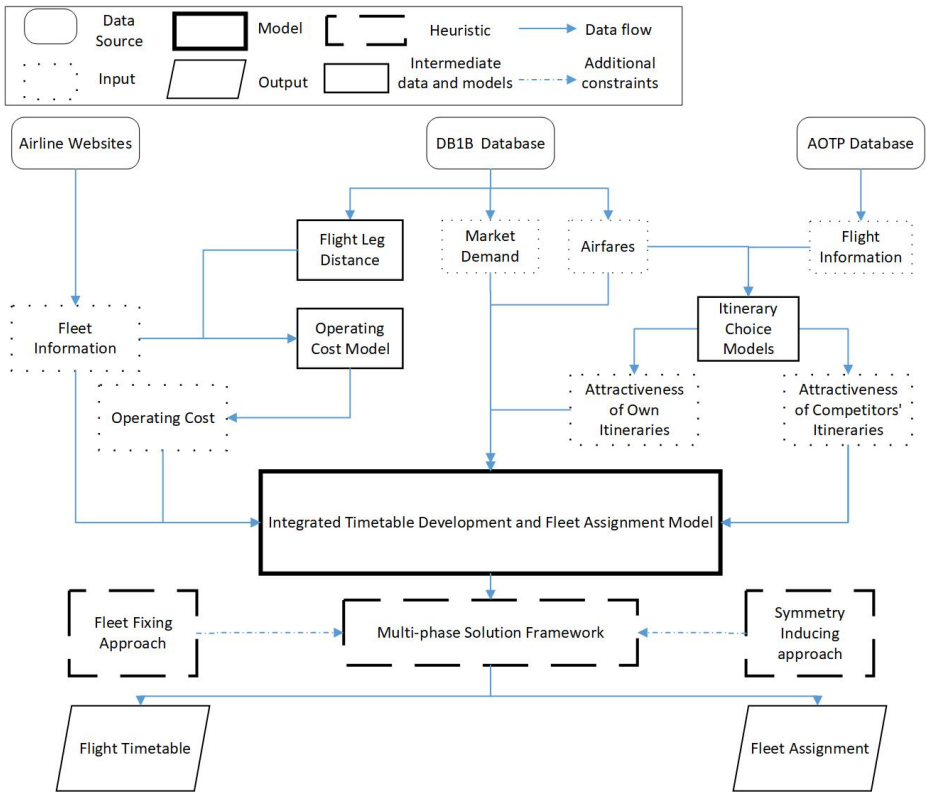


Figure 3.1: Modeling architecture

ond, timetable development and fleet assignment decisions in practice need to take into account a number of additional considerations that lie out of the scope of this model due to the limitations of the publicly available datasets. Examples include airport slot and gate availability constraints—which are typically not too restrictive at most airports in the United States—and airline crew availability constraints. Adding such considerations into the model formulation would be relatively straightforward.

In Section 3.2.1, we first describe the underlying methodology used to integrate a passenger choice into a mixed-integer linear programming model. Next, we formulate our model mathematically in Section 3.2.2.

3.2.1 Passenger Choice Model

A PMM has been the most common approach to model passenger flows in a network of flights. In one of the first prominent PMM studies, Glover et al. [62] computed the passen-

ger flows that maximize operating profit, based on a given fleet assignment solution and passenger demand information. Classical PMM assumes the spill and recapture rates to be known *a priori* for each itinerary. These simplified assumptions about passengers' behaviors have been used in airline scheduling and fleet assignment until recently [81, 18]. However, within the airline revenue management literature, there have been frequent attempts to integrate discrete-choice passenger behavior models into the revenue optimization formulations [96]. While the scheduling model developed in this chapter can leverage a variety of such discrete-choice models, we consider here the Generalized Attraction Model (GAM) recently proposed by Gallego et al. [59].

The GAM was developed to address some inaccuracies stemming from the Independence of Irrelevant Alternatives (IIA) property of the more commonly used multinomial logit model (MNL). These inaccuracies can be described by the so called "Blue-Bus, Red-Bus Paradox" [23]: Consider a person with a 50% probability of traveling by bus and a 50% probability of driving. If a second bus alternative, with identical attributes as the first bus alternative, was added, the IIA property would predict that this person has a 33% probability of taking the existing bus, a 33% probability of taking the newly added bus, and a 33% probability of driving while, in fact, it is more reasonable to expect a 25% probability of taking each bus and a 50% probability of driving. More generally, even if the attributes of the two bus alternatives are not identical, there often exist interdependencies between available and latent alternatives, leading to violations of the IIA property. The GAM addresses this in a systematic manner by assuming each choice probability to be a function of the actual attractiveness of all available alternatives as well as the *shadow attractiveness* of all alternatives (including those not available). In the context of passenger itinerary choice, this can be described as follows.

We consider a set of all itineraries offered by the host airline, denoted by I, and assume that an itinerary set $I' \subset I$ is actually available to choose from. We denote by u_i the passenger utility associated with itinerary $i \in I$. Let A_i be the attractiveness of itinerary i,

defined as $A_i = e^{u_i}$, and let A_O be the attractiveness of the *outside option*, defined as the aggregation of all itineraries of all other airlines as well as the no-fly alternative. According to the MNL, the probability that a passenger will choose any itinerary $i \in I'$ is equal to the ratio of its attractiveness to the total attractiveness of all other alternatives, that is,

$$\pi_i^{MNL} = \frac{A_i}{A_O + \sum_{j \in I'} A_j}, \forall i \in I'; \qquad \pi_i^{MNL} = 0, \forall i \in I \setminus I'. \tag{3.1}$$

In the GAM, by denoting the *shadow attractiveness* of each itinerary $i \in I$ by w_i, the probability that a passenger selects any itinerary $i \in I'$ is modified as follows:

$$\pi_i^{GAM} = \frac{A_i}{A_O + \sum_{j \in I \setminus I'} w_j + \sum_{j \in I'} A_j}, \forall i \in I'; \qquad \pi_i^{GAM} = 0, \forall i \in I \setminus I'. \tag{3.2}$$

If we define adjusted attractiveness values $\tilde{A}_O$ and $\tilde{A}_i$ as $\tilde{A}_O = A_O + \sum_{i \in I} w_i$ and $\tilde{A}_i = A_i - w_i$, Equation (3.2) can be simplified as follows:

$$\pi_i^{GAM} = \frac{A_i}{\tilde{A}_O + \sum_{j \in I'} \tilde{A}_j}, \forall i \in I'; \qquad \pi_i^{GAM} = 0, \forall i \in I \setminus I'. \tag{3.3}$$

Careful choice of the shadow attractiveness values can circumvent the aforementioned "Blue-Bus, Red-Bus Paradox", and thus provide a more accurate representation of passengers' booking decisions. Attraction parameters $\tilde{A}_i$ and the shadow attraction parameters w_i can be obtained based on maximum likelihood estimation and least squares methods [59]. In the remainder of this chapter, we use the GAM model to characterize passenger choice.

3.2.2 Optimization Model Formulation

Throughout this chapter, a "market" is defined as an origin-destination (O-D) pair of airports between which the passengers demand travel. On the other hand, a "segment" corresponds to an ordered pair of airports between which an airline operates nonstop flight(s).

This distinction between markets and segments stems from the fact that some passengers use connecting itineraries with more than one nonstop flight to go from their respective origins to destinations. An "itinerary" represents a sequence of consecutive flights used by a passenger to complete a trip starting at the origin airport and ending at the destination airport. For example, if a passenger flies from San Francisco (SFO) to Denver (DEN) and then to Boston (BOS), the corresponding market is SFO-BOS, the two segments in this trip are SFO-DEN and DEN-BOS, and the itinerary comprises two specific flights, one from SFO to DEN and another from DEN to BOS. A "fare class" is defined as one of the multiple purchasing options provided by the airline in any market, associated with a particular cabin (e.g., economy, business) and various restrictions (e.g., time of purchase). On the demand side, "passenger types" refer to different market segments characterized by similar price sensitivities and time sensitivities. Business and leisure passengers are the two most common passenger types, but they can also be defined at a finer level of granularity.

We represent timetabling and operating decisions in a *time-space network*, where each node is associated with a unique combination of fleet type, airport and time period. The start node of each flight arc corresponds to the scheduled departure time while the end node of each flight arc corresponds to the scheduled arrival time plus the minimum aircraft turnaround time. This is a standard modeling choice in the fleet assignment literature [67]. We consider a full day of operations, which we discretize into 15-minute time periods. We assume that no more than one flight will be scheduled on any given segment during any given 15-minute period, a reasonable assumption in practice. For simplicity of the exposition, we also assume that the flight timetable is repeated daily. Note that this assumption can be easily relaxed in our formulation (e.g., to capture variations in market demand or flight frequencies by day of the week or with seasonality).

We now formulate our integrated model of flight timetabling and fleet assignment.

Sets and Indices

$\mathcal{AP}$: Set of all airports served by the host airline; indexed by ap.

$\mathcal{M}$: Set of all markets served by the host airline; indexed by m.

$\mathcal{S}$: Set of all segments served by the host airline; indexed by s.

$\mathcal{S}_o(ap) \subset \mathcal{S}$: Set of segments served by the host airline with airport ap as origin; indexed by s.

$\mathcal{S}_d(ap) \subset \mathcal{S}$: Set of segments served by the host airline with airport ap as destination; indexed by s.

$\mathcal{T}$: Set of all 15-minute time periods in a day; indexed by t.

$\mathcal{F}$: Set of all fleet types available to the host airline; indexed by f.

$\mathcal{PT}$: Set of all passenger types; indexed by pt.

$\mathcal{CL}$: Set of all fare classes from the host airline; indexed by cl.

$\mathcal{FL}$: Set of red-eye flights of the host airline —that is, set of pairs (s, t) of segments and departure times such that the corresponding flight is en-route at the beginning of the day.

$\mathcal{I}_m$: Set of all itineraries in market m offered by the host airline; indexed by i.

$\mathcal{N}$: Set of all nodes in the time-space network of the host airline; indexed by (f, ap, t), for fleet type f, airport ap, and time period t.

$\tilde{\mathcal{I}}_{m,s,t}$: Set of all itineraries of the host airline in market m which use a flight on segment s with departure period t; indexed by i.

Parameters

$Dem_{m,pt}$: Total demand of type pt passengers in market m.

$A_{i,pt,cl}$: Attractiveness of itinerary i and fare class cl offered by the host airline for passenger type pt.

$\tilde{A}_{i,pt,cl}$: Adjusted attractiveness of itinerary i and fare class cl offered by the host airline for passenger type pt.

$A^0_{m,pt}$: Total attractiveness of all itineraries of other airlines and of the no-fly alternative, in market m, for passenger type pt.

$\tilde{A}^0_{m,pt}$: Adjusted total attractiveness of all itineraries of other airlines and the no-fly alternative, in

market m, for passenger type pt.

$P_{i,cl}$: Estimated ticket price of fare class cl for itinerary i offered by the host airline.

$Ope_{s,f}$: Operating cost of a flight, on segment s, operated using fleet type f by the host airline.

Cap_f : Seating capacity of an aircraft of fleet type f.

$Freq_s$: Number of flights of the host airline per day on segment s.

$Avail_f$: Available number of aircraft of fleet type f for the host airline.

$MinT$: The first time period of the day.

$MaxT$: The last time period of the day.

$D(s,t)$: Scheduled departure time of a flight of the host airline on segment s with end node in time period t (after scheduled en-route time plus minimum aircraft turnaround time).

As described in Section 3.2.1, we use a Generalized Attraction Model (GAM) to characterize passengers' itinerary choice across the set of available alternatives. Note, however, that compared to the u_i and A_i values in Section 3.2.1, we re-define the utility $u_{i,pt,cl}$ and the attractiveness $A_{i,pt,cl}$ here as functions of the itinerary, the passenger type and the fare class. We also re-define the attractiveness (A_O in Section 3.2.1) of the outside option as $A^0_{m,pt}$ so that it is a function of the market and the passenger type. These values are obtained from the passenger utilities derived from the outside option, denoted by $u^0_{m,pt}$. The utilities $u_{i,pt,cl}$ and $u^0_{m,pt}$ are expressed as linear functions of a set of itinerary attributes that affect passenger choice. For each combination of itinerary i and fare class cl, and for each passenger type pt, we denote this attribute vector by $\mathbf{z}_{i,pt,cl}$ and the corresponding vector of linear coefficients by $\beta_{i,pt,cl}$. For instance, the empirical specification provided by Lurkin [91] uses the following attributes: total trip time, number of connections, departure time of the day, ticket price, distance of the itinerary, direction of travel, number of time zones crossed and departure day of the week. Similarly, for each market m and each passenger type pt, we denote the attribute vector associated with the outside option by $\mathbf{z}^0_{m,pt}$ and

the corresponding vector of linear coefficients by $\beta^0_{m,pt}$. Finally, we denote by $w_{i,pt,cl}$ the shadow attractiveness of itinerary i and fare class cl for passenger type pt. We then have

$$u_{i,pt,cl} = \boldsymbol{\beta}_{i,pt,cl} \cdot \mathbf{z}_{i,pt,cl} \qquad \forall i \in I_m, m \in \mathcal{M}, pt \in \mathcal{PT}, cl \in \mathcal{CL} \tag{3.4}$$

$$A_{i,pt,cl} = \exp\left(u_{i,pt,cl}\right) \qquad \forall i \in I_m, m \in \mathcal{M}, pt \in \mathcal{PT}, cl \in \mathcal{CL} \tag{3.5}$$

$$u^0_{m,pt} = \boldsymbol{\beta}^0_{m,pt} \cdot \mathbf{z}^0_{m,pt} \qquad \forall m \in \mathcal{M}, pt \in \mathcal{PT} \tag{3.6}$$

$$A^0_{m,pt} = \exp\left(u^0_{m,pt}\right) \qquad \forall m \in \mathcal{M}, pt \in \mathcal{PT} \tag{3.7}$$

$$\tilde{A}_{i,pt,cl} = A_{i,pt,cl} - w_{i,pt,cl} \qquad \forall i \in I_m, m \in \mathcal{M}, pt \in \mathcal{PT}, cl \in \mathcal{CL} \tag{3.8}$$

$$\tilde{A}^0_{m,pt} = A^0_{m,pt} + \sum_{j \in I_m} \sum_{cl \in \mathcal{CL}} w_{j,pt,cl} \qquad \forall m \in \mathcal{M}, pt \in \mathcal{PT} \tag{3.9}$$

Decision Variables

$$x_{s,f,t} = \begin{cases} 1 & \text{if fleet type } f \text{ is assigned to a flight on segment } s \text{ with departure time period } t \\ 0 & \text{otherwise} \end{cases}$$

$y^-_{f,ap,t}$: Number of aircraft of fleet type f on the ground at airport ap just before period t.

$y^+_{f,ap,t}$: Number of aircraft of fleet type f on the ground at airport ap just after period t.

$\sigma^0_{m,pt}$: Sum of the market shares of all itineraries of the other airlines and of the no-fly alternative in market m for passenger type pt.

$\sigma_{i,pt,cl}$: Market share of passenger type pt, corresponding to the combination of itinerary i and fare class cl.

The full mathematical formulation is provided as follows:

$$\text{Maximize} \quad \sum_{m \in \mathcal{M}} \sum_{i \in I_m} \sum_{pt \in \mathcal{PT}} \sum_{cl \in \mathcal{CL}} (Dem_{m,pt} * P_{i,cl} * \sigma_{i,pt,cl}) - \sum_{s \in \mathcal{S}} \sum_{f \in \mathcal{F}} \sum_{t \in \mathcal{T}} (Ope_{s,f} * x_{s,f,t}) \tag{3.10}$$

subject to: **Aircraft count constraints:**

$$\sum_{ap \in \mathcal{AP}} y^{-}_{f,ap,MinT} + \sum_{(s,t) \in \mathcal{FL}} x_{s,f,t} \leq Avail_f, \quad \forall f \in \mathcal{F}, \tag{3.11}$$

Flow balance constraints:

$$y^{-}_{f,ap,MinT} = y^{+}_{f,ap,MaxT}, \quad \forall f \in \mathcal{F}, ap \in \mathcal{AP}, \tag{3.12}$$

$$y^{-}_{f,ap,(t+1)} = y^{+}_{f,ap,t}, \quad \forall (f, ap, t) \in \mathcal{N}, t \neq MaxT, \tag{3.13}$$

$$y^{-}_{f,ap,t} + \sum_{s \in \mathcal{S}_d(ap)} x_{s,f,D(s,t)} = y^{+}_{f,ap,t} + \sum_{s \in \mathcal{S}_o(ap)} x_{s,f,t},$$
$$\forall (f, ap, t) \in \mathcal{N}, \tag{3.14}$$

Demand and capacity constraints:

$$A^0_{m,pt} * \sigma_{i,pt,cl} \leq A_{i,pt,cl} * \sigma^0_{m,pt},$$
$$\forall i \in I_m, m \in \mathcal{M}, pt \in \mathcal{PT}, cl \in \mathcal{CL}, \tag{3.15}$$

$$\sum_{m \in \mathcal{M}} \sum_{pt \in \mathcal{PT}} \sum_{cl \in \mathcal{CL}} \sum_{i \in \tilde{I}_{m,s,t}} (Dem_{m,pt} * \sigma_{i,pt,cl}) \leq \sum_{f \in \mathcal{F}} Cap_f * x_{s,f,t},$$
$$\forall s \in \mathcal{S}, t \in \mathcal{T}, \tag{3.16}$$

$$\frac{\tilde{A}^0_{m,pt}}{A^0_{m,pt}} * \sigma^0_{m,pt} + \sum_{i \in I_m} \sum_{cl \in \mathcal{CL}} \left(\frac{\tilde{A}_{i,pt,cl}}{A_{i,pt,cl}} * \sigma_{i,pt,cl} \right) = 1,$$
$$\forall m \in \mathcal{M}, pt \in \mathcal{PT}, \tag{3.17}$$

Itinerary selection constraints:

$$\sum_{f \in \mathcal{F}} x_{s,f,t} \geq \sigma_{i,pt,cl},$$
$$\forall i \in \tilde{I}_{m,s,t}, m \in \mathcal{M}, s \in \mathcal{S}, t \in \mathcal{T}, pt \in \mathcal{PT}, cl \in \mathcal{CL}, \tag{3.18}$$

Restrictions on flight leg variables:

$$\sum_{f \in \mathcal{F}} x_{s,f,t} \leq 1, \quad \forall s \in \mathcal{S}, t \in \mathcal{T}, \tag{3.19}$$

$$\sum_{t \in \mathcal{T}} \sum_{f \in \mathcal{F}} x_{s,f,t} = Freq_s, \quad \forall s \in \mathcal{S}, \tag{3.20}$$

Variable value constraints:

$$x_{s,f,t} \in \{0, 1\}, \quad \forall s \in \mathcal{S}, f \in \mathcal{F}, t \in \mathcal{T}, \tag{3.21}$$

$$y^{-}_{f,ap,t}, y^{+}_{f,ap,t} \in Z^{+}, \quad \forall f \in \mathcal{F}, ap \in \mathcal{AP}, t \in \mathcal{T}, \tag{3.22}$$

$$\sigma^0_{m,pt} \geq 0, \quad \forall m \in \mathcal{M}, pt \in \mathcal{PT}, \tag{3.23}$$

$$\sigma_{i,pt,cl} \geq 0, \quad \forall i \in I_m, m \in \mathcal{M}, pt \in \mathcal{PT}, cl \in \mathcal{CL}. \tag{3.24}$$

Expression (3.10) formulates the model's objective function of maximizing the total operating profit of the airline, which is given by the total fare revenue (across all itineraries in all markets, all passenger types and all fare classes) minus total cost of operating all scheduled flights (across all segments, all fleet types and all departure time periods). The constraints are organized in six categories. First, Constraints (3.11) are the aircraft availability constraint which limits the number of assigned aircraft of each type to the available number in the airline's fleet. The second category includes flow balance constraints in the airline network. Because our model is a daily scheduling model, Constraints (3.12) require that the number of aircraft on the ground at any airport at the beginning of the day is equal to that at the end of the day. Constraints (3.13) and (3.14) ensure the flow conservation of aircraft across the network of flights. Constraints (3.15) to (3.17) correspond to the demand and capacity constraints, which belong to the third category of constraints. Constraints (3.15) define the market share of each itinerary-fare class combination to be proportional to its attractiveness. This constraint embeds a linearized version of our discrete-choice model of passengers' booking decisions, similarly to the approach from Wang et al. [142]. To see this, observe that it can be re-written as $\sigma_{i,pt,cl} \leq \frac{A_{i,pt,cl} * \sigma^0_{m,pt}}{A^0_{m,pt}}$. In the absence of aircraft capacity constraints, it would split the demand across all itineraries based on their respective attractiveness according to Equation (3.3). However, this is modified by Constraints (3.16), which ensure that the total number of passengers assigned to all itineraries which use a particular flight does not exceed the capacity of the aircraft type assigned to that flight. In other words, the inequalities in Constraints (3.15) let passengers shift across itineraries based on the attractiveness values and aircraft capacity restrictions. Constraints (3.17) then ensure that the market shares across all alternatives (including the itineraries from the other airlines and the no-fly alternative) sum up to 1. Fourth, Constraints (3.18) state that no passenger can be assigned to an itinerary if any of its flight legs is not operated. Fifth, Con-

straints (3.19) and (3.20) require each segment to be operated at most once per time period and the total number of flights on each segment to be equal to the daily frequency for that segment. Note that, in this formulation, daily flight frequency is given as an input to the model. We relax this assumption in Section 3.6, where we integrate frequency planning decisions into our modeling framework. Last, Constraints (3.21)-(3.24) define the domain of definition of the variables.

3.3 Solution Approach

The size of the mathematical model presented in Section 3.2 is extremely large for problem instances corresponding to the networks of any real-world airlines. As the timetable is discretized into 15 minute time periods, the time window between 6 am and 12 am (midnight) includes $18 \times 4 = 72$ time periods. Our largest case studies reported in this research comprise instances with 299 segments and 7 aircraft types resulting in $299 \times 7 \times 72 = 150,696$ binary variables, each corresponding to a combination of segment, fleet type and time period. As we shall see in our results, direct implementation of this model with commercial solvers does not provide good solutions (or sometimes any solutions at all) in reasonable times, even upon fine-tuning and carefully optimizing the solver parameters. In order to solve this large-scale problem, we have also implemented several well-known, state-of-the-art integer programming solution approaches. However, this did not yield any significant improvements either.

Thus, we propose, implement and demonstrate a new solution approach, as presented in Section 3.3.1, which decomposes the computations into multiple phases. Specifically, the first phase aggregates multiple consecutive time periods at each airport, and aims only to identify the aggregated time period of each flight. The second phase uses this solution as an input and identifies a narrower time period when each individual flight will be scheduled. The third phase is similar to the second, but narrows the time period even further, etc.

Additionally, we develop and integrate two accelerated rule-based heuristic approaches into this multi-phase framework based on variable-fixing and symmetry-inducing ideas. We refer to these heuristic approaches as a *Fleet-Fixing* Approach and a *Symmetry-Inducing* Approach and present them in Section 3.3.2.

3.3.1 Multiple Phase Solution Framework

This multi-phase framework is motivated by the fact that most airlines usually do not operate more than one flight in any given segment of their network within a short continuous period of time (with a length of, say, 60 minutes). We now describe a general version of the multi-phase framework with the number of phases equal to N. Let the width of the time window corresponding to phase i ($i \in 1, \ldots, N$) be W_i minutes and let $Solution_i$ be the solution generated at the end of phase i. In particular, we have $W_N = 15$ minutes. Our multi-phase solution approach to optimize the flight timetable step-by-step is presented as Algorithm 1.

Algorithm 1 Multi-phase solution algorithm

1: Initialization: $i = 1$.
2: Generate the input data to the MILP model (3.10)-(3.24) by setting the width of each time period equal to W_1.
3: Solve the MILP model (3.10)-(3.24), and store $Solution_1$.
4: **for** $i = 2, \ldots, N$ **do**
5: Generate input data to the MILP model (3.10)-(3.24) by setting the width of each time period equal to W_i.
6: Add constraints to ensure consistency with $Solution_{i-1}$ (see Equation (3.25)).
7: Solve the MILP model (3.10)-(3.24) with these additional constraints, and store $Solution_i$.
8: **end for**
9: Return $Solution_N$.

We computationally tested several different variants of the general approach described in Algorithm 1. The best performance was achieved when setting $N = 2$, $W_1 = 60$ minutes and $W_2 = 15$ minutes. In other words, we consider a two-phase approach. In Phase I we solve the model at the hourly level, while in Phase II we use this timetable obtained as Phase

I's solution to identify the exact 15 minute period for each flight. Let $\mathcal{U}$ be the set of 60-minute time periods, indexed by u, and let τ_u denote the 15-minute period corresponding to the top of the hour u (e.g., if u corresponds to the 8:00-9:00 am hour, then τ_u corresponds to the 8:00-8:15 am period). We denote the assignment variables of the Phase I problem by $x^{(I)}_{s,f,u}$ and those of the Phase II problem by $x^{(II)}_{s,f,t}$. In Phase II we *replace* Constraint (3.19) with Constraint (3.25), which ensures that, in any hour, a flight is scheduled in Phase II *if and only if* a flight is scheduled in that hour in Phase I.

$$\sum_{i=0}^{3} \sum_{f \in \mathcal{F}} x^{(II)}_{s,f,\tau_u+i} = \sum_{f \in \mathcal{F}} x^{(I)}_{s,f,u}, \quad \forall s \in \mathcal{S}, u \in \mathcal{U}. \tag{3.25}$$

3.3.2 Rule-Based Accelerated Heuristic Strategies

We now present two acceleration strategies that, when implemented in combination with the multi-phase solution framework, will be shown to considerably improve the computational performance of the solution algorithm.

Fleet-Fixing Approach

The central idea is to fix the fleet type assigned to each flight leg at the Phase I optimal value rather than re-optimizing it in Phase II, thus considerably reducing the number of binary decision variables in the Phase II model formulation. This allows the Phase II model to focus exclusively on timetabling decisions under passenger choice while holding fleet assignment decisions at their Phase I levels. While this does engender the possibility of sub-optimality arising from the neglected potential enhancements from Phase II re-fleeting decisions, it appears reasonable here given our particular emphasis on the timetabling decisions. Moreover, feasibility of the fleet assignment solution is already guaranteed given the corresponding constraints used when generating the Phase I solution. The results of our computational experiments presented in Section 3.4 validate the hypothesis that the considerable computational benefits of this heuristic approach more than offset the relatively

smaller sub-optimality risks. Mathematically, we replace Constraint (3.25) with Constraint (3.26):

$$\sum_{i=0}^{3} x^{(II)}_{s,f,\tau_u+i} = x^{(I)}_{s,f,u}, \quad \forall s \in \mathcal{S}, f \in \mathcal{F}, u \in \mathcal{U} \tag{3.26}$$

Symmetry-Inducing Approach

The central idea is to require the number of flights operated by any aircraft type on any segment to be equal to the number of flights operated by that same aircraft type on the "reverse segment", defined as the segment with the origin and destination equal to the destination and origin, respectively, of the original segment. This idea reduces the number of binary decision variables (i.e., the $x_{s,f,t}$ variables) by half. Similar to the Fleet-Fixing Approach, it also generates the risk of sub-optimality. However, real-world airline schedules are often highly symmetric, suggesting that the computational benefits might again outweigh the risk of sub-optimality. Our computational results, presented in Section 3.4, support this hypothesis. Note that, in addition to accelerating the solution procedure, the Symmetry-Inducing Approach has additional practical benefits because it simplifies airline operations. Mathematically, if $\mathcal{RS} \subset \mathcal{S} \times \mathcal{S}$ is the set of all unordered segment pairs such that the two segments in every pair are the reverse segments of each other, then we enforce Constraint (3.27):

$$\sum_{t \in \mathcal{T}} x_{s,f,t} = \sum_{t \in \mathcal{T}} x_{s',f,t}, \quad \forall f \in \mathcal{F}, (s, s') \in \mathcal{RS}. \tag{3.27}$$

Unlike the Fleet-Fixing Approach which only applies to Phase II, the Symmetry-Inducing Approach can be added to Phase I, Phase II, or both. This results in eight potential combinations: applying the Symmetry-Inducing Approach or not in Phase I, applying the Symmetry-Inducing Approach or not in Phase II, and applying the Fleet-Fixing Approach or not in Phase II. Seven of these eight combinations are listed in Table 3.1. The one involving the Symmetry-Inducing Approach in Phase II but not in Phase I as well as the

Fleet-Fixing Approach in Phase II has been left out because it will either result in infeasibility (if the Phase I decisions are not symmetric in terms of the assigned fleet types) or be identical to Heuristic 7.

Table 3.1: Abbreviated description of each of the seven feasible combinations of heuristics

	Heuristic 1	Heuristic 2	Heuristic 3	Heuristic 4	Heuristic 5	Heuristic 6	Heuristic 7
Phase I				Symmetry	Symmetry	Symmetry	Symmetry
Phase II		Fixing	Symmetry	Symmetry	Fixing		Fixing, Symmetry

Note that, if each optimization problem was solved to optimality, Heuristics 5 and 7 should yield identical optimal solutions because Phase II symmetry is guaranteed by the combination of the Phase I Symmetry-Inducing Approach and the Phase II Fleet-Fixing Approach. However, explicit application of the Symmetry-Inducing Approach in Phase II has an impact on the run-times and memory requirements, and hence on the actual best solutions obtained within reasonable run-times. Therefore, we retain both Heuristic 5 and Heuristic 7 in Table 3.1, rather than eliminating one of them. Next, we test and compare the performances of all seven heuristics.

3.4 Computational Results

We implement the model and the solution approaches to several problem instances based on the network of Alaska Airlines. Section 3.4.1 presents the computational setup used in the remainder of this research (unless otherwise specified). The first goal of the experiments reported in this section is to compare the performances, within a given computational run-time budget, of our solution framework and the various heuristic combinations listed in Table 3.1 with the performance of a commercial MILP solver. This is presented in Section 3.4.2. The second goal is to analyze how model's solutions change over longer computational time horizons. This is shown in Section 3.4.3.

3.4.1 Experimental Setup

Our computational test instances are based on the network of Alaska Airlines. In our datasets, obtained for year 2016, Alaska Airlines was a moderate-sized hub-and-spoke airline carrier in the United States. In 2016, Alaska Airlines operated a mixed fleet consisting primarily of the Boeing 737 family aircraft. It operated its largest hub at Seattle, WA and two secondary hubs in Anchorage, AK and Portland, OR. Alaska Airlines underwent a merger with Virgin America in 2017. As a result, there are interesting opportunities to analyze the effects of the merger on the optimal timetables and fleet assignment solutions of the combined carrier using the model and algorithms developed in this research. We perform this analysis in Section 3.6.3.

We design a series of test instances increasing in size. We refer to these instances as Network 1 to Network 5—containing 5 to 59 airports. Table 3.2 reports the number of airports and flights in each network and the corresponding sizes of the optimization model (Equations (3.10)-(3.24)).

Table 3.2: Descriptions of the five networks under consideration, and respective problem sizes.

			# variables			
Network	# airports	# flights	continuous	binary	integer	# constraints
Network 1	5	92	66,625	9,000	3,600	156,725
Network 2	7	126	130,585	17,640	5,040	290,015
Network 3	14	210	450,385	60,840	9,360	932,820
Network 4	17	232	1,066,000	144,000	14,400	2,140,005
Network 5	59	390	39,665,860	5,358,240	87,840	74,438,915

We now describe how the model's input parameters can be obtained or approximated from available public data sources and from the existing literature. First, the number of daily flights in each nonstop segment, denoted as $Freq_s$ in our model, is computed as the average number of flights per day operated in January 2016, obtained from the Airline On-Time Performance (AOTP) database [28]. The number of available aircraft of each fleet

type can be obtained from the carrier's website. However, the publicly available numbers include aircraft used for international flights (which are not included in our model) and aircraft undergoing repair and maintenance; as such, they do not necessarily coincide with the actual numbers of aircraft used for domestic flights. Therefore, we obtain the number of aircraft of each fleet type used for domestic flights by Alaska Airlines in January 2016 by assigning each tail number to a fleet type, and computing the number of aircraft operating daily for each fleet type. We set the earliest departure time $(MinT)$ at 6 am and the latest departure time $(MaxT)$ at 12 am, since over 98% of Alaska Airlines' flights are scheduled to depart between 6 am and 12 am. We only consider nonstop and one-stop itineraries in this research, which together account for 97.5% of the one-way air passenger trips in the United States [20]. Minimum aircraft turnaround times are all assumed to be 45 minutes. To construct connecting itineraries, we consider passenger connection times within the range of 45 to 180 minutes. We do not consider interline connections (that is, a connection from a first-leg flight flown by the host carrier and a second-leg flight flown by another airline, or vice versa). In practice, these assumptions can be easily relaxed depending on the practical requirements. For instance, minimum aircraft and passenger connection times can be varied as a function of the airport and the time of day; in addition, circuity can be used to reduce the number of feasible itineraries that are generated.

We use the estimates of flight operating costs from Swan and Adler [131] to calibrate the parameter $Ope_{s,f}$ as a function of segment distance (D_s) and aircraft capacity (Cap_f). Equations (3.28) and (3.29) describe this relationship for short-haul flights (defined as those on segments of less than 3,106 Miles, or 5,000 Kilometers in length) and long-haul flights (defined as those on all other segments), respectively.

$$Opt_{s,f} = (1.6D_s + 722) * (Cap_f + 104) * \$0.0190 \quad \text{if } D_s \leq 3,106 \text{ miles} \tag{3.28}$$

$$Opt_{s,f} = (1.6D_s + 2200) * (Cap_f + 211) * \$0.0115 \quad \text{if } D_s > 3,106 \text{ miles} \tag{3.29}$$

The airfares used as inputs to our model are computed, using the Airline Origin and Destination Survey (DB1B) data from the Bureau of Transportation Statistics [29], as the average ticket price values for each combination of year, quarter, origin airport, destination airport, connection airport (if any), first leg carrier, and second leg carrier (if any). Note that we use quarterly average values because ticket price information at a finer granularity is not available publicly. However, the airlines (the intended users of our methodology) do have access to such information for their own flights, and also typically for the flights of their competitors (at least in an approximate manner). Our methodology is able to accommodate such departure-time-dependent price values whenever such data is available.

We obtain passenger demand data on each O-D market from the DB1B database, which we use to define our parameters $Dem_{m,pt}$. We calibrate our GAM model of passenger choice (Equations (3.4) to (3.9)) as follows. For simplicity, we first assume only one passenger type and one fare-class. This means that the set PT of passenger types and the set CL of fare clases are both singletons. Extensions involving multiple passenger types and multiple fare classes are presented in Section (3.6.1). For each itinerary (from the host airline as well as other airlines), we compute the corresponding utility values using the empirical specification from Lurkin [91]—as mentioned earlier, the corresponding attributes are: total trip time, number of connections, departure time of the day, ticket price, distance of the itinerary, direction of travel, number of time zones crossed and departure day of the week. Given the lack of available data and research studies on the attractiveness of the no-fly alternative, we ignore this term in our GAM specification. Note that this assumption is reasonable since our demand estimates correspond to the actual number of flying passengers, and does not include latent not-flown demand. We have conducted a detailed numerical investigation into the effects of this assumption, and concluded that this choice does not change our conclusions to any significant extent. Finally, in the absence of the availability of accurate empirical estimates, we set all the shadow attractiveness values to 0. This, once again is a limitation arising from the fact that we don't have access to pro-

prietary passenger booking data (though the airlines do have access to it), which will allow estimating the shadow attractiveness values using the maximum likelihood estimation and least squares methods as mentioned in Section (3.2.1) [59].

All experiments reported in this chapter are performed on a server with 128GB of RAM and Linux Operating System. We implement the algorithms using Java along with the latest version of the CPLEX MILP solver - CPLEX 12.7 - with default parameter settings.

3.4.2 Comparison of Heuristics

In this section, we evaluate and compare the computational performance of the multi-phase solution approach and the different heuristics presented in Section 3.3. Results are reported in Table 3.3. For each of the five networks, we consider two baselines obtained by implementing the model directly in CPLEX without any of the solution approaches or heuristic ideas introduced in Section 3.3. First, "CPLEX Short" corresponds to the solution obtained after a relatively short run-time (fixed at two hours here). Second, "CPLEX Long" corresponds to the solution obtained after a longer run-time of 48 hours. This essentially provides the best solution that can be obtained directly with a commercial solver. For each of our seven heuristics (referred to as "H 1" to "H 7"), the run-time limit is set at one hour for Phase I plus one hour for Phase II, resulting in a two-hour total run-time limit which makes it comparable with the "CPLEX Short" results. In Table 3.3, we use the "CPLEX Long" solution as the baseline against which all other results are evaluated. The results indicate the percentage improvement in operating profit for each approach when compared to the "CPLEX Long" baseline. In other words, for each solution approach and each test instance, the table reports the *% gap* defined as $\frac{O-O^*}{O^*}$, where O is the objective function value obtained by that particular solution approach and O^* is the corresponding objective function value obtained by directly running CPLEX for 48 hours. Thus, any positive results in Table 3.3 indicate that the particular solution approach results in a larger operating profit value than can be obtained directly with CPLEX even after running it for a much longer

time.

Table 3.3: Relative performance of each solution approach, as compared to the "CPLEX Long" solution

Problem	**CPLEX**		**Our Heuristics**						
	Short	Long	**H 1**	**H 2**	**H 3**	**H 4**	**H 5**	**H 6**	**H 7**
Network 1	-3.58%	0%	0.20%	0.59%	-5.87%	-0.98%	0.54%	-6.32%	0.28%
Network 2	-12.21%	0%	2.05%	2.81%	2.32%	2.51%	3.25%	2.16%	1.94%
Network 3	-11.03%	0%	-0.43%	-1.74%	-0.96%	-0.93%	-1.82%	-1.19%	-1.82%
Network 4	-16.98%	0%	-5.88%	-3.51%	-3.14%	-2.80%	-3.51%	-3.07%	-3.50%
Network 5	No solution	0%	**25.72%**	**21.08%**	**25.08%**	**42.82%**	**40.59%**	**42.11%**	**40.39%**

The main takeaway from Table 3.3 is that the combination of our multi-phase solution approach and the acceleration heuristics provide high-quality and scalable solutions, as compared to a direct CPLEX implementation. For the smallest instance (Network 1), four of our seven heuristics (Heuristics 1, 2, 5 and 7) outperform a direct CPLEX implementation, even when our approaches are run for only two hours while CPLEX is allowed to run for 48 hours. Additionally, Heuristic 4 also outperforms the CPLEX solution with comparable run-times (i.e., two hours). The improvements relative to a direct CPLEX implementation are far more dramatic under Networks 2, 3, 4 and 5, which are larger in size. For these four larger instances, each of our seven heuristics performs substantially better than running CPLEX directly for the same amount of run-time (i.e. for two hours). Typically, the larger the size of the network, the bigger and more obvious is the improvement (barring the better-than-expected performance of our heuristics on Network 2). For the entire domestic network of Alaska Airlines (i.e. Network 5), CPLEX does not generate even a feasible solution within two hours. In contrast, all seven of our heuristics produce an improvement of 20% to 40% within a two hour run-time when compared with the CPLEX implementation run for 48 hours. In summary, all seven of our heuristics outperform a direct CPLEX implementation within a much lower run-time budget for the largest network.

Next, let us compare the relative performance of the seven heuristics with each other, with a particular focus on Network 5. Our heuristics that apply the Symmetry-Inducing

Approach in Phase I (Heuristics 4 to 7) (**42.82%**, **40.59%**, **42.11%**, **40.39%**) significantly outperform those that do not (Heuristics 1 to 3) (**25.72%**, **21.08%**, **25.08%**). This suggests that the Symmetry-Inducing Approach can greatly improve the solution quality in Phase I and thus lead to a superior final solution. This result further validates the effectiveness of the symmetry-inducing heuristic.

3.4.3 Evolution of Objective Function Values With Increasing Run-Times

When we implement our model directly in CPLEX for Network 5, we observe that the best available solution improves very slowly. Specifically, an initial feasible solution is found after little more than two hours, but further improvement is then very small (less than 1%) even after running the solver for up to 15 hours. In contrast, our heuristics improve the solution quality much sooner as we increase their run-time beyond two hours. Thus, it is insightful to plot side-by-side the performance of our heuristics in terms of the evolution of the objective function values of the best available solutions against their run-times. The results shown here are based on the entire domestic network of Alaska Airlines (Network 5) as the test instance and Heuristic 4 as the solution approach, which was the best solution approach based on the results from Table 3.3.

Figure 3.2 presents two horizontal lines and four curves. The black horizontal line corresponds to the objective function value obtained by the direct CPLEX implementation after it is run for 48 hours. The green horizontal line corresponds to the objective function value obtained by running our Heuristic 4 for 24 hours in Phase I and then running its Phase II until it runs out of memory (i.e., for 1.5 hour in this case). Each curve corresponds to a different Phase I run-time (namely, half hour, 1 hour, 2 hours, and 4 hours) for our Heuristic 4. The x-axis corresponds to the total run-time of the overall heuristic (combining the Phase I and Phase II run-times) and the y-axis corresponds to the objective function value corresponding to the best available solution. For each curve, we plot the objective function

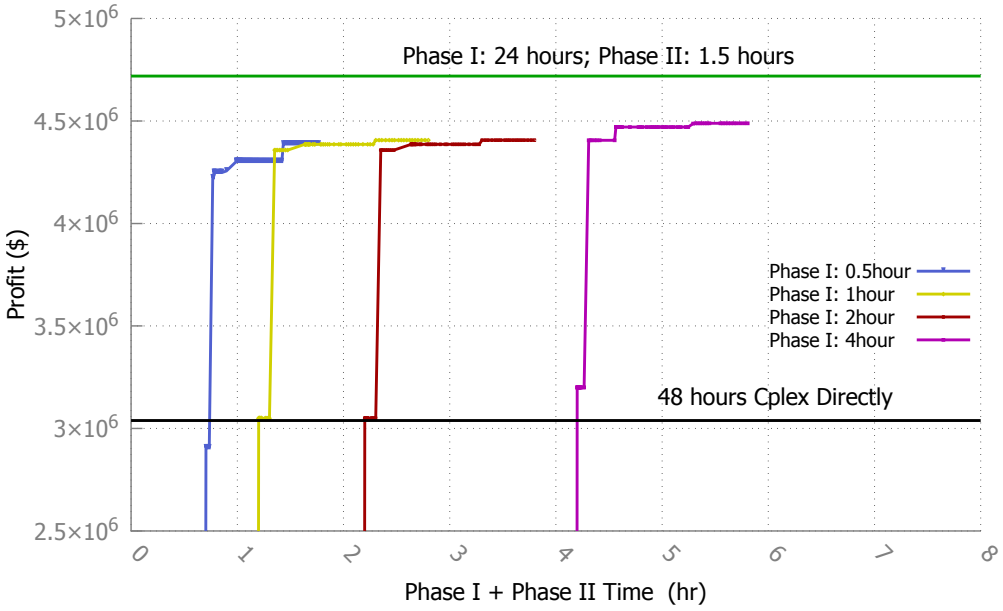


Figure 3.2: Evolution of the objective function value versus run-time (obtained with Heuristic 4 on Network 5)

value until the computer runs out of memory.

First, the left most curve curve demonstrates that our heuristic generates a substantially better solution, in less than one hour of run-time, compared to the solution obtained by a direct CPLEX implementation with a 48 hours run-time. In most cases, as the run-time increases, the solution quality increases and the improvements happen in jumps as and when a better solution than the previous one is obtained by the solver. However, in most cases, the solution quality improvements taper off after a certain point in time, which is an indicator of the quality of the corresponding Phase I solution. We observe that each of the four curves ends because the Phase II model eventually runs out of memory approximately two hours after Phase II begins. In addition, a longer Phase I run-time leads to better quality of the eventual Phase II solution. This makes sense because a longer Phase I run-time allows for a better chance to find a superior solution, which serves as a better starting point for the Phase II model. Note that this relationship is non-linear though. For example, the curves corresponding to one hour and two hour Phase I run-times have the same ultimate

objective function value, which means that the additional hour spent in Phase I does not help in finding a better eventual solution. In contrast, increasing the Phase I run-time from two hours to four hours results in a significant improvement in the ultimate profit value. In fact, the solution obtained after a four-hour run-time in Phase I is within 5% of the one obtained with a 24-hour run-time in Phase I, which indicates the quality of our heuristics in terms of generating high-quality solutions in reasonable computational times.

These results provide valuable insights into the quality of our solution approach. Indeed, even with a much shorter run-time (two hours), our solution approach generates a solution that yields a profit improvement of over 40% when compared with the solution obtained by a direct CPLEX implementation that runs for 48 hours. One of the main contributors to these high-quality solutions is the application of the Symmetry-Inducing Approach in Phase I. These results underscore the benefits that can be derived from the model developed in this research in support of airline timetabling. In addition, they also highlight the trade-off between the run-time and the quality of the generated solution. In particular, a noteworthy aspect of Figure 3.2 is that each of the four curves experiences a sudden and significant jump within the first half hour of Phase II run-time. In general, timetabling problems are strategic in nature and hence allow for large budgets of computational times. However, these jumps indicate that, in the event that the run-time budget is somewhat limited, even a half hour Phase II run-time can yield a good solution. This ability to quickly generate a high-quality timetable can be highly valuable to airlines even in the strategic context of flight timetabling, as it enables repeated runs of the model solution process to conduct sensitivity analyses and scenario analyses as part of the long-term decision-making processes.

3.5 Benefits of Timetabling

The mathematical modeling and computational framework for comprehensive timetable development that integrates fleet assignment decisions and passengers' booking decisions developed in this research differs from existing approaches in two major ways. First, existing schedule design models consider a feasible flight timetable as a starting point, and perform incremental changes to it to determine the new timetable. Second, the airline planning process is typically conducted in a sequential rather than an integrated manner. In other words, (incremental) timetable development takes place first, and then, fleet assignment decisions are optimized using the previously determined timetable as an input. In this section, we compare our integrated approach to these various existing approaches that have been presented in the previous literature and/or used in practice.

3.5.1 Experimental Setup

We aim to evaluate the benefits of our modeling and computational framework by comparing it to a number of baselines. As the first baseline, we use our passenger choice model to evaluate the operating profits corresponding to the actual timetable and the actual fleet assignment solution used by the airline. This implies fixing all timetabling and fleet assignment decision variables at their real-world values, and replicating passengers' booking decisions for these specific flight offerings. We designate this as *Baseline 0*.

Next, we consider the various incremental models replicating the approaches presented in the literature. We design three baselines (*Baseline 1*, *Baseline 2* and *Baseline 3*) corresponding to different levels of incremental changes in the airline planning process. In *Baseline 1*, we only optimize the fleet assignment decisions (i.e., assigning aircraft types to all flight legs to maximize the total operating profits) while holding the timetabling decisions fixed at their real-world values. This aims to replicate many past studies (e.g., Barnhart et al. [18]), which have focussed on the fleet assignment problem with passenger

spill and recapture effects. By comparing the operating profits corresponding to *Baseline 1* with those corresponding to *Baseline 0*, we evaluate the benefits of optimal fleet assignment decisions alone, while incorporating passenger choice.

Then, the focus of *Baseline 2* is to measure the combined effects of optimizing fleet assignment decisions and incremental changes in flight timings. This is motivated by the results from Sherali et al. [120], which suggest that incremental changes to flight departure times can create more connection opportunities for passengers. Specifically, we develop two baselines, referred to as *Baseline 2-a* and *Baseline 2-b*, both of which combine the fleet assignment problem with the problem of flight re-timing within designated time windows. Time windows in *Baseline 2-a* are 30 minutes in width (i.e., the departure time of each flight is allowed to be modified by at most 15 minutes in either direction), while those in *Baseline 2-b* are 60 minutes in width (i.e., the departure time of each flight is allowed to be modified by at most 30 minutes in either direction).

Next, in addition to optimizing fleet assignment and flight re-timing decisions within designated time windows, *Baseline 3* also allows elimination of any subset of the flights designated as optional. This follows multiple previous research studies (e.g., Lohatepanont and Barnhart [89], Sherali et al. [119]) that designate flights as either mandatory or optional, thus allowing the optimization model to consider eliminating some of the flights designated as optional. To the best of our knowledge, none of the existing studies in the literature provides any guidance regarding how to designate any given flight as mandatory or optional. For our experiments under *Baseline 3* we designate all Hub-to-Hub flights as optional. But in order to maintain consistency across our results, we ensure that the total daily frequency in each nonstop segment (including the Hub-to-Hub segments) remains constant. Specifically, on each Hub-to-Hub segment, we generate one additional flight at the "midpoint" of any pair of two consecutive flights, and label all Hub-to-Hub flights as optional. For instance, if on a given segment an airline has two flights scheduled to depart at 2 pm and 6 pm, respectively, then we include in our model a total of five optional flights

on that segment with proposed scheduled departure times of 10 am (the midpoint of 6 am and 2 pm), 2 pm (the other departure time currently offerred), 4 pm (the midpoint of 2 pm and 6 pm), 6 pm (one of the currently offerred departure times) and 9 pm (the midpoint of 6 pm and 12 am). Ultimately, the ratio of the total number of optional flights to the number of optional flights to be selected on each Hub-to-Hub segment is slightly higher than 2:1. Similar to *Baseline 2*, we also divide *Baseline 3* into two different experiments: *Baseline 3-a* has a 30-minute wide time window for each flight (e.g., the aforementioned five flights can be scheduled any time during 9:45-10:15 am, 1:45-2:15 pm, 3:45-4:15 pm, 5:45-6:15 pm, and 8:45-9:15 pm, respectively), while *Baseline 3-b* has a 60-minute wide time window for each flight (e.g., the aforementioned five flights can be scheduled any time during 9:30-10:30 am, 1:30-2:30 pm, 3:30-4:30 pm, 5:30-6:30 pm, and 8:30-9:30 pm, respectively).

We solve each of these six baseline cases (*Baselines 0, 1, 2-a, 2-b, 3-a, 3-b*) by implementing them directly in CPLEX. We compare the results of these with the outputs of our heuristics for solving our model formulation given by (3.11)-(3.24). Additionally, we also compare them with the solution obtained by running for 48 hours a direct CPLEX implementation of our model formulation given by (3.11)-(3.24) (referred to as *CPLEX Directly*). In order to ensure a fair comparison of the results, we run each model for a *total* run-time of 48 hours. Our heuristics are allowed to run for a maximum of 24 hours of Phase I run-time and a maximum of 24 hours of Phase II run-time (or until it runs out of memory). *Baseline 0* and *Baseline 1* are successfully solved to their respective optimal solutions within the 48 hours of their allocated run-time. But provably optimal solution for *Baselines 2-a, 2-b, 3-a, 3-b* was not obtained within 48 hours of run-time. In addition to this so-called *cold start* solution approach, we also experimented by using a *warm start* approach. The general idea of a warm start is to help CPLEX by providing an existing (hopefully good) feasible solution as a starting point. Specifically, in an attempt to generate a better solution for *Baseline 2-a*, we initialize it with the optimal solution from *Baseline 1*. Similarly, we

warm-start *Baseline 2-b* and *Baseline 3-a* both with the best available solution from *Baseline 2-a*, we warm-start *Baseline 3-b* with the best available solution from *Baseline 2-b*, and we warm-start *CPLEX Directly* with the best available solution from *Baseline 3-b*. Each of these warm starts leverages provably feasible solutions obtained from previous problem implementations. No warm start strategy is used when implementing the heuristic solution approaches developed in this research.

Note that there is no guarantee that any warm start approach produces a better solution than a cold start approach. In fact, for both *Baseline 2-a* and *Baseline 2-b*, we found that the warm-start solutions are inferior to their corresponding cold start solutions, while in all other comparisons, warm start solutions were found to be superior to the cold start solutions. For example, the cold start solution of *Baseline 3-b* is even worse than the best available solution for *Baseline 2-b*, but its warm start *Baseline 3-b* is considerably better. When it comes to the *CPLEX Directly* approach, a cold start implementation of our model does not even yield a feasible solution within 48 hours. In contrast, warm start at least guarantees that its solution is not inferior to the best available solution of *Baseline 3-b*. In the remainder of this section, we report, for all baselines and for the *CPLEX Directly* approach, the results obtained by using the warm start or the cold start, whichever produces a superior solution.

3.5.2 Benefits of our Modeling and Computational Framework

Table 3.4 reports the profit generated by each of the baseline approaches, the *CPLEX Directly* approach, and two of our heuristics. These results correspond to the test instance corresponding to the full domestic network of Alaska Airlines (Network 5). For simplicity and to save some space, we only report the results obtained with the heuristic that performs the best (Heuristic 5, here) and the one that performs the worst (Heuristic 6, here) among Heuristics 4-7, which were shown in Table 3.3 to be the better-performing ones. Note that the relative performance of the heuristics is different from the one elicited in Table 3.3, be-

cause Table 3.3 reported results obtained with a one-hour run-time in each phase, while we allow here for a run-time of up to 24 hours in each phase. Moreover, the worst-performing heuristic among Heuristics 4-7 (Heuristic 6) is the only one that does not ensure symmetry in the Phase II solution (see Table 3.1) either through explicit induction of symmetry or through fixing of fleet types in Phase II or both. This further emphasizes the benefits of our Symmetry-Inducing Approach, not only in Phase I but also in Phase II. For each of these tests, Table 3.4 presents a variety of summary statistics characterizing operating profits, revenues and costs. The numbers in parentheses provide changes in comparisons to *Baseline 0*. They are calculated as $\frac{z^* - z^0}{z^0}$, where z^* is the value of the relevant statistic for the solution under consideration and z^0 is the corresponding value for the solution obtained with *Baseline 0*.

Table 3.4: Benefits of our approach compared to various baselines

	Profit ($)	**Revenue**			**Operating Cost**	**Number of Passengers**	
		Total ($)	Nonstop ($)	One-stop ($)		Nonstop	One-stop
Baseline 0	3,125,862 (0%)	9,085,984 (0%)	7,453,217 (0%)	1,632,766 (0%)	5,960,139 (0%)	33,637 (0%)	7,288 (0%)
Baseline 1	3,451,808 (+10.46%)	9,267,774 (+2.00%)	7,600,360 (+1.97%)	1,667,414 (+2.12%)	5,816,174 (-2.42%)	34,259 (+1.85%)	7,444 (+2.14%)
Baseline 2-a	3,823,542 (+22.36%)	9,664,902 (+6.37%)	7,547,487 (+1.26%)	2,155,916 (+32.04%)	5,879,861 (-1.35%)	34,134 (+1.48%)	9,076 (+24.53%)
Baseline 2-b	4,085,654 (+30.75%)	9,981,443 (+9.86%)	7,396,392 (-0.76%)	2,585,051 (+58.32%)	5,895,790 (-1.08%)	33,450 (-0.56%)	10,305 (+41.40%)
Baseline 3-a	3,913,461 (+25.24%)	9,788,934 (+7.74%)	7,512,856 (+0.80%)	2,276,078 (+39.40%)	5,875,979 (-1.41%)	33,890 (+0.75%)	9,546 (+30.98%)
Baseline 3-b	4,138,876 (+32.45%)	10,038,133 (+10.48%)	7,426,230 (-0.36%)	2,612,097 (+59.98%)	5,899,452 (-1.02%)	33,576 (-0.18%)	10,539 (+44.61%)
CPLEX Directly	4,138,876 (+32.45%)	10,038,133 (+10.48%)	7,426,230 (-0.36%)	2,612,097 (+59.98%)	5,899,452 (-1.02%)	33,576 (-0.18%)	10,539 (+44.61%)
Heuristic 6 (worst)	4,312,706 (+38.01%)	10,259,745 (+12.92%)	7,431,617 (-0.29%)	2,828,218 (+73.22%)	5,947,039 (-0.21%)	33,696 (+0.17%)	10,916 (+39.78%)
Heuristic 5 (best)	4,903,946 (+56.93%)	10,838,131 (+19.28%)	7,505,688 (+0.70%)	3,332,443 (+104.10%)	5,934,645 (-0.43%)	34,165 (+1.57%)	12,484 (+71.30%)

As expected, the profit values are found to increase when transitioning from each baseline to the next, as the solution space is progressively expanded to allow for additional flexibility in the flight timetabling and fleet assignment processes. Then, note that when we

use *Baseline 3-b*'s best available solution to initialize the direct CPLEX implementation of our model formulation, that solution still remains the best available solution even after running CPLEX for 48 hours. In contrast, the implementation of *any* of our four heuristics (even the one that performs the worst) results in a larger profit value than the best available baseline. Further profit improvements can be achieved through the implementation of the heuristics that perform better. In summary, the rank ordering of the various solutions listed in Table 3.4 based on their operating profits is *Baseline* 0 $<$ *Baseline* 1 $<$ *Baseline 2-a* $<$ *Baseline 3-a* $<$ *Baseline 2-b* $<$ *Baseline 3-b* $=$ *CPLEX Directly* $<$ *Worst heuristic* $<$ *Best heuristic.* These results suggest that the *combination* of our integrated comprehensive timetabling and fleet assignment formulation and our solution approaches can provide significant profit improvements, as compared to *all* existing approaches. We now compare the solutions obtained under each modeling and solution approach in more detail.

First, compared to *Baseline 0*, *Baseline 1* allows capturing more passengers and hence more revenue in certain markets by allocating larger aircraft. The total number of passengers being carried increases from 40,925 to 41,703 (a 1.90% increase) leading to an increase in total revenue from \$9,085,984 to \$9,267,774 (a 2% increase). Additionally, it also assigns smaller aircraft to certain other flights with empty seats leading to a reduction in the operating cost from \$5,960,139 to \$5,816,174 (a 2.42% decrease). Thus, through fleet assignment optimization in *Baseline 1*, operating profit increases from \$3,125,862 to \$3,451,808 (a 10% increase).

Compared to *Baseline 1*, *Baseline 2-a* allows adjusting flight departure times within time windows of ± 15 minutes. While the number of nonstop passengers decreases slightly (by 125 from 34,259 to 34,134) when compared to *Baseline 1*, the number of one-stop passengers increases significantly from 7,444 to 9,076 leading to a \$488,502 increase in one-stop revenue. This is further amplified in *Baseline 2-b* which allows adjusting flight departure times within time windows of ± 30 minutes. This results in the number of one-stop passengers increasing to \$10,305 but in the number of nonstop passengers decreasing

to $33,450. The one-stop revenue improvement in *Baseline 2-b* leads to a further profit increase. These effects stem from small departure time adjustments that result in slightly less attractive scheduled times for nonstop passengers, but generate more connection opportunities for one-stop passengers.

Going from *Baseline 2-a* to *Baseline 3-a*, and from *Baseline 2-b* to *Baseline 3-b*, the optional flights on Hub-to-Hub segments afford additional timetabling flexibility. These result in an increase in the number of one-stop passengers (by 470 and 234 respectively), and a corresponding increase in one-stop revenues. As a result, the total profit increases by $89,919 and $53,222 respectively. The difference is likely because the ± 30 minutes time windows in case of *Baseline 2-b* already provide significant scheduling flexibility, so the additional gains by allowing optional flights in *Baseline 3-b* are more modest than those obtained by moving from *Baseline 2-a* to *Baseline 3-a*.

The last three rows list the results generated by a direct CPLEX implementation and our heuristics for solving our model formulation. As noted earlier, the direct CPLEX implementation of our model formulation does *not* improve on the solution provided by *Baseline 3-b* even after 48 hours. In contrast, the results obtained with *any* our heuristics yield significant profit improvements over all baseline solutions. Particularly noteworthy is the fact that our heuristics presented in Table 3.4 have nonstop revenues that are very similar to (or sometimes even lower than) the various baselines being considered. Moreover, the operating costs of the solutions from our heuristics are actually slightly higher than all baselines except for *Baseline 0*. However, the solutions given by our heuristics have significantly greater one-stop revenues compared to all the baselines. This highlights the fact that most of the benefits from our approaches are derived from an increase in the availability of attractive one-stop itineraries. All the baseline approaches ranging from *Baseline 0* to *Baseline 3-b*, either do no allow timetabling changes or allow only for certain marginal changes. This limits the airline's ability to provide flight offerings that enable passengers to choose their most desired one-stop itineraries. In contrast, our model's flexibility due

to its comprehensive approach to timetabling, and its explicit capture of passenger choice decisions allows it to increase the number of one-stop passengers, and consequently the total one-stop revenue and the total operating profit, dramatically. In addition to increased profitability, the additional one-stop passenger capture also enhances the airline's market share significantly —an added advantage of our approach. In conclusion, the *combination* of our formulation of the comprehensive integrated timetabling and fleet assignment optimization model with passenger choice with our heuristic solution approaches produces the solutions with, by far, the highest operating profits and market shares.

3.5.3 Comparison with Actual Timetable

We conclude this section with a comparison of the solution generated by our modeling and computational framework to the actual timetable produced by Alaska Airlines in 2016. Unfortunately, a full apples-to-apples comparison is not possible, since our modeling approach necessarily omits a number of practical considerations that play a role in the development of the airline's actual timetable. We thus only report aggregate metrics in Table 3.5; nonetheless, the comparison sheds light on the main differences between the model's solution and the actual solution.

The observations from Table 3.5 are threefold. First, on the supply side, the model reallocates some of the aircraft seats from Hub-Hub segments to Spoke-Hub segments. Interestingly, the number of nonstop passengers on Spoke-Hub segments is actually slightly lower in our solution; this suggests that the added seats on Spoke-Hub segments aim to increase connectivity into the host airline's Hub airports. Obviously, the solutions also differ in terms of timetabling, which is not shown here. Second, the model results in a slight increase in the total number of nonstop passengers—by an estimated 1.57%. This is mainly driven by an 8.2% increase in the number of nonstop passengers on Spoke-Spoke segments. This is primarily achieved by aligning flight timetables with the (estimated) profiles of passenger demand. Third, the main difference between the two timetables lies in

Table 3.5: Comparison of model solution to actual timetable

Metric	Segment	Actual	Model
Number of seats	Hub-Hub	8,539	8,129
	Hub-Spoke	38,435	38,435
	Spoke-Hub	38,373	38,783
	Spoke-Spoke	9,757	9,757
	Total	95,104	95,104
Number of nonstop passengers	Hub-Hub	661	665
	Hub-Spoke	14,408	14,615
	Spoke-Hub	14,605	14,597
	Spoke-Spoke	3,963	4,288
	Total	33,637	34,165
Number of connection opportunities	Hub-Hub-Hub	8	12
	Spoke-Hub-Spoke	1,707	2,714
	Hub-Hub-Spoke	147	286
	Spoke-Hub-Hub	252	275
	Total	2,114	3,287
Number of connecting passengers	Hub-Hub-Hub	153	194
	Spoke-Hub-Spoke	4,754	9,185
	Hub-Hub-Spoke	1,304	1,604
	Spoke-Hub-Hub	1,077	1,501
	Total	7,288	12,484

the dramatic increase in the number of connection opportunities, and the resulting number of connecting passengers. Interestingly, the percentage increase in the number of connecting passengers is even larger than the percentage increase in the number of connection opportunities, suggesting positive ripple effects of network connectivity.

These findings highlight the main drivers of network-wide timetable optimization, namely (i) the alignment of each flight's timetable with the patterns of nonstop passenger demand, and (ii) the coordination of flight timetables across multiple segments to enhance connection opportunities and cater to higher numbers of connecting passengers.

3.6 Extensions

We now propose a number of modeling, computational and practical extensions of our approach, using the full domestic network of Alaska Airlines. This aims to provide additional

insights into the benefits of the model developed in this research and to characterize the optimal scheduling strategies for major airlines. First, in Section 3.6.1, we perform additional computational experiments with multiple passenger types (e.g., business and leisure) and multiple fare classes. While all results presented so far in this chapter have been based on instances involving a single passenger type and a single fare class, the optimization formulation (Equations (3.10)-(3.24)) and the solution heuristics make no such assumption. Therefore, we can use our modeling and computational framework to capture the fact that airlines offer multiple fare classes and cater to passengers with different sensitivities to the itinerary attributes such as price, departure time, etc. Second, in Section 3.6.2, we extend our modeling and computational framework to integrate frequency planning decisions into our timetable development and fleet assignment framework. This is motivated by the significant impact of frequency planning decisions on an airline's profit and the strong interdependencies of frequency planning decisions with timetable development, fleet assignment and passenger choice. Last, the recent merger between Alaska Airlines and Virgin America is expected to have an impact on various aspects of the new merged airline's network, schedules and operations. We use our framework to analyze the effects of this merger on their optimal flight schedules in Section 3.6.3.

Throughout this section, we compare the results using our modeling and computational framework in different test instances. Unlike the previous sections, the focus is less on the comparison of our various heuristics with each other and with various baselines. Instead, we tested all heuristics in Section 3.6.1, and found that Heuristic 7 is best one among all seven heuristics (i.e., the one that leads to the highest profit value) when tested with a runtime of 24 hours in Phase I and 24 hours in Phase II (or until it runs out of memory). Thus, we choose Heuristic 7 as the solution approach in this section to analyze these extensions.

3.6.1 Effects of Multiple Passenger Types and Multiple Fare Classes

So far, as a simplification, our computational experiments have only considered one passenger type and one fare class. In reality, passengers differ in terms of the relative value that they place on different itinerary attributes. For example, business travelers typically place a higher emphasis on schedule convenience and flexibility, while those flying for leisure purposes are often more price sensitive. Airlines, in turn, are also known to offer various fare classes, through different marketing and sales channels, during various time periods prior to the scheduled flight departure time, in an attempt to price different passengers differently. Detailed modeling of airline pricing and revenue management strategies is considerably beyond the scope of this research, not only because of the prohibitive mathematical modeling burden of tackling such analysis but also because of the lack of any public source of the relevant pricing and revenue management data. Instead, in this section, we test our model with multiple passengers types (*market segmentation*) and multiple fare classes (*differential pricing*) to evaluate its potential to handle more complex pricing scenarios if such data were available. Given this relatively modest goal, we introduce three new parameters K_1, K_2 and K_3 to simulate various hypothetical segmentation and pricing scenarios.

In this subsection, we make the simplified assumption that passengers are divided into two categories (business and leisure travelers), that the proportion of business travelers is identical across all markets, and that the airline offers only two fare classes. Specifically, parameter K_1 quantifies the fare differences across the two fare classes. Let us define High Fare = $\mu + K_1\sigma$ and Low Fare = $\mu - K_1\sigma$, where μ and σ denote the average and standard deviation of the fares for each combination of year, quarter, origin airport, destination airport, connection airport (if any), first leg carrier, and second leg carrier (if any) as obtained from the Bureau of Transportation Statistics [29]. Also, parameter K_2 is defined as the fraction of passengers that belong to the category of business passengers in every market (so $1 - K_2$ is the fraction belonging to the category of leisure passengers). In addition, parameter $K_3 \geq 1$ distinguishes between the utility functions of business and leisure pas-

sengers (Equation (3.4)). Particularly, business passengers are assumed to be willing to pay higher fares in return for more convenience. We capture this by multiplying by K_3 the original parameters ($\beta_{i,pt,cl}$) corresponding to departure time of the day, total trip time and number of connections, and by multiplying by $1/K_3$ the original parameter of ticket price. In contrast, leisure passengers have a lower willingness to pay, but are more willing to accept longer travel times and more connections, and are less sensitive to departure time of the day. So we capture this by multiplying by $1/K_3$ the original parameters ($\beta_{i,pt,cl}$) corresponding to departure time of the day, total trip time and number of connections, and by multiplying by K_3 the original parameter of ticket price. Note that K_1 captures the airline's pricing and revenue management decisions, while K_2 and K_3 are parameters characterizing the passenger mix and the extent of heterogeneity of preferences across passenger types, respectively.

Table 3.6 reports various statistics about the solutions obtained with Heuristic 7 for various combinations of parameters K_1, K_2 and K_3 in the following ranges: $0 \leq K_1 \leq 2.5$, $0.25 \leq K_2 \leq 0.75$ and $1 \leq K_3 \leq 3$. Results related to the business passengers who are estimated by the model to purchase a High Fare ticket are reported in the column titled "Business - High Fare", and analogous definitions follow for the columns titled "Business - Low Fare", "Leisure - High Fare" and "Leisure - Low Fare". For each group, we also report the airline's market share in the corresponding segment. For example, the market share in the "Business - Low Fare" column corresponds to the percentage of business passengers who purchase low-fare tickets of the host airline.

In our first set of experiments (rows 2 to 7), we vary K_1 while holding K_2 and K_3 fixed at 0.5 and 2 respectively. These experiments test the effects of an airline changing its pricing strategy by increasing the differential between the fares offerred in different fare classes. Results suggest that, initially, an increase in K_1 value (i.e., stronger price differentiation) enables profit increases. However, beyond $K_1 = 2$, further increase in K_1 seems to decrease profits indicating that $K_1 = 2$ is a "sweet spot" in terms of profit

Table 3.6: Market share ("MS") and revenue with two passenger types and two fare classes

	K_1, K_2, K_3	**Profit ($)**	Business - High Fare		Business - Low Fare		Leisure - High Fare		Leisure - Low Fare	
			MS (%)	Revenue ($)	MS (%)	Revenue ($)	MS (%)	Revenue ($)	MS (%)	Revenue ($)
1	**0, 0.5, 1**	5,126,077	27.7	2,586,802	28.3	2,572,984	29.2	2,613,427	26.6	2,559,731
2	**0, 0.5, 2**	5,464,520	32.0	2,825,312	31.5	2,792,126	25.2	2,177,928	25.3	2,176,134
3	**0.5, 0.5, 2**	6,204,507	66.9	7,542,689	2.6	426,659	18.5	2,266,630	16.7	1,155,479
4	**1, 0.5, 2**	7,677,795	71.2	10,088,041	0.3	304,903	9.0	1,656,606	15.5	694,036
5	**1.5, 0.5, 2**	8,941,834	70.0	12,132,213	0.1	263,742	4.4	1,296,431	13.4	403,797
6	**2, 0.5, 2**	9,570,268	70.3	13,175,734	0.01	226,523	2.1	1,027,068	12.5	280,414
7	**2.5, 0.5, 2**	9,104,495	68.4	11,746,180	≤ 0.01	112,976	1.1	563,873	12.4	201,961
8	**1.5, 0.25, 2**	3,197,389	68.2	5,613,404	0	69,189	3.6	1,888,882	19.5	768,085
9	**1.5, 0.75, 2**	11,123,013	68.1	15,607,341	0	236,678	4.0	328,252	15.3	143,042
10	**2, 0.25, 2**	4,185,653	70.0	6,789,939	≤ 0.01	71,200	2.0	1,835,970	16.4	637,795
11	**2, 0.75, 2**	13,594,989	68.7	18,230,991	0.02	156,213	2.2	327,148	9.4	74,142
12	**2.5, 0.25, 2**	4,660,549	69.1	7,418,060	0	89,448	1.1	1,733,964	12.3	568,131
13	**2.5, 0.75, 2**	15,440,792	67.5	20,163,128	≤ 0.01	138,322	1.1	255,162	8.9	49,716
14	**2, 0.5, 1.5**	9,856,410	65.1	12,883,967	≤ 0.01	228,283	4.8	1,547,928	10.6	353,773
15	**2, 0.5, 3**	10,106,673	77.3	14,251,949	≤ 0.01	147,803	0.5	654,061	12.5	228,943

maximization. Compared to charging a single fare value to all passengers (i.e., $K_1 = 0$), offering two different fares $(K_1 \neq 0)$ may allow for considerable increases in total profit with a maximum profit increase (among the tested values) of over 80%. The incremental revenue clearly comes from the business passengers purchasing higher fare tickets.

In the second set of experiments (rows 8 and 9), we vary K_2 from 0.25 to 0.75 (as compared to 0.5 in row 5), with $K_1 = 1.5$ and $K_3 = 2$. This assesses the effect of changes in the mix of business versus leisure passengers on the optimal timetabling and fleet assignment solutions. Note that the market share for each category does not fluctuate much as the value of K_2 changes (between rows 5, 8 and 9), but the operating profit increases significantly as K_2 increases. This is expected, as higher values of K_2 induce a higher proportion of business passengers. We obtain similar insights in rows 10 and 11 (as compared to row 6) and in rows 12 and 13 (as compared to row 7) for $K_1 = 2$ and $K_1 = 2.5$, respectively. Next (in rows 14 and 15), we vary K_3 from 1.5 to 3 (as compared to 2 in row 6), with $K_1 = 2$ and $K_2 = 0.5$. We find that the operating profit increases with increasing value of K_3. This is because a major driver of operating profits is the "Business - High Fare" revenue, and larger values of K_3 induce a lower price-sensitivity of business passengers (hence increasing the airline's ability to extract more revenue on that segment).

We caution the reader not to interpret the absolute numerical values of results presented in Table 3.6 and in this subsection too literally, because they are based on the aforementioned highly approximate input parameter assumptions. Accurate values of these inputs are available to an airline interested in using these models and algorithms as decision support. But, more importantly, the relative values provide valuable insights. Consistent with the revenue management literature, they underscore the impact of market segmentation and differential pricing on airline operating profitability, even when integrated with timetable development and fleet assignment. Moreover, we find that integrating downstream revenue management dynamics, even in an approximate manner, can improve the flight timetabling and fleet assignment solutions. The model and solution approaches developed in this chapter are endowed with this capability.

3.6.2 Integration with Frequency Planning Decisions

In general, when an airline decides to serve a nonstop segment, frequency plans are established first, usually a year or more before actual departure time. The next step, timetables of departure times and aircraft rotations are established up to 2-6 months before actual departure time [15]. Joint optimization of flight frequency, timetabling and fleet assignment decisions, given their obviously interdependent nature, can potentially yield much larger profit increases than only optimizing the timetabling and fleet assignment decisions. However, flight frequency decisions also depend on a variety of strategic and operational considerations such as, airline business strategy, aircraft orders, airport presence considerations, airport gate and slot availability, etc., which are beyond the scope of this research. It would be naive to assume that airlines can easily make significant changes to their flight frequencies even if these changes indicate a potential for additional profit. In view of these factors, we analyze the impact of some *incremental* frequency planning flexibility by allowing the daily frequency to fluctuate within ± 1 and ± 2 from the existing frequency value on each segment. This is formulated by replacing Constraint (3.20) by Constraint (3.30)

with $\Delta = 1$ and $\Delta = 2$, respectively. Specifically, note that we retain the aircraft count and flow balance constraints to ensure that the schedule remains feasible.

$$Freq_s - \Delta \leq \sum_{t \in \mathcal{T}} \sum_{f \in \mathcal{F}} x_{s,f,t} \leq Freq_s + \Delta, \quad \forall s \in \mathcal{S}, \tag{3.30}$$

Table 3.7 compares the optimal solutions of our integrated approach when frequency is held constant on all segments, when frequency is allowed to fluctuate by at most ± 1, and when frequency is allowed to fluctuate by at most ± 2. For each case, we report several summary statistics to characterize the airline's revenue, operating cost and profit, as well as the number of passengers carried across all Hub-to-Hub ("HH"), Hub-to-Spoke and Spoke-to-Hub ("HS"), and Spoke-to-Spoke ("SS") segments. Here, Seattle (SEA), Anchorage (ANC) and Portland (PDX) are labeled as hub airports and all others are labeled as spoke airports. For each metric, the table reports the relative change from the baseline case where frequency is held constant.

Table 3.7: Effect of integration with frequency planning

	Profit ($)	**Revenue**			**Operating**	**Flights**				**Number of Passengers**	
		Total ($)	Nonstop ($)	One-stop ($)	**Cost**	# Total	# HH	# HS	# SS	Nonstop	One-stop
± 0	4,757,042	9,931,526	6,990,814	2,940,712	5,175,311	395	30	329	36	31,622	11,042
	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)
± 1	5,576,506	12,127,852	7,905,909	4,221,942	6,551,824	468	28	398	42	35,865	15,350
	(+17.23%)	(+22.80%)	(+13.66%)	(+44.58%)	(+31.69%)	(+18.48%)	(-6.67%)	(+20.97%)	(+16.67%)	(+14.14%)	(+39.90%)
± 2	5,869,752	13,029,865	8,384,640	4,645,224	7,160,609	516	28	440	48	38,056	17,383
	(+23.39%)	(+31.94%)	(+20.54%)	(+59.08%)	(+43.93%)	(+30.63%)	(-6.67%)	(+33.74%)	(+33.33%)	(+21.11%)	(+58.43%)

Table 3.7 shows that, from ± 0 to ± 1 margin, total operating profit increases considerably (by 17.23%), while from ± 1 to ± 2 margin, there is a smaller additional increase (from 17.23% to 23.39%). Analogously, the increase in the total number of flights from ± 0 to ± 1 margin is larger (73) than that from ± 1 to ± 2 margin (48). Nonstop revenue increases from \$6.99 million to \$7.91 million to \$8.38 million as we go from ± 0 to ± 1 to ± 2 margin while the one-stop revenue increases from \$2.94 million to \$4.22 million to \$4.65 million. The large percentage increase in one-stop revenue could be explained by the fact that most of the increase in the number of flights is on the Hub-to-Spoke and

Spoke-to-Hub segments. These additional flights contribute moderately to increased nonstop revenue, but very significantly toward a large percentage increase in one-stop revenue by providing better connecting service to many of the low-demand markets that cannot support nonstop flights. This is also reflected in the large percentage increase in the number of one-stop passengers (39.90% and 58.43% increases corresponding to the ± 1 and ± 2 cases, respectively).

As mentioned earlier, frequency planning is a highly complex step in an airline's schedule planning process and it requires paying attention to a variety of strategic and operational considerations. We only consider marginal deviations from existing flight frequency values as a way to demonstrate the ability of our overall modeling and computational approach to provide decision support for the joint optimization problem of frequency planning, timetable development and fleet assignment, by potentially incorporating these strategic and operational considerations. The actual implementation and evaluation of a *comprehensive* frequency planning optimization problem using our modeling and computational framework is an interesting direction for future research.

3.6.3 Impact of an Airline Merger

On December 6th, 2016, the United States Department of Justice approved Alaska Airlines' merger with Virgin America, allowing them to form the fifth largest airline in the United States. Before this merger, Alaska Airlines had its main hubs at Seattle (SEA), Portland (PDX) and Anchorage (ANC) while Virgin America had its main hubs at San Francisco (SFO) and Los Angeles (LAX). In addition, Alaska's original fleet consisted exclusively of the Boeing 737 fleet family, while Virgin's consisted exclusively of the Airbus 320 fleet family. The merger created interesting opportunities to assess the effectiveness of the combined airline in serving the joint network and leveraging the newly heterogeneous fleet to generate new connecting itineraries.

We apply our modeling and computational framework to evaluate and optimize the

timetable of the combined airline. The results are presented in Table 3.8. Columns 2 and 3 correspond to the optimal solution obtained using our modeling and computational framework separately for each of the two airlines. Column 4 reports the total values corresponding to Columns 2 and 3, which serves as the benchmark for comparing the post-merger optimal solutions for the joint airline. After the merger, we evaluate scenarios considering various degrees of integration between the two airlines (Columns 5, 6, 7 and 8). At one end, Column 5 reports a case where the airlines continue to operate the same timetables as before the merger. Any improvement in this case comes from the additional passenger connection opportunities arising from joining of the two networks. Solutions in Columns 4 and 5 correspond to the same flight timetables fleet assignmentd, leading to the same total operating cost value, but have different passenger flows and hence different revenue values. Column 6 reports the benefits of re-optimizing the timetabling as well as the fleet assignment decisions jointly for the combined airline. Additionally, past literature has stated the possibility of unit operating cost reductions achievable due to a merger, resulting from a better reallocation of capital and labor resources leading to more efficient utilization of staff, fuel and maintenance services [103]. Columns 7 and 8 account for this possibility by re-optimizing the timetables and the fleet assignment decisions of the joint airline while assuming a network-wide unit operating cost reduction by 1% and 5%, respectively. All percentage changes reported in Columns 5 through 8 are calculated based on the corresponding numbers in Column 4 used as the baseline.

As expected, compared to Column 4, the one-stop revenue increases significantly (by almost 20%) in Column 5, due to the additional passenger connection opportunities. At the same time, this gets partly offset by lower nonstop revenue because of the displacement of some of the nonstop passengers by the one-stop passengers in this process. This leads to a small increase in the operating profit ($< 1\%$). From Column 5 to Column 6, we observe a significant (7.34%) additional increase in operating profit. This is due to a considerable increase in nonstop as well as one-stop revenue which more than compensates for some

Table 3.8: Assessment of the impact of a merger on the optimal timetabling and fleet assignment decisions

	Pre-Merger			**Post-Merger**			
	Alaska Airlines	Virgin America	Total	Fixed TT No CR	Opt TT No CR	Opt TT 1% CR	Opt TT 5% CR
# of daily flights	395	142	537	537			
# unique destinations	59	21	61	61			
# fleet types	133	54	187	187			
Total passengers	48,691	20,023	68,714 (0%)	66,901 (-2.64%)	68,000 (-1.04%)	67,879 (-1.22%)	67,570 (-1.66%)
Total profit ($)	4,885,476	3,365,813	8,251,289 (0%)	8,318,420 (+0.81%)	8,929,319 (+8.22%)	8,985,121 (+8.89%)	9,206,315 (+11.57%)
Nonstop revenue ($)	7,651,437	4,812,711	12,464,148 (0%)	11,962,136 (-4.03%)	12,242,570 (-1.78%)	12,144,660 (-2.56%)	12,079,316 (-3.09%)
One-stop revenue ($)	2,438,974	525,247	2,964,221 (0%)	3,533,683 (+19.21%)	3,958,583 (+33.55%)	4,059,515 (+36.95%)	4,056,083 (+36.83%)
Operating cost ($)	5,204,934	1,972,465	7,177,399 (0%)	7,177,399 (0%)	7,272,717 (+1.33%)	7,219,288 (+0.58%)	6,929,173 (-3.46%)

*Fixed TT No CR: Fixed timetable and no operating cost reduction.
*Opt TT No CR: Optimized timetable and no operating cost reduction.
*Opt TT 1% CR: Optimized timetable and 1% operating cost reduction.
*Opt TT 5% CR: Optimized timetable and 5% operating cost reduction.

of the corresponding increase in operating cost. In this case, the new fleet assignment solution allows carrying more passengers in profitable markets to increase total revenue. In other words, the merger alone (i.e., the new passenger connection opportunities) results in very small benefits, but the combination of the merger and the re-optimization of flight timetables and fleet assignment solutions results in significant profit improvement.

Under the last two scenarios (i.e., Columns 7 and 8), reductions in unit operating costs result in further improvements in operating profits, as expected. Note, however, that the nonstop revenue decreases while the one-stop revenue increases, compared with Column 6, and that these two effects balance each other out resulting in negligible change in total revenue. Hence the profit enhancements almost exactly reflect the effects of cost reductions. As a result, the profit increase in the 5% cost reduction scenario in Column 8 when compared to that without any cost reduction in Column 6, is approximately five times the profit increase in the 1% cost reduction scenario when compared to that without any cost reduction in Column 6. Interestingly, in each of the post-merger scenarios (i.e., Columns 5 through 8), the total number of passengers stays below the pre-merger value (i.e., Column 4), and yet they all have higher profits than that in the pre-merger scenario. This under-

scores the fact that more passengers need not lead to higher profits. Columns 5 and 6, in particular, highlight the benefits of being able to solve a larger optimization problem thus enabling the capture of more lucrative passengers, which in turn increases the revenue and profit.

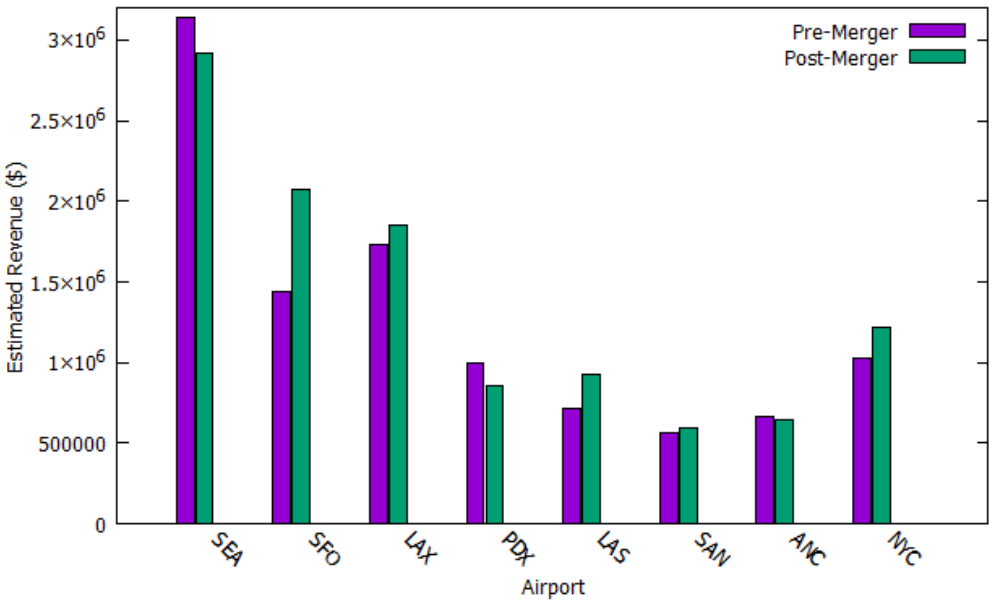


Figure 3.3: In-depth analysis of the estimated fare revenue by airport

Figure 3.3 presents more in-depth analysis of the changes in total revenue corresponding to Columns 4 and 6 of Table 3.8. It displays the top seven airports of the joint Alaska-Virgin network ranked by the estimated number of departing passengers in the optimized solution. Additionally, it also shows the combined data for the three airports in the New York City area (EWR, LGA and JFK, referred to as "NYC"). Interestingly, we observe a post-merger revenue decrease only at the three hubs of the original Alaska Airlines network (namely, SEA, PDX, and ANC). Each of the other airports (including "NYC") shows a post-merger revenue increase between 6% and 44%. In particular, the two hubs of Virgin America (SFO and LAX) as well as San Diego (SAN) and Las Vegas (LAS) show considerable increase in total passenger revenue, which appears to further strengthen the joint airline's position in the West Coast markets. In other words, our integrated flight timetabling

and fleet assignment approach enables re-allocation of aircraft resources to flight segments that leverage the combined network of the two airlines and resulting passenger connection opportunities, in turn leading to increased operating profitability.

3.7 Conclusion

We develop an original modeling and computational framework to comprehensive timetable development and fleet assignment under endogenous passenger choice. Given the flight frequency on each nonstop segment, unconstrained passenger demand in each market and the airline's fleet availability, the approach produces flight timetables and fleet assignment solutions that maximize the airline's profit. Our mathematical model leverages a sales-based linear programming approach that explicitly incorporates the attractiveness of each itinerary and the resulting passenger booking decisions within a large-scale mixed-integer optimization model of airline scheduling. This problem is extremely difficult to solve using off-the-shelf optimization solvers. Therefore, we design a multi-phase solution framework and several additional heuristics to enable practical implementation of the model in reasonable computational times that are consistent with the practical requirements for solving such problems of strategic nature.

In a case study setting leveraging data from a major hub-and-spoke airline carrier in the United States, computational results demonstrate that our algorithm consistently and substantially outperforms a direct implementation of the model using a commercial mixed-integer optimization solver. Most importantly, comparisons with baselines that replicate the various incremental scheduling approaches found in the literature and/or in practice suggest that the real power of the approach developed in this chapter lies in the *combination* of our original modeling framework and our original solution approaches, which can result in significant profit improvements (ranging between 15% and 40%). We also presented a series of additional extensions to integrate frequency planning decisions, to capture the ef-

fects of market segmentation and revenue management practices, and to study the impacts of an airline merger on the optimal timetabling and fleet assignment solutions at various degrees of post-merger integration. These experiments and results demonstrate the versatility and usability of our approach for a variety of strategic planning decisions made by major airlines.

Chapter 4

Transit Planning Optimization under Ride-hailing Competition and Traffic Congestion

Virtually every major city is experiencing rapid growth of ride-hailing as an alternative mode of urban transportation. McKinsey & Company [97] reports that in the United States, the vehicle miles traveled by Uber and Lyft each month increased from 30 million in 2013 to 500 million in 2016—reflecting an annual growth rate of 150%. Even though ride-hailing accounts for only 1% of the nation-wide traffic, its strong spatial concentration results in significant traffic increases in the densest metropolitan areas [41]. This transformation provides clear opportunities to enhance urban mobility, due to the convenience of smartphone technology and cashless transactions, competitive prices, and low wait times enabled by a broad driver base [43]. But it also creates a number of challenges for urban planners and policy-makers.

One of these challenges is declining public transit ridership in many cities worldwide [see, e.g., 146, 136]. This decline can stem from public transit deficiencies, falling gas prices and increased car ownership, but also the new availability of ride-hailing. Graehler

et al. [63] find that the ride-hailing entry reduces transit ridership by 1% to 2% each year—leading to a 12.7% reduction in bus ridership in San Francisco since Uber's entry in 2010. A second challenge is growing traffic congestion, which is responsible for annual costs estimated at $160 billion in the United States in 2014 [115]. While traffic congestion is not a new phenomenon, several reports argue that it is amplified by ride-hailing [see, e.g., 112, 52]. Mangrum and Molnar [93] document a 15.8% increase in median travel times in Midtown Manhattan between 2013 and 2016, most of which is caused by an increase in vehicle supply due to taxi deregulation and ride-hailing growth. A recent study from Uber and Lyft suggests that ride-hailing contributes to a spike in vehicle miles traveled in six major US metropolitan areas [56]. At least in theory, these effects could be explained by the fact that ride-hailing increases demand for roadway usage (due, in part, to the displacement of public transit demand) and induces idle driving between trips. In summary, while ride-hailing creates critical opportunities for urban mobility, it also puts cities at risk of running empty subways and buses while roads become increasingly congested.

What can cities do in this environment? A first (quite drastic) option is to cap the number of ride-hailing vehicles. In 2018, for instance, New York City limited the number of new ride-hailing licenses; Austin, TX even attempted to ban ride-hailing platforms altogether. A second option is to use congestion pricing. For instance, New York City started charging a $2.75 surcharge per ride-hailing trip in January 2019; Singapore's Electronic Road Pricing system charges a toll in the busiest areas; and the San Francisco airport charges extra fees to ride-hailing vehicles. A third (less punitive) option is to partner with ride-hailing providers. For instance, public transit authorities can leverage ride-hailing to provide first- and last-mile services for public transit (as piloted in San Francisco, Atlanta, Philadelphia), to subsidize trips for disabled or low-income residents (as piloted in Pinellas County, FL), and to respond to 911 calls (as proposed in Washington, DC). Another option—which can perhaps be more easily implemented in the short term—is to revise public transit schedules to complement ride-hailing. This is the focus of this research.

Specifically, this research optimizes public transit schedules, given inter-modal competition with ride-hailing and its impacts on road congestion. This approach takes the perspective of a city planner, who adjusts the frequency and schedule of service on each public transit line in order to enhance urban mobility and mitigate congestion. The approach is applied here to tactical adjustments of transit schedules; it could also be applied to strategic network design—by proposing new candidate transit lines and optimizing frequency and timetabling on this new network.

The problem of public transit planning is not new—it has been around for decades. But what is new is the increased availability of ride-hailing that provides a viable transportation option on many origin-destination pairs where previous alternatives to public transit were too expensive or too inconvenient. At the same time, public transit remains more attractive on other origin-destination pairs with severe road congestion or high ride-hailing prices. The growth of ride-hailing thus raises the following question: how shall a public transit authority re-allocate its resources to provide better mobility options where ride-hailing is least competitive, and perhaps divert resources away from areas that are well-served by ride-hailing? To answer this question, this research provides a new scheduling toolbox that captures the interdependencies between public transit schedules, ride-hailing, traffic congestion, and travelers' mode choices.

This research comes at a time of major opportunities for city planning authorities. Despite the overall decline in public transit, transit ridership has been on the rise in a few cities that underwent major overhauls of their transit networks [137]. Notably, Seattle and Houston added new rail services and re-oriented buses to add frequency on critical lines. Shortly thereafter, transit ridership increased in these two cities by 4.1% and 2.3%—the largest growth numbers in the United States [129]. This suggests opportunities to comprehensively revise transit networks to meet changing demand patterns in the new era where travelers' behaviors are shaped by public transit offerings, ride-hailing offerings, and traffic congestion. Yet, there is limited research that provides a blueprint on *how* to achieve these

objectives. Our research does just that.

The remainder of the chapter is organized as follows. Section 4.1 reviews the relevant literature. Section 4.2 presents our mathematical model. Section 4.3 describes our iterative solution algorithm. Section 4.4 reports results from our case study in New York City. Section 4.5 confirms the robustness of our results to model specifications. Section 4.6 synthesizes our findings.

4.1 Literature Review

This research lies at the intersection of public transit, traffic congestion and ride-hailing operations. Each of these fields includes vast streams of studies. Thus, we stay focused on the literature that is most closely related to our central question: how can we enhance transit schedules while accounting for the presence of ride-hailing to improve the overall urban transportation ecosystem?

4.1.1 Public Transit Planning

Public transit planning is often broken down into three main steps: 1) network design, which determines stop locations and routes; 2) line planning or frequency setting, which determines the set of transit lines and line frequencies; and 3) timetabling, which determines the transit vehicle departure times [65]. From the 1970s, transit network design has captured researchers' attention [see, e.g. 36]. Objective functions include social welfare maximization [39, 45], minimization of the generalized travel costs of walking, waiting, schedule displacement, and transfers [47], minimization of total costs of passenger travel, vehicle operations and stop construction [98], etc. Computationally, the network design problem is highly complex, often requiring metaheuristics like genetic algorithms, local search, simulated annealing and tabu search [see, e.g., 54, 46]. Line planning and frequency setting also have a long history [see, e.g., 58, 34]. Recent studies have focused on

improving system-wide efficiency or minimizing social costs [114, 113, 27, 69]. They also rely on metaheuristics [11, 133], or branch-price-cut algorithms [40].

Within the timetabling literature, a major stream of studies has focused on improving passenger transfers and transit vehicle synchronization [70, 35, 130, 149, 66, 87]. Another stream of studies minimizes the weighted sum of passenger costs and operational costs [see, e.g., 33, 21] or maximizes a social welfare function combining travel times reduction, passenger transfer opportunities, vehicle utilization and operator profitability [e.g., 79, 80, 104]. Various analytical [see, e.g., 153] and computational [see, e.g., 37] approaches have been developed for welfare maximization in transit timetabling.

This research contributes to the literature on public transit planning by explicitly integrating the effects of traffic congestion and ride-hailing on service frequency and timetabling—thus providing insights on how transit systems can evolve to best respond to contemporary urban challenges.

4.1.2 Relationship between Traffic Congestion and Public Transit

Modeling traffic flows and road congestion has a long history [see, e.g., 44, 94]. However, the optimization literature on understanding the effects of transit network changes on road congestion is relatively scarce. Farahani et al. [55] review the related problems of road network design and public transit network design. Anderson [10] points out the potential benefits of public transit for congestion mitigation—even for commuters who rarely use public transit, since transit passengers would otherwise have predominantly traveled on the most congested roads at the most congested times. Small [124] and Winston and Maheshri [148] point out that public transit investments involve high construction and operating costs, are not necessarily the most economical congestion mitigation options. Therefore, this research focuses on transit frequency planning and timetabling, instead of network redesign. Related to our research, Beaudoin et al. [22] show that the optimal design of public transit systems should consider not only the welfare of public transit users, but also the effects

on automobile users and resulting congestion. Their result indicates that a 10% increase in transit capacity reduce the number of vehicles on the road by 0.7%. In this research, we leverage these interdependencies by designing adjustments to public transit schedules that can bring system-wide benefits, given traffic congestion patterns.

There is also some work focusing on the effects of road congestion on public transit delays and passenger waiting times [9]. However, this type of interdependence is out of the scope of this research. Instead, we assume that transit operations are not significantly affected by road traffic flows. So this research is most directly relevant for transit systems for which this property holds—including trams, passenger trains, rapid transit, and buses that use dedicated lanes. For other bus services, the approach developed in this research serves as an approximation.

4.1.3 Ride-hailing Operations

The emergence of ride-hailing platforms has generated a fast-growing field of research in operations management, information systems and transportation science. For instance, Cachon et al. [31] and Banerjee et al. [14] compare dynamic and static pricing for two-sided ride-hailing platforms with strategic customers and self-scheduled capacity. Taylor [135] and Bai et al. [12] optimize pricing policies in strategic queuing settings where customers balance prices and wait times. Bimpikis et al. [26] show that differentiating prices based on origin and destination can enhance the platform's revenues.

At the operational level, the optimization of ride-hailing operations can be formulated as a pickup and delivery problem—which is highly difficult to solve [88, 109]. Alonso-Mora et al. [8] decompose the network into many sub-networks to reduce the dimensionality. Fagnant et al. [53], Steiner and Irnich [127] focus on the impact of ride-hailing on public transit, congestion, and the environment. Agatz et al. [3], Wang et al. [144], Lee and Savelsbergh [85], Chen and Nie [38], Stiglic et al. [128] consider ride-hailing as a feeder system for public transit to address the "first-mile" and "last-mile" problems. In this re-

search, we instead revise public transit schedules in view of ride-hailing operations, in order to maximize the resulting transit benefits.

4.2 Model Description

In this section, we present our Mixed Integer Nonlinear Programming (MINLP) model of transit scheduling with endogenous passenger choice and road congestion. We first specify our setting and assumptions. We then define the notation and formulate the mathematical model.

4.2.1 Setting and Assumptions

We define a *transit line* as an ordered sequence of transit stops consecutively traversed by a transit vehicle, and an *edge* in the network as the ordered pair of transit stops sequentially traversed by a transit vehicle. A *transit passenger itinerary* is a sequence of transit stops, and the corresponding sequence of departure times, that could be sequentially traversed by a passenger along the transit network (not necessarily on the same vehicle due to possible line transfers). A road network consists of a set of *nodes* and *road segments* (continuous stretches of road between two consecutive nodes). An *origin-destination (OD) pair* is defined as an ordered pair of nodes. A road *travel path* is made up of a sequence of road segments that connects the two end nodes corresponding to an OD pair.

Figure 4.1 explains these concepts using a simple example with three transit stops. Stop a is on line 1, Stop c is on line 2, and Stop b is on both lines. For the OD pair $Origin$-$Dest$, the transit passengers need to walk from $Origin$ to Stop a, take transit line 1 from Stop a to Stop b, transfer at Stop b to line 2, travel via line 2 from Stop b to Stop c, and then walk to $Dest$ from Stop c. The road network in this small example consists of four road segments: $Origin - E$, $E - Dest$, $Origin - D$ and $D - Dest$. If the passenger chooses ride-hailing, the ride-hailing vehicle can take either of two paths: $Origin - E - Dest$ and

$Origin - D - Dest.$

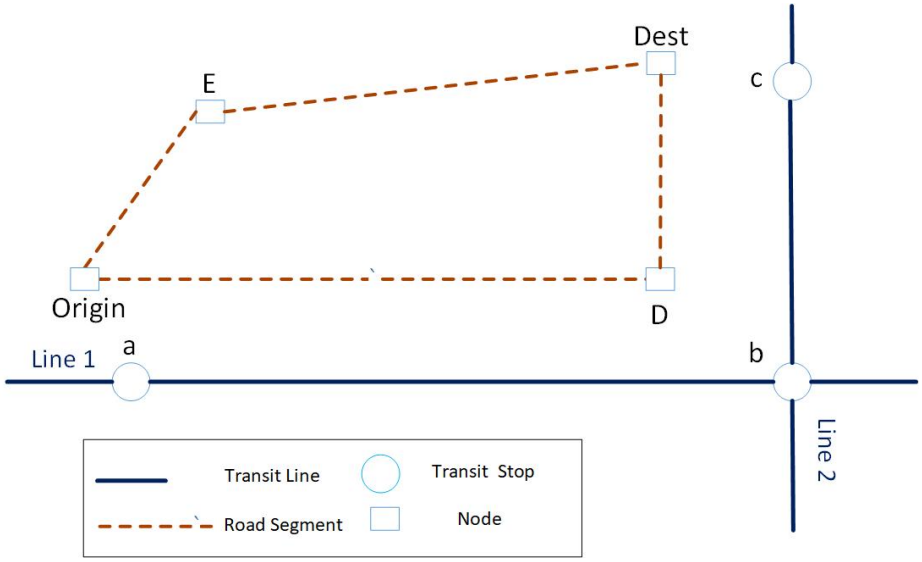


Figure 4.1: An example network

Besides ride-hailing and public transit, we define an *outside option* that encompasses all other modes of transportation (e.g., walking, self-driving, biking) and the option to not travel at all. We assume that the service level of this outside option is exogenous. Its "market share" corresponds to the fraction of passengers that travel neither by transit nor with ride-hailing.

4.2.2 Mathematical Notations

Sets

$\mathcal{Q}$: Set of OD (origin-destination) pairs, indexed by (o, d).

$\mathcal{H}$: Set of passenger types, indexed by h.

$\mathcal{VT}$: Set of vehicle types in the transit system, indexed by v.

$\mathcal{ST}$: Set of first and last stops of all transit lines.

$\mathcal{G}$: Set of all transit lines, indexed by g.

$\mathcal{V}_g$: Subset of vehicle types compatible with transit line g.

$\mathcal{G}^o_{st,v}$: Subset of transit lines compatible with vehicle type v with stop st as origin.

$\mathcal{G}^d_{st,v}$: Subset of transit lines compatible with vehicle type v with stop st as destination.

$\mathcal{E}_g$: Set of edges of the transit network that constitute transit line g, indexed by e.

$\mathcal{PT}'$: Set of all transit timeslots during the modeling horizon, indexed by t.

$\mathcal{PT}$: Subset of all transit timeslots during the optimization horizon, indexed by t.

$\mathcal{I}^{\text{all}}$: Set of possible transit itineraries, indexed by i.

$\mathcal{I}_{od}$: Subset of transit itineraries for passengers traveling on OD pair (o, d).

$\tilde{\mathcal{I}}_{e,g,t}$: Subset of transit itineraries which use edge e in transit line g, with line departure time t.

$\mathcal{TT}'$: Set of all traffic timeslots during the modeling horizon, indexed by tt.

$\mathcal{TT}$: Set of all traffic timeslots during the optimization horizon, indexed by tt.

$\mathcal{P}^{\text{all}}$: Set of possible travel paths through the road network, indexed by p.

$\mathcal{P}_{od}$: Subset of travel paths corresponding to OD pair (o, d).

$\mathcal{SN}^{all}$: Set of road segments in the road network, indexed by s.

$\mathcal{SN}_p$: Subset of road segments in travel path p.

A few comments are noteworthy. First, each *passenger type* $h \in \mathcal{H}$ defines a group of passengers with similar preferences and attributes. In this research, we consider passenger heterogeneity along two dimensions: time sensitivity (more time-sensitive vs. less time-sensitive passengers) and preferred departure time (e.g., passengers who wish to travel ideally at 7 am, at 8 am, at 9 am, etc.). Second, the subsets $\mathcal{V}_g$ reflect that each vehicle can only operate on a subset of transit lines, due to design and engineering characteristics. Third, a *transit timeslot* and a *traffic timeslot* are defined as the smallest units of time discretization used for transit timetabling and traffic flow modeling, respectively (they do not necessarily coincide). Fourth, the *line departure time* is defined as the planned departure time of a transit vehicle from the first stop on a transit line, and the *line end time* is defined

as the sum of the line departure time, scheduled transit time, and minimum maintenance time (if any). Fifth, the *optimization horizon* is defined as the period for which we optimize transit schedules. However, transit vehicles whose line departure times are just before the optimization horizon also continue to run in the early part of the optimization horizon, and thus have an impact on decisions during the optimization horizon. Similarly, transit vehicles whose line departure times are at the end of the optimization horizon need to account for passenger demand immediately thereafter. Therefore, the *modeling horizon* starts before the optimization horizon and ends after the end of the optimization horizon. This explains the difference between $\mathcal{PT}'$ and $\mathcal{PT}$ and between $\mathcal{TT}'$ and $\mathcal{TT}$. Note that we only optimize the timetable in the optimization horizon; for the other time periods in the modeling horizon, we take the actual timetable as an input.

Parameters

$Dem_{od,h}$: Demand of passengers of type h on OD pair (o, d).

$Avail_v$: Number of available transit vehicles of type v.

$MinPT/MaxPT$: First (resp. last) transit timeslot of the optimization horizon.

$Ope^{TR}_{g,v}$: Transit unit operating cost for line g for vehicle type v.

Cap_v : Passenger-carrying capacity of a transit vehicle of type v.

$Veh_{s,tt}$: Number of vehicles on road segment s in traffic timeslot tt (excluding ride-hailing).

$DC_{od,h}$: Cost of the outside option for passengers with type h on OD pair (o, d).

$A^O_{od,h}$: Attractiveness of the outside option for passengers with type h on OD pair (o, d).

$D(g, t)$: Line departure time of a transit vehicle on line g whose line end time is in transit timeslot t.

$ETV_{v,g,t}$: Number of operations on transit line g with line departure time t using transit vehicle type v that start before the optimization horizon (i.e., $t < MinPT$) and that end after the start time of the optimization horizon ("early transit vehicles").

Direct Decision Variables

$x_{v,g,t}$: Number of services on line g with line departure time t using vehicle type v.

Indirect Decision Variables

$a^{TR}_{i,h}/s^{TR}_{i,h}/tc^{TR}_{i,h}$: Attractiveness (resp., market share and generalized travel cost) corresponding to transit itinerary i for passengers with type h.

$a^{R}_{od,h,tt}/s^{R}_{od,h,tt}/tc^{R}_{od,h,tt}$: Attractiveness (resp., market share and generalized travel cost) corresponding to ride-hailing trips departing in traffic timeslot tt for passengers with type h on OD pair (o,d).

$s^{O}_{od,h}$: Market share of the outside option for passengers with type h on OD pair (o,d).

$y^{-}_{v,st,t}/y^{+}_{v,st,t}$: Number of transit vehicles of type v located at stop st just before (resp., just after) timetslot t.

w_i : Passenger waiting time for transit itinerary i.

$ope^{R}_{od,tt}$: Operating cost of ride-hailing vehicles starting in traffic timeslot tt on OD pair (o,d).

$ext^{R}_{od,tt}$: External cost of ride-hailing vehicles starting in traffic timeslot tt on OD pair (o,d), including environmental, public health, and safety-related cost.

$tra_{s,tt}$: Per-vehicle travel time on road segment s in traffic timeslot tt.

$tc^{V}_{s,tt}$: Per-vehicle travel time cost of other vehicles on road segment s in traffic timeslot tt.

$ext^{V}_{s,tt}$: Per-vehicle external cost of other vehicles on road segment s during traffic timeslot tt, including environmental, public health, and safety-related costs.

$ope^{V}_{s,tt}$: Per-vehicle operating cost for other vehicles on road segment s in traffic timeslot tt.

$ratio^{R}_{od,p,tt}$: Of all ride-hailing providers serving a passenger on OD pair (o,d) and starting service in traffic timeslot tt, the fraction who use travel path p.

$rhfare_{od,tt}$: Ride-hailing fare for OD pair (o,d) starting in traffic timeslot tt.

The direct decision variables determine the frequency of each transit line in any transit timeslot. All other variables are not directly controlled by the transit planners. Still, they

depend on transit timetabling and, vice versa, impact the timetabling solution. These interdependencies are captured through the itinerary attractiveness variables, which impact traffic flow assignments, travel costs, and, ultimately, the market share variables. Note that the ride-hailing fares are defined as decision variables because they depend on travel times, and hence on traffic congestion; however, the unit price (per unit of time) is treated as exogenous.

4.2.3 Model Formulation

The mathematical formulation of the model is given as follows.

$$\begin{aligned}\text{Minimize} \quad & \sum_{h\in\mathcal{H}}\sum_{(o,d)\in\mathcal{Q}}\sum_{i\in\mathcal{I}_{od}} Dem_{od,h}\times s^{TR}_{i,h}\times tc^{TR}_{i,h} + \sum_{v\in\mathcal{V}}\sum_{t\in\mathcal{PT}'}\sum_{g\in\mathcal{G}} Ope^{TR}_{g,v}\times x_{v,g,t} + \\ & \sum_{tt\in\mathcal{TT}}\sum_{h\in\mathcal{H}}\sum_{(o,d)\in\mathcal{Q}} Dem_{od,h}\times s^{R}_{od,h,tt}\times (tc^{R}_{od,h,tt} + ope^{R}_{od,tt} + ext^{R}_{od,tt}) + \\ & \sum_{tt\in\mathcal{TT}}\sum_{s\in\mathcal{SN}^{all}} (ope^{V}_{s,tt} + ext^{V}_{s,tt} + tc^{V}_{s,tt})\times Veh_{s,tt} + \\ & \sum_{h\in\mathcal{H}}\sum_{(o,d)\in\mathcal{Q}} Dem_{od,h}\times s^{O}_{od,h}\times DC_{od,h} \qquad (4.1)\end{aligned}$$

subject to: **Vehicle Availability and Consistency Constraints:**

$$x_{v,g,t} = 0,\ \forall g\in\mathcal{G}, v\in\mathcal{V}\setminus\mathcal{V}_g, t\in\mathcal{PT}' \qquad (4.2)$$

$$\sum_{st\in\mathcal{ST}} y^{-}_{v,st,MinPT} + \sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{PT}'} ETV_{v,g,t} \le Avail_v, \quad \forall v\in\mathcal{VT} \qquad (4.3)$$

Transit Vehicles Flow Balance Constraints:

$$y^{-}_{v,st,(t+1)} = y^{+}_{v,st,t}, \quad \forall v\in\mathcal{VT}, st\in\mathcal{ST}, t\in\mathcal{PT}\setminus\{MaxPT\} \qquad (4.4)$$

$$\begin{aligned}& y^{-}_{v,st,t} + \sum_{g\in\mathcal{G}^{d}_{st,v}} x_{v,g,D(g,t)} + \sum_{g\in\mathcal{G}^{d}_{st,v}} ETV_{v,g,D(g,t)} = y^{+}_{v,st,t} + \sum_{g\in\mathcal{G}^{o}_{st,v}} x_{v,g,t}, \\ & \forall v\in\mathcal{VT}, st\in\mathcal{ST}, t\in\mathcal{PT} \qquad (4.5)\end{aligned}$$

Demand Constraints:

$$\sum_{i\in\mathcal{I}_{od}} s^{TR}_{i,h} + \sum_{tt\in\mathcal{TT}} s^{R}_{od,h,tt} + s^{O}_{od,h} = 1, \quad \forall (o,d)\in\mathcal{Q}, h\in\mathcal{H} \qquad (4.6)$$

$$A^{O}_{od,h}\times s^{R}_{od,h,tt} = a^{R}_{od,h,tt}\times s^{O}_{od,h}, \quad \forall (o,d)\in\mathcal{Q}, h\in\mathcal{H}, tt\in\mathcal{TT} \qquad (4.7)$$

$$A^{O}_{od,h} \times s^{TR}_{i,h} \leq a^{TR}_{i,h} \times s^{O}_{od,h}, \quad \forall i \in \mathcal{I}_{od}, (o,d) \in \mathcal{Q}, h \in \mathcal{H} \tag{4.8}$$

Capacity Constraints:

$$\sum_{(o,d)\in\mathcal{Q}} \sum_{h\in\mathcal{H}} \sum_{i\in\tilde{\mathcal{I}}_{e,g,t}\cap\mathcal{I}_{od}} (Dem_{od,h} \times s^{TR}_{i,h}) \leq \sum_{v\in\mathcal{V}} Cap_v \times (x_{v,g,t} + ETV_{v,g,t}),$$
$$\forall e \in \mathcal{E}_g, g \in \mathcal{G}, t \in \mathcal{PT}' \tag{4.9}$$

Traffic Flow Assignment: $[\mathbf{ratio^R}, \mathbf{tra}, \mathbf{tc^V}] = f_{Rout}(\mathbf{s^R})$ (4.10)

Dynamics of the ride-hailing and public transit systems

Ride-hailing Fare: $\mathbf{rhfare} = f_{rhfare}(\mathbf{ratio^R}, \mathbf{tra})$, (4.11)

Public Transit Waiting Time: $\mathbf{w} = f_{WaitPT}(\mathbf{x})$ (4.12)

Generalized Travel Cost and Attractiveness Calculations:

$$\mathbf{tc^R} = f_{CostRH}(\mathbf{ratio^R}, \mathbf{tra}) \tag{4.13}$$

$$\mathbf{a^R} = f_{AttrR}(\mathbf{tc^R}, \mathbf{rhfare}) \tag{4.14}$$

$$\mathbf{tc^{TR}} = f_{CostPT}(\mathbf{w}) \tag{4.15}$$

$$\mathbf{a^{TR}} = f_{AttrR}(\mathbf{tc^{TR}}) \tag{4.16}$$

Operating Costs and External Costs:

$$\mathbf{ope^R} = f_{OpeR}(\mathbf{tra}, \mathbf{ratio^R}) \tag{4.17}$$

$$\mathbf{ext^R} = f_{ExtR}(\mathbf{tra}, \mathbf{ratio^R}) \tag{4.18}$$

$$\mathbf{ope^V} = f_{OpeV}(\mathbf{tra}) \tag{4.19}$$

$$\mathbf{ext^V} = f_{ExtV}(\mathbf{tra}) \tag{4.20}$$

Domain of Definition:

$\mathbf{x}, \mathbf{y}^-, \mathbf{y}^+$ non-negative integer (4.21)

$\mathbf{s^{TR}}, \mathbf{s^R}, \mathbf{s^O} \geq 0$ (4.22)

$\mathbf{w}, \mathbf{tc^{TR}}, \mathbf{tc^R}, \mathbf{ope^R}, \mathbf{ext^R}, \mathbf{tra}, \mathbf{tc^V}, \mathbf{ext^V}, \mathbf{ope^V}, \mathbf{ratio^R}, \mathbf{rhfare} \geq 0$

(4.23)

$\mathbf{a^{TR}}, \mathbf{a^R} \geq 0$ (4.24)

Objective Function: The model minimizes the system-wide cost of the closed system consisting of the public transit operator, the ride-hailing operator, all passengers (those taking transit, ride-hailing and outside option), and all vehicles on the road network. The first term of the objective function is the cost to transit passengers. The second term is the transit operating cost (e.g., crew, fuel, maintenance). The third term is the ride-hailing cost, including the cost to ride-hailing passengers, the operating cost of ride-hailing vehicles and their external cost. The fourth term includes the time, operating and external costs corresponding to all other vehicles on the roads. The last term is the cost borne by the outside option passengers, who choose other travel modes.

Vehicle Availability and Consistency Constraints: Constraints (4.2) restrict service to compatible vehicles. Constraints (4.3) enforces the number of available transit vehicles of each type.

Transit Vehicles Flow Balance Constraints: Constraints (4.4) and (4.5) ensure flow conservation of transit vehicles across the network, for each vehicle type. We ensure flow balance at the first and last stop of each line—which automatically ensures flow balance at intermediate stops.

Demand Constraints: Constraints (4.6) ensure that the market shares across all alternatives (i.e., public transit itineraries, ride-hailing departure timeslots, and the outside option) sum up to 1 for each passenger type on every OD pair. Next, we use the Sales Based Linear Programming (SBLP) model from Gallego et al. [59] to incorporate passenger choice into our model: Constraints (4.7) and (4.8), respectively, ensure that the market share of every ride-hailing and transit alternative is proportional to its attractiveness for every passenger type on every OD pair. Constraints (4.8) are inequalities to reflect that the transit market share may be lower than that dictated by its attractiveness due to the passenger-carrying restrictions of transit vehicles.

Capacity Constraints: Constraints (4.9) impose that the number of passengers traveling on any edge of the public transit line cannot exceed the capacity of the corresponding

transit vehicles.

Traffic Flow Assignment: Constraints (4.10) assign paths to each ride-hailing trip, given the ride-hailing market share ($s^R_{od,h,tt}$). To this end, we use a traffic flow assignment model, detailed in Section 4.3.2. It returns the fraction of ride-hailing vehicles on each path ($ratio^R_{od,p,tt}$), the travel times ($tra_{s,tt}$) and the travel costs ($tc^V_{s,tt}$).

Dynamics of the ride-hailing and public transit systems: Constraints (4.11)-(4.20) capture the impact of the public transit schedule on the dynamics of the ride-hailing and public transit systems. We describe them briefly below and provide additional details in Appendix B.1.

Ride-hailing Fare: Constraints (4.11) calculate the fare for each ride-hailing trip ($rhfare_{od,tt}$) as a function of the distance (given by $ratio^R_{od,p,tt}$) and the in-vehicle time ($tra_{s,tt}$). We assume that the waiting time of ride-hailing passengers is not affected by the transit schedule changes.

Public Transit Waiting Time: The waiting time w_i of itinerary i is calculated by adding the waiting time at the first station (as a function of transit frequency, provided by $x_{v,g,t}$) and the transfer times at the intermediate stops (time difference between the start time of the next service and the arrival time of the current service).

Generalized Travel Cost and Attractiveness Calculations: Constraints (4.13) to (4.16) calculate the cost and attractiveness of each transit itinerary and each ride-hailing trip, for each passenger type. In Constraints (4.14) and (4.16), the attractiveness values of ride-hailing ($a^R_{od,h,tt}$) and transit ($a^{TR}_{i,h}$) are obtained from passengers' utilities, expressed as linear functions of a set of attributes that affect passengers' choice. We denote these attribute vectors by $\boldsymbol{z}^R_{od,h,tt}$ and $\boldsymbol{z}^{TR}_{i,h}$ and the corresponding vectors of linear coefficients by $\boldsymbol{\beta}^R_{od,h,tt}$ and $\boldsymbol{\beta}^{TR}_{i,h}$, for ride-hailing and transit respectively. In each case, the utility consists of the following attributes: walking time, waiting time, in-vehicle travel time, trip cost, start time,

and transport mode [88]. We have:

$$a^R_{od,h,tt} = \exp\left(\boldsymbol{\beta}^R_{od,h,tt} \cdot \mathbf{z}^R_{od,h,tt}\right) \qquad \forall (o,d) \in \mathcal{Q}, h \in \mathcal{H}, tt \in \mathcal{TT} \tag{4.25}$$

$$a^{TR}_{i,h} = \exp\left(\boldsymbol{\beta}^{TR}_{i,h} \cdot \mathbf{z}^{TR}_{i,h}\right) \qquad \forall i \in \mathcal{I}_{od}, (o,d) \in \mathcal{Q}, h \in \mathcal{H} \tag{4.26}$$

Since the ride-hailing and transit fares are the internal costs within the system, they get canceled out when calculating the system-wide costs. Therefore, the cost to ride-hailing passengers $tc^R_{od,h,tt}$ and the cost to transit passengers $tc^{TR}_{i,h}$ are their utility values excluding the trip fares.

Operating Costs and External Costs: Constraints (4.17) to (4.20) compute the operating and external costs of ride-hailing and other vehicles, as a function of travel times ($tra_{s,tt}$), and the travel paths of ride-hailing vehicles (given by $ratio^R_{od,p,tt}$).

Domain of definition: Constraints (4.21) to (4.24) define the domain of all decision variables.

Additional constraints on transit operations: In order to guarantee a certain level of service for transit passengers, the transit operator may impose a minimum number of operations via Constraints (4.27). Here, $MinFreq_{g,t}$ is defined as the minimum frequency of transit line g with starting transit timeslot t. Alternatively, a minimum level of service requirement may be enforced with a lower bound $MinTotalFreq_g$ on the total frequency of a line g throughout the morning rush hour, via Constraints (4.28). A transit operator may also be concerned about not changing the aggregate transit timetable too much in terms of the total transit vehicle-miles. To this end, Constraints (4.29) and (4.30) ensure that the aggregate schedule does not fall outside of a reasonable range given by $MinTotalMiles$ and $MaxTotalMiles$. Here, Len_g is defined as the length (in miles) of transit line g. Note that not all of these minimum level of service constraints may be relevant for a given transit operator. Our computational experiments presented in Section 4.4 have been performed under various combinations of Constraints (4.27)–(4.30).

$$\sum_{v \in \mathcal{V}_g} x_{v,g,t} \geq MinFreq_{g,t}, \ \forall g \in \mathcal{G}, t \in \mathcal{PT} \tag{4.27}$$

$$\sum_{v \in \mathcal{V}_g} \sum_{t \in \mathcal{PT}} x_{v,g,t} \geq MinTotalFreq_g, \quad \forall g \in \mathcal{G} \tag{4.28}$$

$$\sum_{g \in \mathcal{G}} \sum_{v \in \mathcal{V}_g} \sum_{t \in \mathcal{PT}} Len_g \times x_{v,g,t} \geq MinTotalMiles \tag{4.29}$$

$$\sum_{g \in \mathcal{G}} \sum_{v \in \mathcal{V}_g} \sum_{t \in \mathcal{PT}} Len_g \times x_{v,g,t} \leq MaxTotalMiles \tag{4.30}$$

4.3 Solution Approach

The MINLP presented in Section 4.2 optimizes public transit schedules, while capturing interdependencies between transit, ride-hailing and other vehicles. The performances of these components are interrelated due to passenger choice decisions and traffic flows. However, solving this integrated model to optimality is complicated by three main factors: large network size, large number of discrete variables, and nonlinear non-convex model structure. First, the size of the mathematical model is extremely large for realistic problem instances; for example, in the case study network in Section 4.4, passenger demand is sparsely distributed across over 16 million node pairs. Second, the model includes a very large number of integer variables. Our case study network considered in Section 4.4 involves over 1 million transit passenger itinerary variables. Third, many components of our model are highly nonlinear, and some are even non-convex. For instance, several terms in the objective function involve a product of a flow variable (e.g., number of transit passengers) and a cost variable (e.g., transit time and travel cost)—both of which are endogenous to the passenger choice model and resulting traffic flows. Similarly, the equations defining the dynamics of the system (Constraints (4.10)–(4.20)) are all nonlinear. Moreover, the use of the Multinomial Logit (MNL) model of passenger choice results in non-convex attractiveness calculations in equations (4.25) and (4.26). For all these reasons, direct implementation of

this MINLP with commercial solvers typically does not even yield any feasible solution. We thus propose a new solution approach to derive high-quality solutions in reasonable computational times for this large-scale, mixed-integer, non-convex model. This solution approach is outlined in Figure 4.2, detailed in the remainder of this section, and implemented and evaluated in Section 4.4.

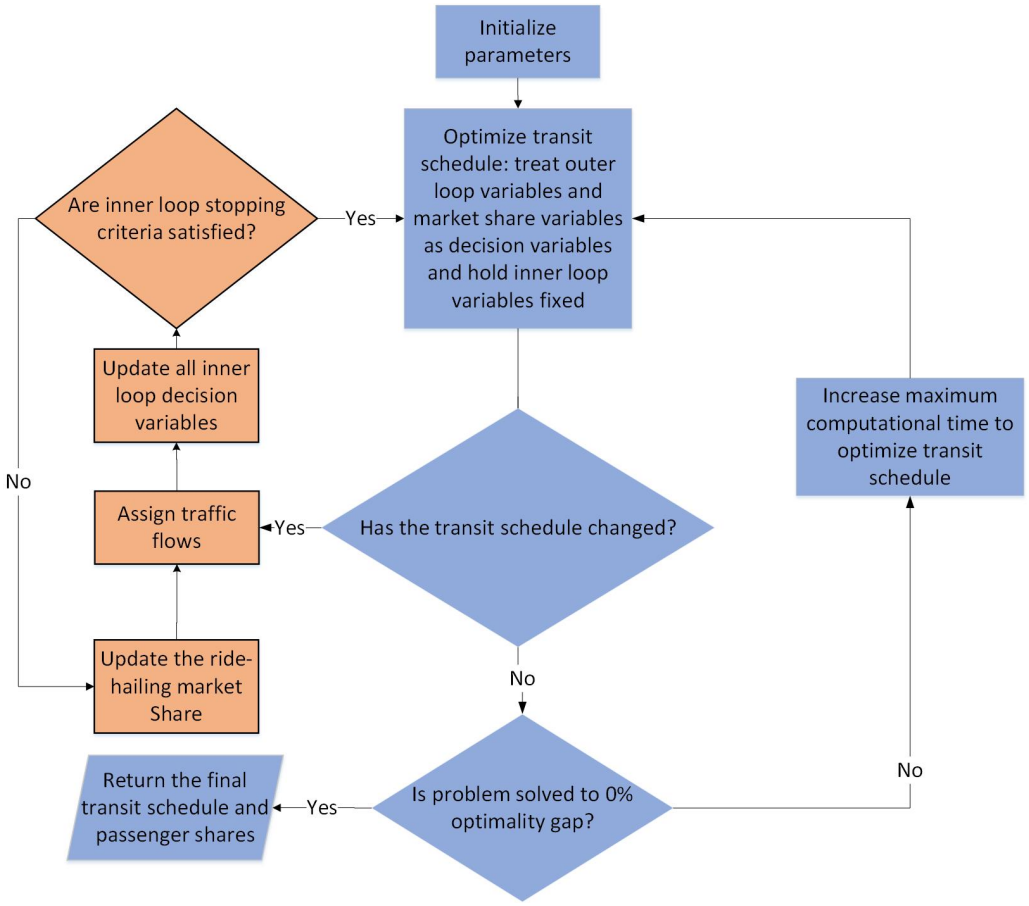


Figure 4.2: Solution approach

In this solution approach, we divide the variables into three categories:

Outer Loop Variables: strategic transit scheduling decisions ($x_{v,g,t}$, $y^{-}_{v,st,t}$ and $y^{+}_{v,st,t}$).

Passenger Share Variables: passenger shares on each mode ($s^{TR}_{i,h}$, $s^{R}_{od,h,tt}$, and $s^{O}_{od,h}$).

Inner Loop Variables: traffic flows ($ratio^{R}_{od,p,tt}$), transit waiting times (w_i), travel costs ($tc^{TR}_{i,h}$, $tc^{R}_{od,h,tt}$, $ope^{R}_{od,tt}$, $ext^{R}_{od,tt}$, $tra_{s,tt}$, $tc^{V}_{s,tt}$, $ext^{V}_{s,tt}$, $ope^{V}_{s,tt}$), ride-hailing fares

(i.e., $rhfare_{od,tt}$), and attractiveness of transit and ride-hailing ($a^{TR}_{i,h}$ and $a^{R}_{od,h,tt}$).

In the outer loop (blue boxes in Figure 4.2), all *Inner Loop Variables* are fixed; we optimize the *Outer Loop Variables* and the *Passenger Share Variables* by minimizing the objective function (4.1) subject to constraints (4.2)-(4.9) and (4.21)-(4.22). This updates the transit schedule, and the passenger decisions which are based on the resulting attractiveness of the different options. The inner loop (orange boxes in Figure 4.2) uses the resulting schedule to estimate traffic congestion at the equilibrium. In the inner loop, all *Outer Loop Variables* are fixed; we iteratively update the *Inner Loop Variables* and the ride-hailing market share, until convergence—that is, until the road travel times for ride-hailing vehicles and other vehicles vary by less than a pre-determined percentage from one iteration to another. The inner loop thus captures the impact of the transit schedule on the dynamics of the system (i.e., traffic flows, transit waiting times, travel costs, and ride-hailing fares). Then, we proceed to the outer loop, and re-optimize the transit schedule. This process continues until the transit schedule does not change between consecutive iterations of the outer loop.

This solution approach provides a decomposition algorithm for this large-scale MINLP, bringing three benefits. First, by treating the Inner Loop Variables as parameters in the outer-loop model, this transit schedule optimization sub-problem becomes a Mixed Inter Linear Program (MILP)—which is easier to solve than the originally MINLP. Second, it greatly reduces the size of the model by eliminating constraints (4.10)-(4.20) and (4.23)-(4.24). Third, the non-linear constraints (4.10)-(4.20) are captured in the inner loop, which is descriptive: by replicating the effects of the transit schedule on system dynamics, rather than optimizing transit schedules, the inner loop can be run in reasonable computational times for large-scale instances. Ultimately, this solution approach helps us transform the large-scale MINLP in a way that can be easily implemented without losing the interrelationships among the different components of the model.

4.3.1 Outer Loop

We initialize the travel times on all road segments to their corresponding free-flow values (i.e., those in the absence of congestion). Next, by treating all Inner Loop Variables as parameters, we optimize the public transit schedule by solving the model given by Equations (4.1)-(4.9) and (4.21)-(4.22) to optimality or until a pre-determined time limit. The algorithm terminates when two conditions have been met: 1) the outer loop optimization is solved to 0% optimality gap, and 2) the resulting solution is in equilibrium with the inner loop decision variable values. The latter condition is enforced by ensuring that the transit schedules obtained in two successive outer loop iterations are identical. A key idea in this algorithm, which accelerates its performance, is that initially, we do not attempt to solve the outer loop problem to full optimality. Instead, we apply a runtime budget of 1 hour. This allows the inner loop variables to come closer to their optimal values. Then, we progressively increase the outer loop runtime budget (in 1 hour increments), thus bringing all the decision variables closer to the overall optimum. This idea was found to be instrumental in reaching overall optimality more expeditiously, tremendously cutting down the overall runtime.

4.3.2 Inner Loop

We use a traffic assignment model to replicate traffic flows and travel times on all road segments. Outer loop optimization calculates the ride-hailing market share on each OD pair in each traffic timeslot. To assign ride-hailing vehicles to different travel paths across the road network, we use a User Equilibrium Model (UEM) defined from Wardrop's first principle—all routes used by travelers leaving from the same origin at the same time to the same destination have equal and minimal travel time [145]. It is formulated as follows

[150, 105]:

$$\text{Minimize} \quad \sum_{tt \in \mathcal{TT}'} \sum_{s \in \mathcal{SN}^{all}} \int_0^{z_s^{tt}} tra_{s,tt}(\tilde{z}_s^{tt}) d\tilde{z}_s^{tt} \tag{4.31}$$

$$\text{subject to:} \quad \sum_{p \in \mathcal{P}_{od}} flow_p^{tt'} = D_{od}^{tt'}, \quad \forall (o,d) \in \mathcal{Q},\ tt' \in \mathcal{TT}' \tag{4.32}$$

$$z_s^{tt} = \sum_{tt' \in \mathcal{TT}'} \sum_{(o,d) \in \mathcal{Q}} \sum_{p \in \mathcal{P}_{od}} (\delta_{p,s}^{tt',tt} \times flow_p^{tt'}), \quad \forall s \in \mathcal{SN}^{all},\ tt \in \mathcal{TT}' \tag{4.33}$$

$$flow_p^{tt'} \geq 0, \quad \forall p \in \mathcal{P}_{od}, (o,d) \in \mathcal{Q},\ tt' \in \mathcal{TT}' \tag{4.34}$$

$$tra_{s,tt} \geq 0, \quad s \in \mathcal{SN}^{all},\ tt \in \mathcal{TT}' \tag{4.35}$$

Here, z_s^{tt} is defined as the number of vehicles entering segment s during traffic timeslot tt. The function $tra_{s,tt}$ (a monotonic, continuously differentiable function) represents the travel time on road segment s in traffic timeslot tt, as a function of traffic flows z_s^{tt}. Next, $flow_p^{tt'}$ is the total number of vehicles taking path p and departing in traffic timeslot tt', $D_{od}^{tt'}$ is the total vehicle flow of OD pair (o,d) with departure time tt', and $\delta_{p,s}^{tt',tt}$ is given by:

$$\delta_{p,s}^{tt',tt} = \begin{cases} 1 & \text{if the vehicles taking path } p \text{ and departing in traffic timeslot } tt' \\ & \text{enter road segment } s \text{ during traffic timeslot } tt \\ 0 & \text{otherwise} \end{cases}$$

Equation (4.31) applies the Wardrop principle. Constraints (4.32) ensure that, for each OD pair, all vehicles get assigned to one of corresponding paths. Constraints (4.33) define traffic flows on each segment in each traffic timeslot. Constraints (4.34) and (4.35) define the domains of the variables.

In order to estimate $tra_{s,tt}(\tilde{z}_s^{tt})$, we use the speed-flow functions given by Schrank et al. [115], listed in Table 4.1. They are piece-wise linear continuous functions with slopes dependent on traffic volumes per lane and on the road type (freeway vs. arterial road).

Table 4.1: Segment travel speed as function of traffic volume per lane

Road Type	Congestion Level	Traffic Volume per Lane	Speed Equation
Freeway	Uncongested	Under 15,000	60
	Medium	15,001-17,500	70-(0.9× ADT/Lane)
	Heavy	17,501-20,000	78-(1.4× ADT/Lane)
	Severe	20,001-25,000	96-(2.3× ADT/ Lane)
	Extreme	Over 25,000	76-(1.46× ADT/Lane)
Arterial Street	Uncongested	Under 5,500	35
	Medium	5,501-7,000	33.58-(0.74× ADT/Lane)
	Heavy	7,001-8,500	33.80-(0.77× ADT/Lane)
	Severe	8,501-10,000	31.65-(0.51× ADT/Lane)
	Extreme	Over 10,000	32.57-(0.62× ADT/Lane)

Note. ADT/Lane is average daily traffic per lane, in thousands.

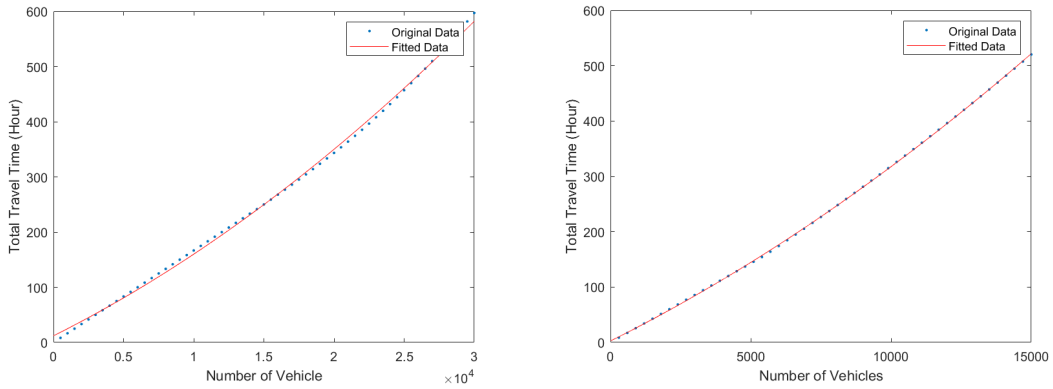
The UEM problem has a continuously differentiable objective function, and could thus be solved by the method of Frank and Wolf [57]—by iteratively solving a linear program with an approximated objective function and a line search, until convergence. One major challenge is that this technique requires significant customization and large computer memory to store path-flows. As an alternative, we propose a faster numerical approach, by approximating the objective function with polynomials—using the least-squares method. A polynomial of order 2 was found to strike a good balance between accuracy and simplicity. We obtained R^2 values of 99.99% (3 parameters fitted to 30,000 data points) for freeways and 99.88% for arterial streets (3 parameters fitted to 15,000 data points). Figure 4.3 demonstrates the accuracy of our approximation approach by plotting the original objective function and the fitted quadratic approximation together for freeways (Figure 4.3a) and arterial streets (Figure 4.3b), each for a one-lane segment of 1 mile length.

Ultimately, this approach replaces the traffic flow assignment model by the following problem:

$$\text{Minimize} \quad \sum_{s \in \mathcal{SN}^{all}} LE_s(\alpha_s z_s^{tt} \times z_s^{tt} + \beta_s z_s^{tt} + \gamma_s) \tag{4.36}$$

$$\text{subject to:} \quad \text{Constraints } (4.32) - (4.35), \tag{4.37}$$

where α_s, β_s, γ_s are fitted parameters for road segment s, determined by the road type and



(a) Freeways (b) Arterial Street

Figure 4.3: Quadratic approximation of the UEM objective function

number of lanes of that road segment. This approximate flow assignment model was solved to 0% optimality gap in 10 seconds in our largest case study—thus enabling tractability of the overall model. Upon solving the traffic assignment problem, we update the travel time values on each segment for each traffic timeslot, and update the Inner Loop Variables accordingly.

4.4 Case Study

We use the Manhattan road network and the New York Subway system as a case study to demonstrate the effectiveness of our model and solution approach.

4.4.1 Experimental Setup

We consider the road network of Manhattan, consisting of 4,092 nodes and 9,453 road segments. We are interested in optimizing the transit schedules during the 6:30–9:30 am rush hours—our optimization horizon. To capture boundary effects immediately before and after the rush hours, our modeling horizon is set to 5:30–10:30 am. We divide this five-hour modeling horizon into 20 15-minutes traffic timeslots. We assume that all other vehicles except ride-hailing vehicles follow their original travel paths; so our traffic assignment

model only captures ride-hailing demand. This assumption can be easily relaxed, if one has access to accurate OD-level data for other vehicles. All data on road traffic are taken from Donovan and Work [49], Bertsimas et al. [24] and Moss et al. [101]. The procedures used for estimating passenger demand of each OD pair and the number of other vehicles on each road segment in each traffic timeslot are detailed in Appendix B.2.

We obtain all relevant information on each subway line, including length, stops, actual schedule, passenger-carrying capacity, vehicle compatibility, vehicle availability, and fare, from the MTA [102]. We divide the five-hour modeling horizon into 20 15-minutes transit timeslots. MTA buses are not included in the model but considered as part of the outside option.

We tune the utility coefficients to generate the attractiveness values ($A^O_{od,h}$, $a^{TR}_{i,h}$ and $a^R_{od,h,tt}$) for each transport mode. To this end, we use an iterative procedure to match the actual market share of each transportation mode in New York City reported in Trottenberg [138]. The calibration details are reported in Appendix B.3. Given the blurring of the boundaries between ride-hailing model and traditional taxicabs, we combine ride-hailing and taxis into the same category (as done in New York City Mobility report [138]).

We use six passenger types in our computational experiments, corresponding to a combination of one of the two trip purposes (price-sensitive vs. time-sensitive passengers) and one of the three ideal departure time preferences (7 am, 8 am and 9 am). Passenger type definitions are detailed in Appendix B.4. All other parameters are described in Appendix B.5.

Note that this computational study focuses predominantly on Manhattan, but we take precautions to capture all relevant interdependencies with other boroughs. Specifically, we optimize the entire transit schedule corresponding to all subway lines that are at least partly in Manhattan. Also, we include passenger demand with part of the trip (origin, destination, candidate travel paths, or candidate transit itineraries) possibly in Manhattan. Finally, we model the entire road network of all five boroughs of New York City but we model the

road congestion effects of transit scheduling changes only in Manhattan—for the other boroughs, we fix the congestion levels.

4.4.2 Computational Results

We now demonstrate the benefits of our modeling and solution approach across multiple demand distributions and different assumptions regarding the required minimum level of transit service.

Demand Distribution Approaches: Our first, and most straightforward, approach (we name it *"Default"*) for calculating the geographical distribution of passenger demand assumes that it is the same as the distribution of taxi trips. Although simple, this distribution may be biased for two reasons. First, bridges and tunnels connecting Manhattan to other boroughs have tolls which are included in ride-hailing fare, so the demand distribution estimated from taxi data may be undercounting inter-borough trip demand. In our second approach (we name it *"Borough"*), we drastically compensate for this effect by increasing demand of all inter-borough ODs by 50% and decreasing the demand for all intra-borough ODs by a fixed percentage, while keeping the total network-wide demand constant. Second, the New York City subway fare does not depend on trip distance but the taxi/ride-hailing fare does, so the demand distribution estimated from taxi data may be undercounting longer-distance trips. In our third approach (we name it *"Distance"*), we use another drastic compensation scheme by increasing the demand for the long-distance ODs by 50% and decreasing that for the remaining ODs by 50% (we define long-distance and short-distance OD pairs based on their shortest path on the road network, so that total demand is equal among the two groups). The goal is to test the robustness of our findings to the underlying demand distribution.

Transit Service Level Scenarios: As mentioned in Section 4.2.3, different transit operators may have different preferences regarding the minimum required transit service level and/or the allowable deviation from the existing schedule. So, we evaluate our model and

solution algorithm under three scenarios. All three scenarios enforce Constraints (4.27) with the minimum frequency $MinFre_{g,t}$ set to 1 for all transit lines and all transit timeslots. The first scenario (we name it *"Miles"*) also enforces Constraints (4.29) and (4.30), so that the transit vehicle-miles do not deviate from the current ones by more than 5%. The second scenario (we name it *"Frequency"*) enforces Constraints (4.28) instead of Constraints (4.29) and (4.30), so that the total frequency across the rush hours does not drop more than 25% below the current frequency for any transit line. The third scenario (we name it *"Neither"*) imposes none of these constraints beyond Constraints (4.27).

Results: For each of the nine demand and transit level scenario combinations, the total computational time lies between 8 and 15 hours. These runtimes are consistent with the strategic transit planning decisions considered in our modeling framework.

Table 4.2 summarize the main practical findings from this section. Column 4 of the table lists the reduction in the annualized system-wide cost (in $ million) in the morning rush hours resulting from our approach when compared with the actual transit schedule. The next four columns break it down into four categories: passengers, service providers, other vehicles, and emissions costs (the only type of external cost considered in our computational experiments). Positive (resp., negative) values indicate decrease (resp., increase) in costs. Figure 4.4 shows this information graphically.

Table 4.2: Breakdown of the annualized cost reductions (in USD $M) across nine scenarios

Scen. Comb.	Dem. Distr.	Service Level	Total	To Passengers	To Operators	To Other Veh.	Emissions
S1	Default	Miles	408.4	-5.6	414.0	-3.0	3.1
S2	Default	Frequency	1,275.3	251.3	1,024.0	-3.2	1.8
S3	Default	Neither	1,756.1	307.9	1,453.4	-6.9	3.3
S4	Borough	Miles	329.2	276.5	56.8	1.7	-5.8
S5	Borough	Frequency	852.2	347.5	509.9	0.9	-6.1
S6	Borough	Neither	1,720.5	555.9	1,166.7	-1.7	-0.4
S7	Distance	Miles	564.3	292.9	272.2	-0.7	-0.2
S8	Distance	Frequency	1,092.4	370.9	723.5	-1.5	-0.5
S9	Distance	Neither	1,617.0	458.8	1,165.6	-5.3	-2.2

Note. Scen. Comb.: Scenario Combination, Dem. Distr.: Demand Distribution, Veh.: Vehicles.

Note that, irrespective of demand distribution and service level scenario, the optimized transit schedule leads to a significant system-wide cost reduction—ranging from $329.2

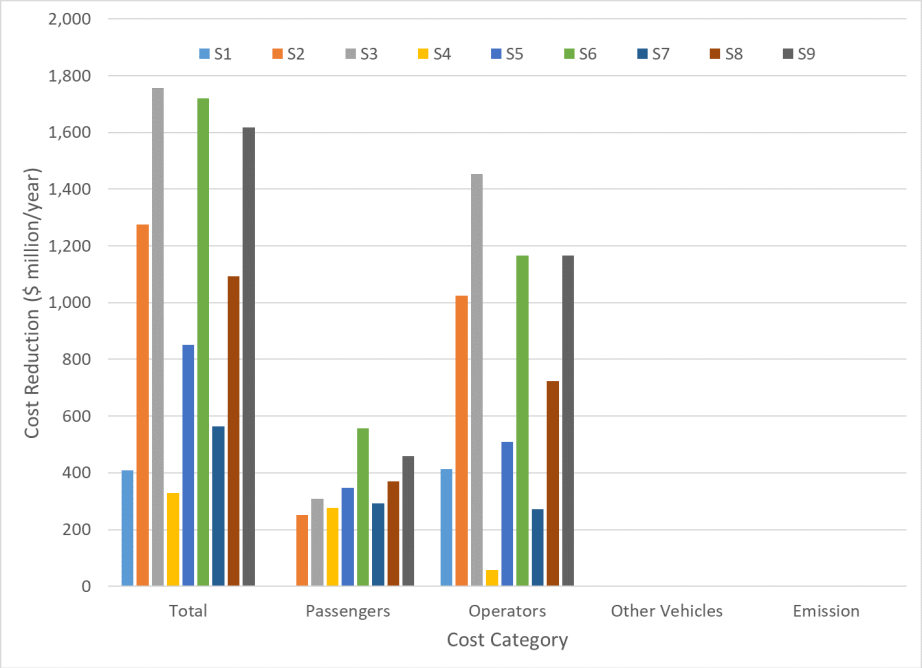


Figure 4.4: Breakdown of annualized cost reduction by category

M/year to $1,756.1 M/year. A major driver of this cost reduction is a considerable reduction in cost to the passengers. In S1, the passenger cost remains essentially unchanged (a $\sim$0.01% change); in the other eight scenario combinations, the passenger cost reduction ranges from $251.3 M/year $555.9 M/year. Another significant driver of the cost reduction is the reduction in the cost to the operators (transit and ride-hailing)—ranging from $56.8 M/year to $1,453.4 M/year. Finally, the costs to other vehicles and the emissions costs both remain essentially unchanged—with an average decrease of -$2.2 M/year and -$0.8 M/year, respectively. Note that this should not be misconstrued as equivalent to saying that other vehicles' costs and emission costs can be ignored in the optimization. This is a subtle but important point, highlighted in Section 4.4.3.

Additional details of the solutions are provided in Appendix B.6 (Tables B.5-B.4). In particular, these results highlight that our solutions do not significantly modify each mode's market shares. On average, the transit market share decreases by only 1.4% from the actual to the optimal schedule, and the market shares of ride-hailing and of the outside option increase +0.16% and +1.24%, respectively. Stated differently, our approach achieves a

significant cost reduction for multiple stakeholders without much change in the aggregate mode shares.

Ultimately, our solution leads to consistently and significantly large reductions in the system-wide costs (by $0.4\% - 2.5\%$), the costs to passengers (by $0.0\% - 0.9\%$) and the costs to operators (by $1.6\% - 40.85\%$), while keeping the costs to the other vehicles and the emissions costs unchanged. This means that adapting transit frequencies and schedules can result in win-win-win outcomes for the passengers, service-operators and the urban transportation system as a whole—without causing additional harm to the environment or to the other vehicles on the road.

In the remainder of this chapter, we focus on the distance-based demand adjustment (scenario combinations S7, S8 and S9). However, the main takeaways remain essentially unchanged when similar analyses are repeated under other scenario combinations.

4.4.3 Benefits of Our Comprehensive Modeling Approach

We now evaluate the benefits of explicitly modeling passengers' decisions and road traffic congestion in the transit schedule optimization. To do so, we create two baselines. The first one (*Ignore Congestion*) ignores the effects on traffic congestion, by fixing the road travel time on each segment to that under the actual schedule. The second one (*Ignore Passenger Choice*) ignores passenger choice considerations in the outer loop optimization, by fixing the transit market share in each OD market to that under the actual schedule. We evaluate each baseline by running our full solution approach after fixing the transit schedule—so the baselines ignore passenger choice or road congestion in the schedule optimization, but the schedules are assessed under both passenger choice and road congestion.The results (in scenario combination S9) are reported in Table 4.3, where the numbers in parentheses indicate the relative change, as compared to the *Actual* schedule.

The optimal schedule reduces total system-wide costs by approximately $4.4 million per day. Table 4.3 shows that 7% and 31% of this reduction would be washed away when

Table 4.3: Benefits assessment of individual components of our optimization model

	Actual	Ignore Passenger Choice	Ignore Congestion	Optimized Schedule
Total Cost ($)	175,497,569 (0.00%)	172,439,923 (-1.74%)	171,358,632 (-2.36%)	171,067,522 (-2.52%)
Total Cost to Passengers ($)	141,685,551 (0.00%)	140,289,171 (-0.99%)	141,200,587 (-0.34%)	140,428,436 (-0.89%)
Cost to Transit Passengers ($)	10,975,221 (0.00%)	9,435,050 (-14.03%)	8,047,640 (-26.67%)	9,761,367 (-11.06%)
Cost to Ride-hailing Passengers ($)	11,141,282 (0.00%)	11,568,298 (3.83%)	11,533,091 (3.52%)	11,731,933 (5.30%)
Cost to Outside Option Passengers ($)	119,569,047 (0.00%)	119,285,823 (-0.24%)	121,619,856 (1.72%)	118,935,136 (-0.53%)
Total Cost to Transit and Ride-hailing Operators ($)	8,833,828 (0.00%)	7,157,791 (-18.97%)	5,166,537 (-41.51%)	5,640,334 (-36.15%)
Cost to Transit Operator ($)	7,251,348 (0.00%)	5,536,175 (-23.56%)	3,549,253 (-51.05%)	4,001,311 (-44.82%)
Cost to Ride-hailing Operator ($)	1,582,480 (0.00%)	1,621,616 (2.47%)	1,617,284 (2.20%)	1,639,023 (3.57%)
Total Cost to Other Vehicles ($)	24,528,097 (0.00%)	24,538,926 (0.04%)	24,537,879 (0.04%)	24,542,732 (0.06%)
Time Cost to Other Vehicles ($)	22,932,078 (0.00%)	22,933,229 (0.01%)	22,933,178 (0.00%)	22,933,009 (0.00%)
Operating Cost to Other Vehicles ($)	1,596,019 (0.00%)	1,605,697 (0.61%)	1,604,701 (0.54%)	1,609,723 (0.86%)
Total Emission Cost ($)	450,092 (0.00%)	454,036 (0.88%)	453,629 (0.79%)	456,020 (1.32%)
Cost of Emissions from Ride-hailing Vehicles ($)	50,278 (0.00%)	51,797 (3.02%)	51,640 (2.71%)	52,773 (4.96%)
Cost of Emissions from Other Vehicles ($)	399,814 (0.00%)	402,239 (0.61%)	401,989 (0.54%)	403,247 (0.86%)
Market Shares				
Transit Market Share	34.84%	34.10%	30.04%	32.05%
Ride-hailing Market Share	5.11%	5.37%	5.34%	5.46%
Outside Option Market Share	60.05%	60.53%	64.61%	62.49%

ignoring congestion and passenger choice, respectively. This underscores the value of explicitly accounting for congestion and passenger choice in our modeling and computational framework.

Diving further into the results in Table 4.3, note that the biggest cost difference between *Ignore Passenger Choice* and *Optimized Schedule* pertains to the transit operator. By ignoring the alternative transportation modes available to the passengers, we force the optimization model to accommodate a fixed number of passengers in each OD pair. This leads to a significantly suboptimal transit schedule, thus causing unnecessary increase in transit operating costs ($\sim$27%)—much larger than the corresponding increase in transit market share ($\sim$6%).

The situation under *Ignore Congestion* is completely different but equally interesting. In that baseline, many more passengers switch to ride-hailing (since congestion costs are

underestimated), enabling the transit operator to reduce transit service to an even lower level than in the optimal solution. This lower level of transit service, in actuality, just pushes more passengers to the outside option, increasing the total cost to outside option passengers by $2.7 million per day. This is the biggest factor contributing to the suboptimality of the schedule obtained under *Ignore Congestion*.

4.4.4 Detailed Comparison with Actual Schedule

We now compare the optimized schedule to the actual schedule (in scenario combination S8) to identify the main improvements provided by our schedule. We report the service frequencies in Figure 4.5 for operations starting in the rush hours. In this figure, we cluster the lines which share the same primary trunk line, and separate south-bound vs. north-bound operations.

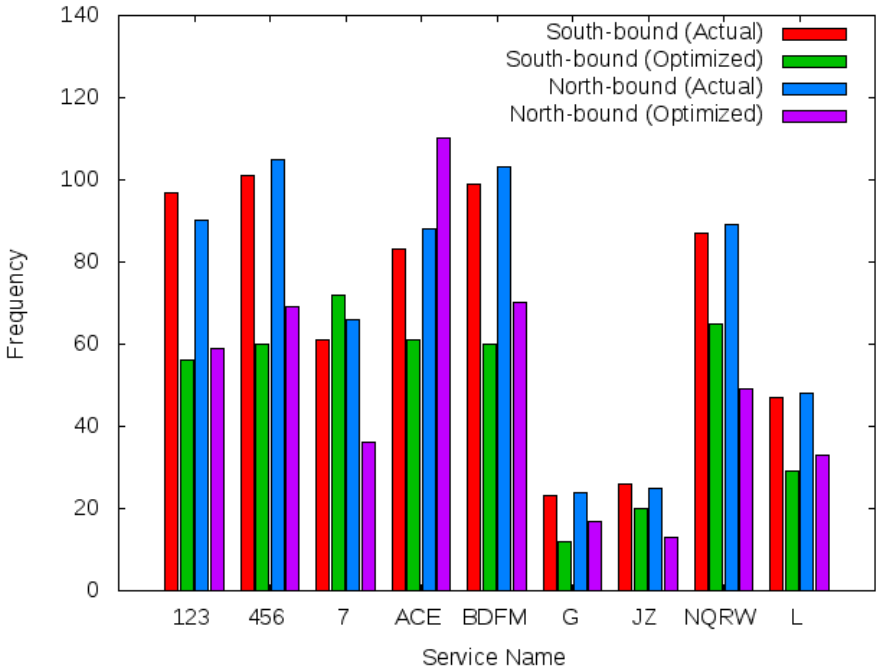


Figure 4.5: Frequency by line

Figure 4.5 suggests that service frequencies are reduced on most lines. But there are two notable exceptions: frequency increases on line "7" South-bound and on lines "ACE"

North-bound—two lines originating in Queens and then passing through Manhattan. This is likely an effect of the strong morning demand from Queens toward Manhattan—many commuters reside in Queens and work in Manhattan and other boroughs. This also stems from the fact that ride-hailing services are not particularly attractive to travel from Queens to Manhattan, due to traffic congestion on the corresponding paths, long travel distances and high fares. In summary, the optimized schedule reduces service frequency on many lines but increases on others—thus spatially reconfiguring transit supply to match passenger demand, given ride-hailing alternatives and traffic congestion patterns.

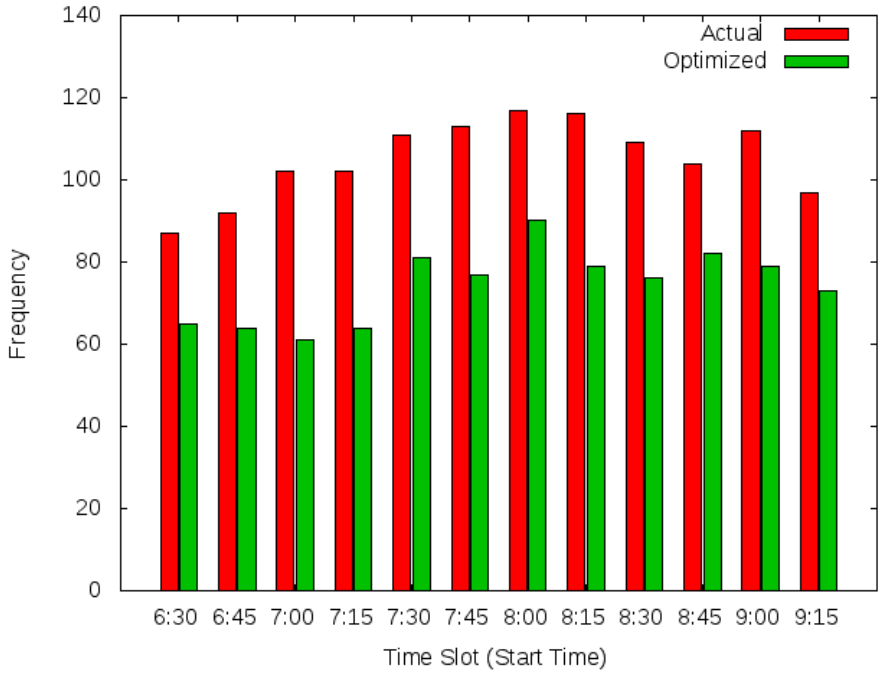


Figure 4.6: Frequency by time slot

Next, Figure 4.6 plots total service frequencies across all transit lines, by line departure time. The main observation is that the temporal patterns are consistent between the actual and optimal schedules. In the actual schedule, the frequency is distributed quite evenly between $6:30$ and $9:30$ with a peak around 8:00-8:15 corresponding to 35% higher frequency than in the lowest-frequency timeslot. The same pattern holds under our optimized schedule with a peak around 8:00-8:15 corresponding to 38% higher frequency than in the

lowest-frequency timeslot. Moreover, both schedules have a slightly higher frequency at the end of the rush hour than at the start.

Last, we compare the passenger choice results under the optimized and actual transit schedules as shown in Figure 4.7 (in scenario combination S7). We classify the OD pairs along two dimensions: travel distance and congestion level. We divide OD pairs into long-distance and short-distance trips based on their shortest path on the road network, so that total demand is equal among the two groups. Similarly, we divide the OD pairs into "Congested" (top 50%) and "Less Congested" (bottom 50%) based on the average vehicle-to-capacity ratios across all travel paths connecting the OD pair.

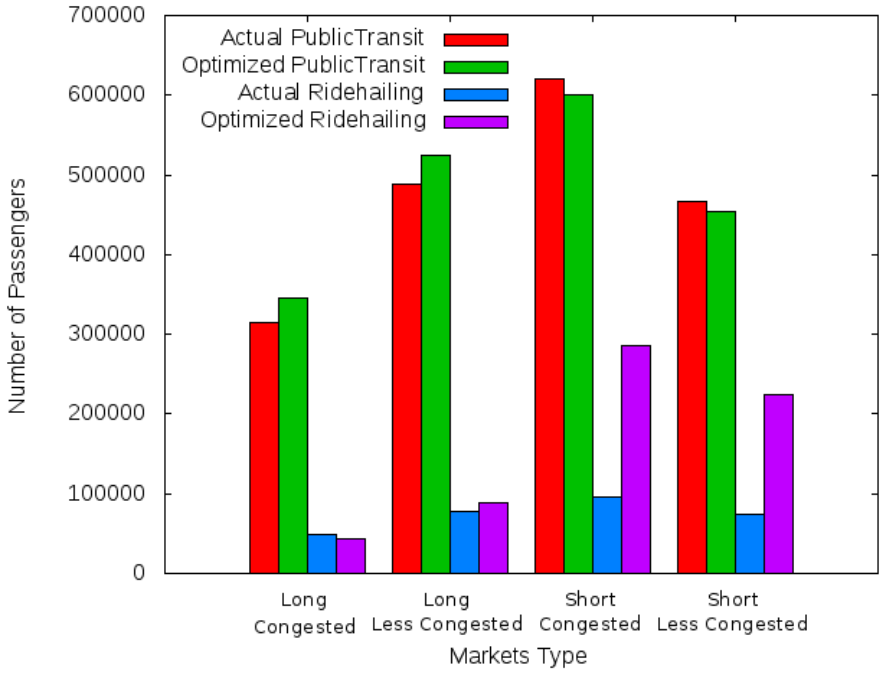


Figure 4.7: Passenger choice comparison between actual schedule and optimal schedule

Note that the number of transit passengers increases in "Long" markets and slightly decreases in "Short" markets. Thus, the optimal transit schedule encourages passengers to take transit for long-distance trips and slightly discourages them from taking transit for short-distance trips. At the same time, the number of short-distance passengers taking ride-hailing increases significantly. Among the longer-distance markets, the number of

ride-hailing passengers decreases slightly in congested markets, and increases slightly in less congested markets. Thus our model, which explicitly captures congestion, encourages more passengers to take ride-hailing in shorter markets, while also discouraging them from taking ride-hailing in the longer-distance and congested markets.

In summary, our model results in a better matching between passengers and suitable travel modes. For example, public transit is especially suitable for the longer OD markets in New York City where the subway system provides a high level of access and cheap fares. In contrast, ride-hailing is more attractive in the shorter and less congested markets, allowing passengers to avoid the comparatively long transit access and egress times at a more moderate price premium.

4.5 Robustness Tests

The results presented in Section 4.4 relied on a number of assumptions. An important one is that we only considered the ride-hailing vehicle-miles that are traveled with passengers—thus ignoring empty vehicle driving between trips and underestimating congestion induced by ride-hailing. Another assumption is that the transit fare structure remains the same as that currently in place in New York City. We now relax these two assumptions by investigating the impact of empty vehicle driving and different transit fare structures on the optimal transit schedule. We also establish the robustness of our findings, by showing that the main benefits identified in Section 4.4 hold even if some system complexities are ignored when optimizing the public transit schedule.

4.5.1 Empty Ride-hailing Vehicles

Following passenger drop offs, ride-hailing vehicle may drive empty—without passengers—while waiting for the next trip request and driving toward the next pickup location. Henao and Marshall [68] estimated that empty vehicle driving increases average vehicle-miles

traveled in ride-sharing systems by 69%. Obviously, there is significant uncertainty regarding the exact paths along which empty vehicle driving occurs—depending on driver decisions, drop-off locations, traffic conditions, etc. We therefore use an approximate method (detailed in Appendix B.7) to reconstruct the paths of empty ride-hailing vehicles, while ensuring that empty driving amounts to 69% of the actual distance driven with passengers. This approach should not be overly interpreted at the disaggregate level (e.g., on each road segment, at each node, in each traffic timeslot). Still, by accounting for the aggregate effects of empty driving, this algorithm helps paint a more accurate picture of overall congestion caused by the ride-hailing vehicles.

Table 4.4 lists the results with and without empty driving (in scenario combination S9). The second column reports the results from Section 4.4, obtained and evaluated without empty driving. The next column uses the actual schedule and evaluates it with empty vehicle driving. The next one uses the transit schedule obtained without empty vehicle driving (as in column 2) but evaluates it with empty driving—by recomputing the traffic assignment accordingly. The last column re-optimizes the transit schedule and evaluates it while explicitly accounting for empty vehicle driving.

Table 4.4: Comparison of results with and without empty driving

	Without Empty Driving	With Empty Driving		
		Actual	Optimized Schedule	Re-optimized Schedule
Total Cost ($)	171,067,522	175,497,616	172,258,760	172,255,435
Total Cost to Passengers ($)	140,428,436	141,685,530	141,412,110	141,402,620
Cost to Transit Passengers ($)	9,761,367	10,975,182	9,838,755	9,869,186
Cost to Ride-hailing Passengers ($)	11,731,933	11,141,321	11,206,571	11,206,232
Cost to Outside Option Passengers ($)	118,935,136	119,569,027	120,366,784	120,327,202
Total Cost to Transit and Ride-hailing Operators ($)	5,640,334	8,833,835	5,655,812	5,662,031
Cost to Transit Operator ($)	4,001,311	7,251,349	4,001,311	4,007,675
Cost to Ride-hailing Operator ($)	1,639,023	1,596,020	1,654,501	1,654,356
Total Cost to Other Vehicles ($)	24,542,732	24,528,157	24,710,287	24,710,241
Time Cost to Other Vehicles ($)	22,933,009	22,932,138	22,970,595	22,970,572
Operating Cost to Other Vehicles ($)	1,609,723	1,596,020	1,739,692	1,739,669
Total Emission Cost ($)	456,020	450,093	480,551	480,543
Cost of Emissions from Ride-hailing Vehicles ($)	52,773	50,278	44,745	44,743
Cost of Emissions from Other Vehicles ($)	403,247	399,815	435,806	435,800
Market Shares				
Transit Market Share	32.05%	34.84%	32.30%	32.34%
Ride-hailing Market Share	5.46%	5.11%	5.33%	5.33%
Outside Option Market Share	62.49%	60.05%	62.37%	62.33%

The results show that empty driving, evaluated using our approximate approach, leads

to a daily cost increase of around $\$1.2$ M. That is partly due to the additional $\$170,000$ of cost to other vehicles due to increased traffic congestion. In addition, this extra congestion lowers the ride-hailing market share from 5.46% to 5.33%. This effect results in a significant increase in the cost to passengers (by $\sim$ \$1 M per day), primarily driven by a large increase in the costs to the passengers displaced from ride-hailing to the outside option.

More importantly, the cost values are similar in the last two columns of the table. So even transit schedules optimized without considering empty driving end up being close to optimal in the presence of empty driving. Also, when we compare the *Actual* and *Optimized Schedule* results, the main takeaways from Section 4.4 still hold: the system-wide costs decrease significantly (by \$1,182.18 M/year in this case) and this improvement is driven primarily by a cost reduction to the passengers (\$99.80 M/year) and to the operators (\$1,159.58 M/year). These results establish the robustness of our modeling and computational approach, showing that our optimized schedule provides win-win-win outcome for passengers, operators and the system as a whole even when failing to account for some real-world complexities (e.g., additional congestion caused by empty vehicle driving, here).

4.5.2 Fare Structures

We now relax the assumption that transit fares follow the current structure in New York City—a flat \$2.75 per trip. First, we vary the public transit fare with trip length (*Trip-Length Based Fare*)—a common practice in many cities. Specifically, we set the fare to $\$2$ for trips of 4 stops or less, $\$4$ for trips of $5-10$ stops, and $\$6$ for trips of more than 10 stops. Second, we consider an extreme scenario where public transit is offered at no charge (*Free Transit*)—as currently experimented in a few cities worldwide (e.g., Chapel Hill, Dunkirk, Tallinn). The results are reported in Table 4.5 (in scenario combination S9), using the same structure as in Table 4.4. For each transit fare structure, we report the costs using (i) the actual schedule evaluated under the new fare structure, (ii) the optimal schedule corresponding to a flat \$2.75 fare evaluated under the new fare structure, and (iii)

the optimal schedule obtained and evaluated under the new fare structure.

Table 4.5: Comparison of results under different fare structures

	Without Fare Change	Trip-Length Based Fare			Free Transit		
		Actual	Optimized	Re-optimized	Actual	Optimized	Re-optimized
Total Cost ($)	171,067,522	175,618,653	171,060,431	170,986,099	175,454,936	170,725,850	170,701,059
Total Cost to Passengers ($)	140,428,436	141,801,687	140,449,957	140,397,801	141,648,421	140,123,509	140,079,834
Cost to Transit Passengers ($)	9,761,367	10,918,078	9,632,337	9,571,563	11,267,431	9,814,530	9,865,998
Cost to Ride-hailing Passengers ($)	11,731,933	11,170,627	11,542,211	11,538,664	11,070,577	11,460,147	11,462,579
Cost to Outside Option Passengers ($)	118,935,136	119,712,982	119,275,409	119,287,574	119,310,413	118,848,832	118,751,257
Total Cost to Transit and Ride-hailing Operators ($)	5,640,334	8,837,574	5,618,251	5,596,268	8,830,009	5,612,452	5,631,235
Cost to Transit Operator ($)	4,001,311	7,251,349	4,001,311	3,979,437	7,251,349	4,001,311	4,020,003
Cost to Ride-hailing Operator ($)	1,639,023	1,586,225	1,616,940	1,616,831	1,578,660	1,611,141	1,611,232
Total Cost to Other Vehicles ($)	24,542,732	24,528,926	24,538,401	24,538,332	24,526,764	24,536,718	24,536,801
Time Cost to Other Vehicles ($)	22,933,009	22,932,126	22,933,300	22,933,301	22,931,842	22,933,051	22,933,078
Operating Cost to Other Vehicles ($)	1,609,723	1,596,799	1,605,101	1,605,031	1,594,922	1,603,667	1,603,723
Total Emission Cost ($)	456,020	450,465	453,822	453,697	449,742	453,170	453,188
Cost of Emissions from Ride-hailing Vehicles ($)	52,773	50,455	51,732	51,625	50,203	51,440	51,444
Cost of Emissions from Other Vehicles ($)	403,247	400,010	402,090	402,072	399,540	401,730	401,744
Market Shares							
Transit Market Share	32.05%	34.74%	31.57%	31.49%	35.74%	32.32%	32.38%
Ride-hailing Market Share	5.46%	5.13%	5.35%	5.35%	5.06%	5.30%	5.30%
Outside Option Market Share	62.49%	60.13%	63.09%	63.16%	59.20%	62.38%	62.32%

First, the comparison of the *Without Fare Change* column with both the *Optimized* columns shows that the changes are as expected. The trip-length based fare leads to a slight reduction in transit market share (by 0.5%) as the new structure induces a slightly cheaper fare on short trips, but a significantly more expensive fare on longer trips. Short-distance passengers get displaced from ride-hailing to transit, causing a tiny reduction (0.1%) in ride-hailing market share—and resulting in lower congestion and lower costs related to ride-hailing and emissions. But many more longer-distance passengers get displaced from transit to the outside option, causing a more significant increase (0.6%) in the outside option market share—and resulting in higher costs to other vehicles. These two effects balance out and lead to the overall costs staying essentially unchanged. This fare structure change also leads to an increase in transit revenue of $436 M/year.

A similar comparison shows that, as expected, *Free Transit* increases the transit market share at the expense of ride-hailing and outside option. This leads to small reductions in all four cost categories: cost to passengers, operators, and other vehicles, as well as emission cost, summing to a total reduction of $125 million/year. Obviously, this comes at the expense of a significant reduction in transit revenues, of the order of $1,710 million/year.

Most importantly, the comparison of the *Actual* and *Optimized* schedules under both

fare structures shows the robustness of our approach. Indeed, the *Optimized* schedule leads to a significantly lower total system-wide cost (by $4.6–4.7 M/day) than the *Actual* schedule, driven by lower costs to passengers (by $1.4–1.5 M/day) and costs to operators (by $3.2 M/day). This confirms that even if the transit fare structure changes drastically after we optimized the transit schedules, the same win-win-win outcomes can be expected relative to the existing transit schedule.

4.6 Summary

The increasing availability of on-demand transportation options creates opportunities to enhance urban mobility, but also contributes to urban planning challenges—including declining public transit ridership and increasing traffic congestion. This research proposes to redesign public transit schedules to address these challenges. To this end, we have developed an integrated modeling framework that optimizes public transit frequencies and timetables, while capturing passengers' endogenous mode choices under ride-hailing competition and road congestion. This problem is formulated as a Mixed Integer Nonlinear Program (MINLP). To solve it, we have developed a bi-level solution algorithm that iteratively optimizes public transit schedules and estimates passengers' mode choices and traffic congestion, until convergence.

Case study results in New York City suggest that optimized transit schedules can reduce daily system-wide transportation costs by millions of dollars. Importantly, this cost reduction is shared by urban transportation stakeholders—resulting in win-win-win outcomes for passengers, transportation service providers, and the system as a whole. These benefits are primarily achieved by re-allocating public transit resources away from OD pairs that are well served by alternative transportation modes (e.g., short trips on mildly congested roadways) toward OD pairs where public transit has a stronger edge over road transportation (e.g., longer trips on more congested roadways). Ultimately, these findings suggest

that public transit and ride-hailing can co-exist and provide complementary services in a way that enhances mobility across the urban ecosystem.

Chapter 5

Conclusions and Future Directions

In this thesis, we develop and solve optimization models for transportation planning incorporating the uncertainty in supply and demand faced by airline industry and urban transportation systems. In this chapter, we conclude with a detailed summary of the contributions and future research directions.

5.1 Chapter 2: Modeling Crew Itineraries and Delays in the National Air Transportation System

This chapter explains the complex reality of crew scheduling, and provides an inside peak at the techniques used by airlines to absorb system-wide delay propagation. It achieves this by developing optimization and estimation methods that are shown to successfully solve the inverse of one of the most challenging integer programming problems, for airline networks that are among the largest in the world.

From a methodological standpoint, for the first time, the inverse of the robust crew pairing generation problem is presented, formulated and solved in order to gain insights into the extent of robustness of real-world airline scheduling practices. The problem is formulated as one of learning the parameters of the robust optimization objective function

using real-world airline crew scheduling samples. A heuristic solution approach is developed and implemented. It involves solving the forward problem repeatedly to minimize a similarity measure between the solution of the robust crew pairing problem and the actual airline crew schedule samples by identifying the optimal set of objective function parameters. The forward problem minimizes the sum of planned crew cost and the penalty costs which penalize the crew pairings for six different features that make them vulnerable to the propagation of delays and disruptions. A sequence of exact methods and heuristic ideas is used to solve this robust crew pairing problem to near-optimality. This allows the overall parameter estimation problem to be solved in a reasonable amount of time.

From a practical standpoint, accurate estimation of crew delays is critical for an overall understanding of aviation system performance and to inform government policy and airline carrier decisions. This chapter demonstrates that sophisticated crew scheduling practices allow airlines to avoid 60% to 80% of crew-propagated delays. Furthermore, we find that the crew pairings estimated using four different airline networks perform similar to each other, and much better than the deterministic crew pairing solutions, in terms of their closeness to the actual crew schedules, even when the estimation and evaluation is not conducted on the same network. This suggests that the estimated parameters and the overall estimation approach are relatively stable in the face of changes in data availability. Finally, this research demonstrates that the ratio of the penalty costs (representing the costs of the crew-propagated delays and disruptions) and the crew salary costs consistently lies between 0.5% and 4% providing a measure of the relative importance placed by the airlines on planned costs versus the costs of ensuring robustness to disruptions.

In addition to these contributions, this research makes the estimated crew pairing solutions available for further research and analysis. We have made this entire model calibration code as well as the resulting calibrated crew pairing solutions publicly available for future research. These estimated crew pairing solutions are useful to gauge the extent of delays and disruptions that propagate across the airline networks through crew connections. While

these crew pairing estimates do give a starting point to estimate the crew-propagated delays and disruptions, the important next step toward accurately estimating historical delay propagation is to develop a similar understanding of the crew recovery strategies used by the airlines in the real world. Once we have access to a historical sample of actual crew recovery actions, a framework similar to the one developed in this chapter could be used to learn the airline crew recovery optimization process as well. This will be a valuable next step in this research stream.

5.2 Chapter 3: Airline Timetable Development and Fleet Assignment Incorporating Passenger Choice

This chapter optimizes the airline timetable by explicitly capturing the endogeneity of fleet assignment decisions and passenger booking decisions. It achieves this by developing a mixed-integer programming model and a multi-phase solution framework that are shown to successfully optimize airline timetables by integrating a model of passenger choice that approximately reflects the airline revenue management practices.

From a methodological standpoint, we incorporate a sales-based linear programming framework to accurately capture the itinerary-level demand substitution effects into a comprehensive timetable development and fleet assignment optimization model. Specifically, passenger choice is captured by a discrete-choice Generalized Attraction Model, defined in the constraint requiring the market share of each itinerary-fare class combination to be proportional to its attractiveness. This constraint embeds a linearized version of a discrete-choice model of passenger booking decisions. To our knowledge, ours is the first research study to capture an airline's comprehensive timetable development problem under passenger choice. Also, we design an effective multi-phase solution approach to solve this large-scale mixed-integer linear programming problem. When combined with several variable-fixing and symmetry-inducing accelerated heuristics, we can achieve high quality solutions

within reasonable computational times by narrowing down the flight's departure time range step-by-step.

From a practical standpoint, we use a case study from a major hub-and-spoke airline carrier to demonstrate the benefits of our approach. When compared to previous methods that only consider incremental changes to existing timetables and typically ignore passenger choice, our approach results in between 15% and 40% profit improvements. Also, our computational results for a real-world airline merger case study show that this combination can lead to 10% additional profits by jointly re-optimizing the timetables and fleet assignments.

In future research, our framework can be extended to integrate other airline planning considerations such as route development, aircraft maintenance routing, and crew scheduling. In addition, even though our accelerated heuristic strategies considerably improve the computational performance, an optimality gap may still exist, and very large airline networks (such as those of the mega-carriers created by the mergers of the major U.S. carriers in the last decade) may still be impossible to solve within acceptable computational runtimes. This motivates further research into even faster heuristics or exact methods to handle these extremely large problem instances.

5.3 Chapter 4: Transit Planning Optimization under Ride-hailing Competition and Traffic Congestion

This chapter develops an original integrated model that optimizes transit frequencies and timetables under endogenous passenger mode choice shaped by ride-hailing service offerings and road traffic congestion. The problem is formulated as a mixed-integer nonlinear optimization model, and solved using a new bi-level solution algorithm that iteratively optimizes public transit schedules and estimates passenger mode choices and traffic congestion, until convergence.

From a methodological standpoint, we contribute a new highly nonlinear model as well as a bi-level solution approach to solve it in a reasonable computational time. The solution approach includes two interconnected procedures. An outer loop optimizes public transit schedules, given travelers' mode choices and estimates of traffic congestion on road networks. Given a transit schedule, an inner loop leverages a user-equilibrium traffic flow assignment model to replicate the paths of ride-hailing vehicles and update traffic congestion levels by fitting a high-resolution polynomial curve to data-driven estimates of travel times. The proposed approach results in win-win-win outcomes for passengers, transportation providers, and the urban transportation system as a whole for the computational experiments with New York City datasets.

From a practical standpoint, this research is potentially relevant for many cities worldwide that are struggling to respond to the interrelated challenges of declining public transit ridership and increasing traffic congestion — driven, in part, by the increasing popularity of ride-hailing. A few cities (e.g., Seattle, Houston) have managed to successfully revamp their public transit systems in this environment. However, there is limited academic research providing blueprints on how to revise transit networks to meet changing demand patterns in the new era where travelers' behaviors are shaped by public transit offerings, ride-hailing offerings, and traffic congestion. This chapter addresses this gap by matching the travelers with suitable transportation options. Ultimately, our findings suggest that public transit and ride-hailing can co-exist and provide complementary services in a way that enhances mobility across the urban transportation ecosystem.

In future research, our framework can be extended to integrate other transit planning considerations such as network design, line planning or rolling stock scheduling. In addition, under our current framework, the very large networks — such as the combined subway and bus system for the entire New York City — may still be difficult to solve within acceptable computational times. This motivates further research to effectively handle these extremely large problem instances.

Appendix A

Supporting Material for Chapter 2

A.1 Two-step Approach in Solving the Pricing Problem to Optimality

This appendix details the two-step approach used to solve the pricing problem to optimality.

For the pricing problem, the objective is to minimize the reduced cost of the chosen pairing. The reduced cost is equal to the objective function coefficient of the chosen pairing minus the sum of the dual variables corresponding to all flights included in this pairing. Irnich and Desaulniers [72] frame this problem as an SPPRC and utilize a dynamic programming approach to solve it. The core idea is to build paths in a flight network by extending them in all feasible directions, and to identify those paths that represent feasible crew pairings with negative reduced costs. The efficiency of this approach depends on being able to identify and eliminate paths such that these paths themselves and all their extensions are guaranteed to be sub-optimal. These nonuseful paths are discarded by using a dominance subalgorithm based on a set of dominance rules. However, the standard dominance algorithm is too slow for the computational requirements for our large-scale networks. To accelerate the dominance algorithm, we tested different variations of the dominance rules. We found that if we remove the rule that requires both the dominant and

dominated paths to have the same crew base (Algorithm B), the speed of the dominance step significantly increases. However, this simplification risks eliminating some paths that could have negative reduced costs, and hence Algorithm B is not guaranteed to identify all paths with negative reduced costs. Algorithm B, therefore, serves as an intermediate step that helps us identify some negative reduced cost pairings in a very short computational time. However, if it is unable to find any such pairings, then we revert to the full implementation of the dominance algorithm (named Algorithm A) to perform a comprehensive search for negative reduced cost pairings. In our computational experiments, we find that in most of the iterations, Algorithm B is able to identify enough variables with negative reduced costs and add them to the RMP's column pool, thus requiring us to implement Algorithm A much more sparingly and hence speeding up the pricing problem solution process dramatically. The two dominance algorithms are presented below.

Algorithm A This is a dominance algorithm with an exact implementation, similar to that described by Irnich and Desaulniers [72], wherein only a path starting with the same crew base can dominate another path.

Algorithm B This is a dominance algorithm with an implementation similar to that described by Irnich and Desaulniers [72] except that a path starting with either the same or a different crew base can dominate another path.

The exact set of labels used by Algorithm A in our robust crew-pairing implementation is as follows. Note that Algorithm B uses all but the last label listed:

1. The number of duties covered so far by the path
2. The total flying time so far in the current duty of the path
3. The total elapsed time so far in the current duty of the path
4. A constant multiple (ζ) of the total elapsed time so far in the path minus the sum of the dual contributions of all flights included so far in the path

5. The total flying time so far in the current duty plus the sum of the costs of the previous duties in the path minus the sum of the dual contributions of all flights included so far in the path
6. The minimum guaranteed pay of the current duty plus the sum of the costs of the previous duties in the path minus the sum of the dual contributions of all flights included so far in the path
7. A constant multiple of (ϵ) of the total elapsed time so far in the current duty plus the sum of the costs of the previous duties in the path minus the sum of the dual contributions of all flights included so far in the path
8. Crew base (the starting point) of the path

A.2 Local Search Heuristic for the Calibration Problem

We describe the local search heuristic used to solve our calibration problem. It is as follows.

Algorithm 2 Multi-phase solution algorithm

Initialization:
Set all parameters to zero, i.e., $\alpha_i = \beta_i = 0, i \in \{1, 2, 3, 4\}$ and $\gamma_i \in \{5, 6\}$.
while If parameter values got updated in the last iteration **do**
    for $i = 1, 2, 3, 4$ **do**
        Perform local grid-search by varying (α_i, β_i) values to minimize $\sum_{i=1}^{6} |F^{\hat{x}}(i) - F^{x}(i)|$. Update (α_i, β_i) values.
    end for
    for $i = 5, 6$ **do**
        Perform local grid-search by varying γ_i values to minimize $\sum_{i=1}^{6} |F^{\hat{x}}(i) - F^{x}(i)|$. Update γ_i values.
    end for
end while

A.3 Data Preprocessing Steps

Preprocessing consisted of two major steps. The first step was to get aircraft tail numbers for all flights in the crew scheduling samples by matching each flight in the sample with exactly one flight in the AOTP database. This is performed by matching the departure airport, arrival airport, scheduled departure time, scheduled arrival time and the airline code. We are able to match around 95% of all flights in the airline crew scheduling samples.

The second preprocessing step was data filtering to account for the limitations of the AOTP database. Because tail number information is missing for some flights in the AOTP, we use, as input to our models, only those flights for which the tail number is present in the AOTP database. Since only domestic flights information is provided in the ATOP, we removed all international flights from our crew scheduling sample as well. Typically, cockpit crews are assigned to operate aircraft belonging to only one fleet family within a given pairing. Indeed, almost all the crew pairings in our confidential crew scheduling data contained flights operated by a single fleet family. We eliminated the few crew pairings (and their corresponding flights) which cut across multiple fleet families in our crew schedule data. As a result, the crew pairing problem can be considered separately for each fleet family. For the regional carrier, we used the first week of March 2014 as our calibration dataset and the first week of April 2014 as the validation dataset. For the network legacy carrier, we used the first week of January 2014 as our calibration dataset and the first week of one month in each quarter, namely, February 2014, April 2014, July 2014, and October 2013, as our validation datasets. We also eliminated all crew pairings (and all flights in those crew pairings) such that at least one flight in that crew pairing was already removed for any of the reasons mentioned above. Overall, this resulted in the removal of approximately 15-20% of all crew pairings and approximately 10-15% of all flights in our network legacy carrier crew schedule sample across different time periods in the sample. Also, it resulted in the removal of approximately 10-15% of all our crew pairings and approximately 10-15% of all flights in our regional carrier crew schedule sample across

different time periods in the sample.

Appendix B

Supporting Material for Chapter 4

B.1 Detailed Formulation of Constraints (4.11)-(4.20)

In this appendix, we provide functional expressions of constraints (4.11)-(4.20). The numerical calibration of these expressions is detailed in Section 4.4.1 and in Appendices B.2, B.3, B.4 and B.5.

Ride-hailing Fare: $\mathbf{rhfare} = f_{rhfare}(\mathbf{ratio^R}, \mathbf{tra})$

Recall that $tra_{s,tt}$ denotes the travel time of ride-hailing vehicles taking road segment s and departing in traffic timeslot tt and that $ratio^R_{od,p,tt}$ is the fraction who use travel path p of all ride-hailing providers serving a passenger on OD pair (o, d) and starting in service in traffic timeslot tt. choosing path p in traffic timeslot tt. The ride-hailing fare is given by:

$$rhfare_{od,tt} = \sum_{p \in \mathcal{P}_{od}} ratio^R_{od,p,tt} \times \min\left(P_{\text{min}}, Surge_{tt} \times \left(P_{\text{time}} \times \sum_{s \in \mathcal{SN}_p} tra_{s,tt} + P_{\text{dist}} \times \sum_{s \in \mathcal{SN}_p} LE_s\right)\right)$$
$$, \forall (o, d) \in \mathcal{Q}, tt \in \mathcal{TT} \tag{B.1}$$

In this equation, LE_s denotes the length of road segment s, P_{min} denotes the minimum fare of each trip, P_{dist} denotes the unit price per unit of distance, P_{time} denotes the unit price per unit of time, and $Surge_{tt}$ denotes the surge pricing multiplier in traffic timeslot tt.

Public Transit Waiting Time: $\mathbf{w} = f_{WaitPT}(\mathbf{x})$

For each itinerary i, the waiting time $\mathbf{w_i}$ is given as follows:

$$w_i = \frac{SlotDur}{2 \times \sum_{v \in \mathcal{V}} x_{v,g_i,t_i}} + TR_i, \quad \forall i \in \mathcal{I}^{\text{all}} \tag{B.2}$$

Here, $SlotDur$ denotes the duration of each transit timeslot, g_i is the first line of itinerary i, t_i is its departure timeslot, and TR_i is the transfer time between consecutive lines in itinerary i (which is zero for non-connecting itineraries). The first term is the average waiting time at the first station (its denominator is two times the departure frequency per transit timeslot) and the second term is the total transfer time at all transfer stations in an itinerary.

Generalized Travel Cost and Attractiveness Calculations:

$\mathbf{tc^R} = f_{CostRH}(\mathbf{ratio^R}, \mathbf{tra}), \mathbf{a^R} = f_{AttrR}(\mathbf{tc^R}, \mathbf{rhfare}), \mathbf{tc^{TR}} = f_{CostPT}(\mathbf{w}),$
$\mathbf{a^{TR}} = f_{AttrR}(\mathbf{tc^{TR}})$

The attractiveness calculation consists of the following attributes: walking and waiting time, in-vehicle time, trip start time, transport mode and trip cost (i.e., fare). Start time is represented by a set of 0-1 dummy values corresponding to each transit timeslot. Transport mode is defined by another set of 0-1 dummies, one for each mode: transit, ride-hailing or the outside option.

We define the attributes of public transit attractiveness, $\mathbf{a^{TR}}$, as follows. The walking time is computed by multiplying the distance from the trip's origin to the nearest transit stop and the distance to the trip's destination from the nearest transit stop, with an average walking speed of 3.1 miles/hour. The waiting time w_i is given by Equation (B.2). In-vehicle time can be directly obtained from the corresponding itinerary's taking transit lines timetable. The trip fare is obtained from public information (we use a flat fare of $\$2.75$ per trip in New York City).

We define the attributes of ride-hailing attractiveness, $\mathbf{a^R}$, as follows. We assume a

constant value of the walking and waiting times (equal to 6 minutes). We calculate the in-vehicle travel time as $InVehicleTime_{od,tt} = \sum_{p \in \mathcal{P}_{od}} (ratio^{R}_{od,p,tt} \times \sum_{s \in \mathcal{SN}_p} tra_{s,tt})$. The trip cost is given by Equation (B.1).

Based on these specifications, the attractiveness values $\mathbf{a^R}$ and $\mathbf{a^{TR}}$ are obtained from Equations (4.25) and (4.26). We provide further details on the calibration of these equations in Appendix B.3.

As mentioned in the chapter, the ride-hailing and transit fares are internal costs within the system, they cancel each other out. So the cost to ride-hailing passengers $tc^{R}_{od,h,tt}$ and the cost to transit passengers $tc^{TR}_{i,h}$ are equal to their utility values excluding the trip fare.

Operating Costs and External Costs:

$\mathbf{ope^R} = f_{OpeR}(\mathbf{tra}, \mathbf{ratio^R}), \mathbf{ext^R} = f_{ExtR}(\mathbf{tra}, \mathbf{ratio^R}), \mathbf{ope^V} = f_{OpeV}(\mathbf{tra}), \mathbf{ext^V} = f_{ExtV}(\mathbf{tra})$

First, we define the operating cost and the external cost on road segment s in transit time slot tt (denoted by $ope_{s,tt}$ and $ext_{s,tt}$, respectively) by the following piece-wise linear relationships:

$$ope_{s,tt} = \widehat{Ope_s}(1 + \alpha_1 \times \max(vc_{s,tt} - \widehat{VC_s}, 0)), \quad \forall s \in \mathcal{SN}^{all}, tt \in \mathcal{TT}' \tag{B.3}$$

$$ext_{s,tt} = \widehat{Ext_s}(1 + \alpha_2 \times \max(vc_{s,tt} - \widehat{VC_s}, 0)), \quad \forall s \in \mathcal{SN}^{all}, tt \in \mathcal{TT}' \tag{B.4}$$

In these equations, $\widehat{Ope_s}$ is the free-flow operating cost on road segment s (obtained from the road segment length LE_s and the free-flow speed), $\widehat{Ext_s}$ is the free-flow external cost on road segment s (obtained from the road segment length LE_s, estimates of Carbon Dioxide emissions per mile, and estimates of the social cost of Carbon Dioxide), and $vc_{s,tt}$ is the vehicle-to-capacity ratio on road segment s in transit time slot tt at taffic timeslot tt. If the vehicle-to-capacity ratio exceeds a threshold $\widehat{VC_s}$, the travel time and resulting costs increase linearly. We use results from Skabardonis and Dowling [123] and Greenwood et al. [64] to calibrate α_1 and α_2.

We now translate these expressions into the operating costs and external costs of ride-hailing vehicles and other vehicles on the road. Let us denote by $VehicleNum_{s,tt}$ the number of vehicles on road segment s. We use the following expressions:

$$ope^R_{od,tt} = \sum_{p \in \mathcal{P}_{od}} ratio^R_{od,p,tt} \times \left(\sum_{s \in \mathcal{SN}_p} ope_{s,tt} \right), \quad \forall (o,d) \in \mathcal{Q}, tt \in \mathcal{TT}' \tag{B.5}$$

$$ext^R_{od,tt} = \sum_{p \in \mathcal{P}_{od}} ratio^R_{od,p,tt} \times \left(\sum_{s \in \mathcal{SN}_p} ext_{s,tt} \right), \quad \forall (o,d) \in \mathcal{Q}, tt \in \mathcal{TT}' \tag{B.6}$$

$$ope^V_{s,tt} = VehicleNum_{s,tt} \times ope_{s,tt}, \quad \forall s \in \mathcal{SN}^{all}, tt \in \mathcal{TT}' \tag{B.7}$$

$$ext^V_{s,tt} = VehicleNum_{s,tt} \times ext_{s,tt}, \quad \forall s \in \mathcal{SN}^{all}, tt \in \mathcal{TT}' \tag{B.8}$$

B.2 Demand Estimation

To estimate total demand for each OD pair, we use the publicly available taxi dataset in Manhattan to get the taxi demand for each OD pair $TaxiDem^{2013}_{od,h}$ [49]. From Moss et al. [101], the number of taxi trips in Manhattan dropped 34.8% from 2013 to 2017 (168.8 millions to 110.0 million). However, ride-hailing recorded an all time high of 158.0 million trips in 2017, more than offsetting the decline in taxi trips. We combine the ride-hailing service and taxi as one transport mode. So the total demand of each OD pair can be estimated as follows:

$$Dem_{od,h} = TaxiDem^{2013}_{od,h} \times (110/168) \times (1 + (158/110))/(4.62\%), \quad (o,d) \in \mathcal{Q} \tag{B.9}$$

In Equation (B.9), first, using the taxi demand from 2013 (the latest data we can access), we get the estimated number of taxi trips in 2017 for each OD pair as $TaxiDem^{od}_{2013} \times (110/168)$. Next, multiplying that number with $(1 + (158/110))$ we obtain the number of total passengers combining ride-hailing and taxi passengers. Then, by dividing by the corresponding market share of 4.62%, we obtain the total demand, $Dem_{od,h}$, of each market.

To estimate the traffic volume on each road segment, we get the average speed for each traffic timeslot and each road segment from Bertsimas et al. [24]; we estimate the number of vehicles on each road segment at any traffic timeslot through the speed-volume relationship in Table 4.1.

B.3 Passenger Utility Calibration

Trottenberg [138] provides the market share of each travel mode in New York City in 2017 as follows: *Car* 32%, *Walk* 28%, *Subway* 23%, *Bus* 8%, *Bike* 3%, *For-Hire Vehicles* 3% and *Other* 3%. We note that the probability of people switching between driving and not driving is relatively low given long-term effects like availability of parking, car ownership, etc. *Other* is not a precisely defined transportation mode in this report. We therefore include *Walk*, *Bus* and *Bike* in the definition of our outside option.

The coefficients for walking and waiting time, in-vehicle travel time and trip cost are obtained from Liu et al. [88]. We calculate the distribution of passengers' departure time preferences across the morning rush hours based on the distribution of departure times in the taxi trip data [49]. We let the utility of outside option be a linear function of the corresponding market distance. We tune the slope and intercept parameters of this linear function along with the alternative specific constants for the travel modes to ensure that the mode shares are similar to the values reported by Trottenberg [138] under the actual transit schedule in New York City. Table B.1 summarizes all utility coefficients. In particular, the attributes *Travel Start Time* and *Transport Mode* are tuned in our study and all other attributes' coefficients in Table B.1 are directly borrowed from Liu et al. [88].

B.4 Definition of Passenger Types

From Trottenberg [138], we first divide the total demand into two types based on trip purpose: time-sensitive and price-sensitive passengers. Time-sensitive passengers are those

Table B.1: Passenger utility coefficient estimates

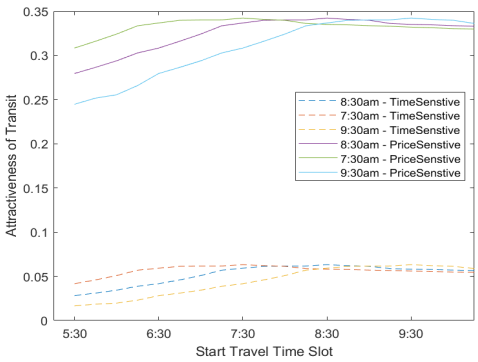
Attributes	Coefficient
Walking and waiting time	-0.032
In-vehicle travel time	-0.023
Trip cost	-0.074
Transport Mode	
Ride-hailing	-3.580
Public transit	0
Outside option	-0.2
Travel Start Time	
5:30-6:45	-0.159
5:45-6:00	-0.159
6:00-6:15	-0.159
6:15-6:30	-0.109
6:30-6:45	-0.159
6:45-7:00	-0.159
7:00-7:15	-0.159
7:15-7:30	-0.109
7:30-7:45	-0.052
7:45-8:00	-0.033
8:00-8:15	-0.015
8:15-8:30	-0.013
8:30-8:45	-0.012
8:45-9:00	0
9:00-9:15	-0.009
9:15-9:30	-0.015
9:30-9:45	-0.036
9:45-10:00	-0.042
10:00-10:15	-0.015
10:15-10:30	-0.013

whose likely trip purposes include commute to work, school, medical visits and business—which are more sensitive to in-vehicle time, waiting time, transfer time and travel start time. They account for almost 50% of all trips in New York City [138]. The remaining 50% trips are for other purposes such as shopping, running errands, social visits and recreation which are less sensitive to these time-related attributes but more sensitive to the trip fare; so they could be classified as *Price Sensitive*. We use a parameter K to calibrate the

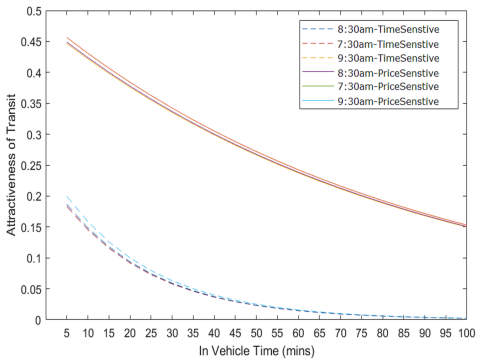
utility functions of price sensitive and time sensitive passengers. Particularly, time sensitive passengers are assumed to be willing to pay higher fares in return for more convenience. We capture this by multiplying by K the original parameters ($\beta^R_{od,h,tt}$ and $\beta^{TR}_{i,h}$) corresponding to *Walking and waiting time*, *In-vehicle travel time* and *Travel Start time* in Table B.1 and by multiplying $\frac{1}{K}$ the original parameters of *Trip cost* in Table B.1. In contrast, price sensitive passengers have a lower willingness to pay, but are more willing to accept longer travel times and or less preferred start times. So we capture this by multiplying by $\frac{1}{K}$ the original parameters ($\beta^R_{od,h,tt}$, $\beta^{od}_{i,h}$) corresponding to *Walking and waiting time*, *In-vehicle travel time* and *Travel start time* in Table B.1, and by multiplying by K the original parameters of *Trip cost* in Table B.1. In all our computational experiments, we set $K = 2$ to differentiate time sensitive passengers and price sensitive passengers.

As mentioned earlier, we obtain the passenger travel start time preferences from the taxi trip data by Donovan and Work [49]. It shows that the highest demand is during the 8:30-8:45 am timeslot and it tapers on both sides. In order to make the case study more realistic, we generate passengers belonging to multiple time preference types. Each type has a preference curve with the exact same shape as the overall taxi trip demand shape, but the mean is shifted. In particular, we consider three time preference types, with means separated by 1 hour each. We calculate the percentage of each type to ensure that the combined distribution matches the taxi trip demand pattern. Figure B.1 shows the attractiveness comparison in multiple passenger types, helping us better understand how the utility patterns differ across the passenger types. Particularly, in Figure B.1a, we fix all other utilities and only change the starting travel time in multiple passenger types. As expected, each curve's highest points match their most preferable starting travel time. For time-sensitive passengers, the lowest attractiveness value is only 30% of the highest attractiveness value. In contrast, for price-sensitive passengers, this percentage is about 70%. That means that the starting travel time has a larger impact on time-sensitive passengers than price-sensitive passenger. Also Figure B.1b shows that, as the in-vehicle time increases from 5 minutes to

100 minutes for example, the attractiveness of time-sensitive passengers decreases by 95% but the corresponding decrease is only 60% for the price-sensitive passengers.



(a) Transit Attractiveness in Travel Start Time



(b) Transit Attractiveness in In-vehicle Time

Figure B.1: Attractiveness comparison in multiple passenger types

B.5 Other Parameter Values

Table B.2 reports the other parameters used in the case study.

Table B.2: Other parameters used in the case study

Description	Value	Unit
Value of Passenger Travel Time	15.6	\$/person-hour
Vehicle Occupancy	1.25	persons/vehicle
Walking Speed	3.10	miles/hour
Ride-hailing Passengers' Average Waiting Time	6.00	min
Transit Fare	2.75	\$/person
Transit Train Crew Cost	275.00	\$/hour
Transit Train Direct Operating Cost	15.10	\$/mile
Minimum Per-Unit Operating Cost	0.1774	\$/mile
Vehicle Carbon Dioxide Emission	404	gram/mile
Social Cost of Carbon Dioxide Emissions	105	\$/ton
Minimum Fare of a Ride-hailing Trip	7.19	\$/trip
Ride-hailing Price Rate	0.66	\$/min
	1.46	\$/mile

From Liu et al. [88], we set the value of time for local commute by car in downstate New York to $\$15.6$/hour. The average vehicle occupancy (average number of persons per

vehicle), is equal to 1.25 [22]. We assume that average walking speed is 3.10 miles/hour [100] and the average waiting time of ride-hailing is 6 minutes [12]. We use the Uber price calculator [139] to calculate the ride-hailing trip fare: we use the ride-hailing price rate of $\$0.66$/minute plus $\$1.46$/mile, with a minimum fare per trip of $\$7.19$. Moreover, we apply the *surge multipliers* from Korolko et al. [83] to calculate the ride-hailing fares in different traffic timeslots—in order to account for the dynamic pricing strategies of the ride-hailing operators. We use the flat fare of $2.75 per trip from the New York City subway. The parameters for calculating the subway operating costs are obtained from Levy [86].

In our experiments, the only external cost that we consider is the cost of Carbon Dioxide emissions. We assume that the average vehicle emits 404 grams of Carbon Dioxide per mile, with a social cost of Carbon Dioxide of $105/metric ton [140].

B.6 Detailed Computational Results

Tables B.3, B.4 and B.5 provide the detailed results of the experiments described in Section 4.4.2 under the "Default", "Borough" and "Distance" demand distributions, respectively. The tables report the costs incurred under the actual transit schedule and the optimized transit schedules for each of the three transit service level requirements ("Miles","Frequency", and "Neither"). The main observations are threefold. First, the optimized transit schedule can reduce system-wide costs very significantly—by 1.85% to 2.52%, amounting to millions of dollars. Second, even with transit regularity constraints that impose a minimum level of service or a maximum deviation from the existing schedule, the optimized transit schedule can still reduce system-wide costs by 0.35% to 1.82%. Third, these large cost reductions are predominantly driven by reductions in the costs to passengers and transportation operators, without significantly affecting the costs to other vehicles and the emission costs—leading to the win-win-win outcomes described in Section 4.4.2.

Table B.3: Computational results under "Default" demand distribution, for all 3 transit service level scenarios

	Actual	Miles	Frequency	Neither
Total Cost ($)	191,101,109 (0.00%)	189,982,172 (-0.59%)	187,607,191 (-1.83%)	186,289,850 (-2.52%)
Total Cost to Passengers ($)	156,363,016 (0.00%)	156,378,478 (0.01%)	155,674,622 (-0.44%)	155,519,468 (-0.54%)
Cost to Transit Passengers ($)	11,066,971 (0.00%)	11,567,248 (4.52%)	10,117,475 (-8.58%)	9,049,702 (-18.23%)
Cost to Ride-hailing Passengers ($)	12,721,316 (0.00%)	12,905,459 (1.45%)	12,900,545 (1.41%)	13,411,437 (5.42%)
Cost to Outside Option Passengers ($)	132,574,729 (0.00%)	131,905,771 (-0.50%)	132,656,602 (0.06%)	133,058,329 (0.36%)
Total Cost to Transit and Ride-hailing Operators ($)	9,747,636 (0.00%)	8,613,524 (-11.63%)	6,942,263 (-28.78%)	5,765,794 (-40.85%)
Cost to Transit Operator ($)	7,251,348 (0.00%)	6,495,693 (-10.42%)	4,829,909 (-33.39%)	3,610,877 (-50.20%)
Cost to Ride-hailing Operator ($)	2,496,288 (0.00%)	2,117,831 (-15.16%)	2,112,354 (-15.38%)	2,154,917 (-13.68%)
Total Cost to Other Vehicles ($)	24,518,187 (0.00%)	24,526,509 (0.03%)	24,527,013 (0.04%)	24,537,134 (0.08%)
Time Cost to Other Vehicles ($)	22,924,836 (0.00%)	22,924,692 (0.00%)	22,925,196 (0.00%)	22,925,616 (0.00%)
Operating Cost to Other Vehicles ($)	1,593,351 (0.00%)	1,601,817 (0.53%)	1,601,817 (0.53%)	1,611,518 (1.14%)
Total Emission Cost ($)	472,270 (0.00%)	463,661 (-1.82%)	463,293 (-1.90%)	467,454 (-1.02%)
Cost of Emissions from Ride-hailing Vehicles ($)	73,124 (0.00%)	62,394 (-14.67%)	62,026 (-15.18%)	63,757 (-12.81%)
Cost of Emissions from Other Vehicles ($)	399,146 (0.00%)	401,267 (0.53%)	401,267 (0.53%)	403,697 (1.14%)
Market Shares				
Transit Market Share	37.53%	37.94%	35.77%	34.02%
Ride-hailing Market Share	5.81%	6.04%	6.04%	6.33%
Outside Option Market Share	56.66%	56.02%	58.19%	59.65%

B.7 Path Generation for Empty Ride-hailing Vehicles

We approximate the paths of empty ride-hailing vehicles, while ensuring that empty driving amounts 69% of the distance driven by the ride-hailing vehicles *with* passengers. Specifically, let us consider a path p taken by a passenger-carrying vehicle, and construct an empty driving path p'. The path p' ends in the origin of the path p. Moreover, we assume that the length of path p' amounts to 69% of the length of path p (i.e., we apply the 69% ratio at the level of each trip rather than at the aggregate level) and we assume that p' and p overlap

Table B.4: Computational results under "Borough" demand distribution, for all 3 transit service level scenarios

	Actual	Miles	Frequency	Neither
Total Cost ($)	255,284,312 (0.00%)	254,382,290 (-0.35%)	252,949,572 (-0.91%)	250,570,509 (-1.85%)
Total Cost to Passengers ($)	220,247,758 (0.00%)	219,490,139 (-0.34%)	219,295,608 (-0.43%)	218,724,618 (-0.69%)
Cost to Transit Passengers ($)	11,079,184 (0.00%)	12,115,479 (9.35%)	10,919,963 (-1.44%)	9,951,322 (-10.18%)
Cost to Ride-hailing Passengers ($)	12,685,063 (0.00%)	12,788,175 (0.81%)	12,900,484 (1.70%)	12,914,974 (1.81%)
Cost to Outside Option Passengers ($)	196,483,511 (0.00%)	194,586,486 (-0.97%)	195,475,162 (-0.51%)	195,858,322 (-0.32%)
Total Cost to Transit and Ride-hailing Operators ($)	10,036,256 (0.00%)	9,880,677 (-1.55%)	8,639,293 (-13.92%)	6,839,765 (-31.85%)
Cost to Transit Operator ($)	7,251,348 (0.00%)	6,502,763 (-10.32%)	5,252,301 (-27.57%)	4,045,003 (-44.42%)
Cost to Ride-hailing Operator ($)	2,784,908 (0.00%)	3,377,914 (21.29%)	3,386,992 (21.62%)	2,794,762 (0.35%)
Total Cost to Other Vehicles ($)	24,517,424 (0.00%)	24,512,799 (-0.02%)	24,515,092 (-0.01%)	24,522,138 (0.02%)
Other Vehicles' Time Cost ($)	22,925,939 (0.00%)	22,926,410 (0.00%)	22,926,587 (0.00%)	22,926,850 (0.00%)
Other Vehicles' Operating Cost ($)	1,591,485 (0.00%)	1,586,390 (-0.32%)	1,588,504 (-0.19%)	1,595,287 (0.24%)
Total Emission Cost ($)	482,874 (0.00%)	498,674 (3.27%)	499,580 (3.46%)	483,988 (0.23%)
Cost of Emissions from Ride-hailing Vehicles($)	84,195 (0.00%)	101,272 (20.28%)	101,648 (20.73%)	84,357 (0.19%)
Cost of Emissions from Other Vehicles ($)	398,679 (0.00%)	397,402 (-0.32%)	397,932 (-0.19%)	399,631 (0.24%)
Market Shares				
Transit Market Share	35.80%	35.81%	34.27%	32.66%
Ride-hailing Market Share	5.66%	5.55%	5.62%	5.81%
Outside Option Market Share	58.54%	58.64%	60.11%	61.54%

heavily (so that the busy areas for passenger-carrying trips are also the busy areas for empty trips). The algorithm essentially retraces the path p backward, as much as possible. However, reverse segments of some parts of path p may not be available in the network (due to one-way streets, for example). When such segment is encountered, the path p' chooses one of the available segment randomly with equal probability. This approximate method is detailed in Algorithm 3.

Table B.5: Computational results under "Distance" demand distribution, for all 3 transit service level scenarios

	Actual	**Miles**	**Frequency**	**Neither**
Total Cost ($)	175,497,569 (0.00%)	173,951,600 (-0.88%)	172,504,666 (-1.71%)	171,067,522 (-2.52%)
Total Cost to Passengers ($)	141,685,551 (0.00%)	140,882,980 (-0.57%)	140,669,333 (-0.72%)	140,428,436 (-0.89%)
Cost to Transit Passengers ($)	10,975,221 (0.00%)	12,173,868 (10.92%)	10,925,188 (-0.46%)	9,761,367 (-11.06%)
Cost to Ride-hailing Passengers ($)	11,141,282 (0.00%)	11,215,173 (0.66%)	11,298,515 (1.41%)	11,731,933 (5.30%)
Cost to Outside Option Passengers ($)	119,569,047 (0.00%)	117,493,938 (-1.74%)	118,445,631 (-0.94%)	118,935,136 (-0.53%)
Total Cost to Transit and Ride-hailing Operators ($)	8,833,828 (0.00%)	8,088,088 (-8.44%)	6,851,542 (-22.44%)	5,640,334 (-36.15%)
Cost to Transit Operator ($)	7,251,348 (0.00%)	6,501,057 (-10.35%)	5,255,382 (-27.53%)	4,001,311 (-44.82%)
Cost to Ride-hailing Operator ($)	1,582,480 (0.00%)	1,587,031 (0.29%)	1,596,160 (0.86%)	1,639,023 (3.57%)
Total Cost to Other Vehicles ($)	24,528,097 (0.00%)	24,529,939 (0.01%)	24,532,194 (0.02%)	24,542,732 (0.06%)
Time Cost to Other Vehicles ($)	22,932,078 (0.00%)	22,932,291 (0.00%)	22,932,490 (0.00%)	22,933,009 (0.00%)
Operating Cost to Other Vehicles ($)	1,596,019 (0.00%)	1,597,649 (0.10%)	1,599,704 (0.23%)	1,609,723 (0.86%)
Total Emission Cost ($)	450,092 (0.00%)	450,593 (0.11%)	451,597 (0.33%)	456,020 (1.32%)
Cost of Emissions from Ride-hailing Vehicles ($)	50,278 (0.00%)	50,370 (0.18%)	50,859 (1.16%)	52,773 (4.96%)
Cost of Emissions from Other Vehicles ($)	399,814 (0.00%)	400,223 (0.10%)	400,738 (0.23%)	403,247 (0.86%)
Market Shares				
Transit Market Share	34.84%	35.51%	33.90%	32.05%
Ride-hailing Market Share	5.11%	5.15%	5.21%	5.46%
Outside Option Market Share	60.05%	59.34%	60.89%	62.49%

Algorithm 3 Path Generation for Empty Ride-hailing Vehicle Driving

for each ride-hailing path $p \in \mathcal{P}^{\text{all}}$ **do**
    Initialize s := first road segment in path p.
    Initialize p' := ϕ.
    while s has a road segment s' in the opposite direction **and** total length of path p' is less than 69% of the total length of path p **do**
        Add road segment s' to path p'.
        Set s := the next road segment in path p.
    end while
    Set n := starting node of segment s
    while total length of path p' is less than 69% of the total length of path p **do**
        Select a segment s' randomly among all road segments in $\mathcal{SN}^{all}$ ending in node n.
        Add road segment s' to path p'.
        Set n := starting node of segment s'.
    end while
    Save path p' as the prior driving path before picking up passengers and taking path p.
end for

Bibliography

[1] Jeph Abara. Applying integer linear programming to the fleet assignment problem. *Interfaces*, 19(4):20–28, 1989.

[2] Ahmed Abdelghany, Khaled Abdelghany, and Farshid Azadian. Airline flight schedule planning under competition. *Computers & Operations Research*, 87:20–39, 2017.

[3] Niels Agatz, Alan Erera, Martin Savelsbergh, and Xing Wang. Dynamic ride-sharing: A simulation study in metro atlanta. *Transportation Research Part B: Methodological*, 45(9):1450–1464, 2011.

[4] Shervin AhmadBeygi, Amy Cohn, Yihan Guan, and Peter Belobaba. Analysis of the potential for delay propagation in passenger airline networks. *Journal of Air Transport Management*, 14(5):221–236, 2008.

[5] Ravindra Ahuja and James Orlin. Inverse optimization. *Operations Research*, 49 (5):771–783, 2001.

[6] Ravindra Ahuja, Jon Goodstein, Amit Mukherjee, James Orlin, and Dushyant Sharma. A very large-scale neighborhood search algorithm for the combined through-fleet-assignment model. *INFORMS Journal on Computing*, 19(3):416–428, 2007.

[7] Shawn Allan, Tony Beesley, James Evans, and Stephen Gaddy. Analysis of delay causality at newark international airport. In *4th USA/Europe Air Traffic Management R&D Seminar*, Santa Fe, NM, 2001.

[8] Javier Alonso-Mora, Samitha Samaranayake, Alex Wallar, Emilio Frazzoli, and Daniela Rus. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, 114(3):462–467, 2017.

[9] Mohammad Amin-Naseri and Vahid Baradaran. Accurate estimation of average waiting time in public transportation systems. *Transportation Science*, 49(2):213–222, 2015.

[10] Michael Anderson. Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *American Economic Review*, 104(9):2763–2796, 2014.

[11] Renato Arbex and Claudio Cunha. Efficient transit network design and frequencies setting multi-objective optimization by alternating objective genetic algorithm. *Transportation Research Part B: Methodological*, 81:355–376, 2015.

[12] Jiaru Bai, Kut So, Christopher Tang, Xiqun (Michael) Chen, and Hai Wang. Coordinating supply and demand on an on-demand service platform with impatient customers. *Manufacturing Service Operation Management*, 21(3):556–570, 2019.

[13] Michael Ball, Cynthia Barnhart, Martin Dresner, Mark Hansen, Kevin Neels, Amedeo Odoni, Everett Peterson, Lance Sherry, Antonio Trani, and Bo Zou. Total delay impact study: A comprehensive assessment of the costs and impacts of flight delay in the united states. Technical report, National Center of Excellence for Aviations Operations Research (NEXTOR), College Park, MD, 2010.

[14] Siddhartha Banerjee, Ramesh Johari, and Carlos Riquelme. Pricing in ride-sharing platforms: A queueing-theoretic approach. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, page 639, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450334105.

[15] Cynthia Barnhart and Vikrant Vaze. Airline Schedule Optimization. In Peter Belobaba, Amedeo Odoni, and Cynthia Barnhart, editors, *The Global Airline Industry*, pages 189–222. John Wiley & Sons, West Sussex, UK, 2015.

[16] Cynthia Barnhart and Vikrant Vaze. Irregular Operations: Schedule Recovery and Robustness. In Peter Belobaba, Amedeo Odoni, and Cynthia Barnhart, editors, *The Global Airline Industry*, pages 263–287. John Wiley & Sons, West Sussex, UK, 2015.

[17] Cynthia Barnhart, Ellis Johnson, George Nemhauser, Martin Savelsbergh, and Pamela Vance. Branch-and-price: Column generation for solving huge integer programs. *Operation Research*, 46(3):316–329, 1998.

[18] Cynthia Barnhart, Timothy Kniker, and Manoj Lohatepanont. Itinerary-based airline fleet assignment. *Transportation Science*, 36(2):199–217, 2002.

[19] Cynthia Barnhart, Amr Farahat, and Manoj Lohatepanont. Airline fleet assignment with enhanced revenue modeling. *Operation Research*, 57(1):231–244, 2009.

[20] Cynthia Barnhart, Douglas Fearing, and Vikrant Vaze. Modeling passenger travel and delays in the national air transportation system. *Operation Research*, 62(3): 580–601, 2014.

[21] Eva Barrena, David Canca, Leandro Coelho, and Gilbert Laporte. Single-line rail rapid transit timetabling under dynamic passenger demand. *Transportation Research Part B: Methodological*, 70:134–150, 2014.

[22] Justin Beaudoin, Hossein Farzin, and Cynthia Lawell. Public transit investment and sustainable transportation: A review of studies of transit's impact on traffic congestion and air quality. *Research in Transportation Economic*, 52:15–22, 2015.

[23] Moshe Ben-Akiva and Steven Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, MA, 1985.

[24] Dimitris Bertsimas, Arthur Delarue, Patrick Jaillet, and Sebastien Martin. Travel time estimation in the age of big data. *Operation Research*, 67(2):498–515, 2019.

[25] Chandra Bhat, Naveen Eluru, and Rachel Copperman. Flexible Model Structures for Discrete Choice Analysis. In David Hensher and Kenneth Button, editors, *Handbook of Transport Modelling*, pages 75–104. Elsevier, Amsterdam, 2008.

[26] Kostas Bimpikis, Ozan Candogan, and Daniela Saban. Spatial pricing in ride-sharing networks. *Operation Research*, 67(3):744–769, 2019.

[27] Ralf Borndorfer, Martin Grotschel, and Marc Pfetsch. A column-generation approach to line planning in public transport. *Transportation Science*, 41(1):123–132, 2007.

[28] Bureau of Transportation Statistics. Airline on-time performance data. `https://www.transtats.bts.gov/Tables.asp?DB_ID=120`, 2017. [Online; accessed 29-June-2018].

[29] Bureau of Transportation Statistics. The Airline Origin and Destination Survey (DB1B). `https://www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=125t`, 2017. [Online; accessed 29-June-2018].

[30] Bureau of Transportation Statistics. Airline on-time performance and causes of flight delays. `https://www.bts.gov/topics/airlines-and-airports/airline-time-performance-and-causes-flight-delays`, 2019. [Online; accessed 29-July-2019].

[31] Gerard Cachon, Kaitlin Daniels, and Ruben Lobel. The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing Service Operation Management*, 19(3):368–384, 2017.

[32] Luis Cadarso, Vikrant Vaze, Cynthia Barnhart, and Angel Marin. Integrated airline scheduling: Considering competition effects and the entry of the high speed rail. *Transportation Science*, 51(1):132–154, 2017.

[33] Lorenzo Castelli, Raffaele Pesenti, and Walter Ukovich. Scheduling multimodal transportation systems. *European Journal of Operational Research*, 155(3):603–615, 2004.

[34] Avishai Ceder. Bus frequency determination using passenger count data. *Transportation Research Part A: Policy and Practice*, 18(5):439 – 453, 1984.

[35] Avishai Ceder. Methods for creating bus timetables. *Transportation Research Part A: Policy and Practice*, 21(1):59 – 83, 1987.

[36] Avishai Ceder and Nigel Wilson. Bus network design. *Transportation Research Part B: Methodological*, 20(4):331 – 344, 1986.

[37] Linxi Chen, Paul Schonfeld, and Elise Miller-Hooks. Welfare maximization for bus transit systems with timed transfers and financial constraints. *Journal of Advanced Transportation*, 50(4):421–433, 2016.

[38] Peng Chen and Yu Nie. Connecting e-hailing to mass transit platform: Analysis of relative spatial position. *Transportation Research Part B: Emerging Technologies*, 77:444–461, 2017.

[39] Steven Chien and Lazar Spasovic. Optimization of grid bus transit systems with elastic demand. *Journal of Advanced Transportation*, 36(1):63–91, 2002.

[40] James Chu. Mixed-integer programming model and branch-and-price-and-cut algorithm for urban bus network design and timetabling. *Transportation Research Part B: Methodological*, 108:188–216, 2018.

[41] Regina Clewlow and Gouri Shankar Mishra. Disruptive transportation: The adoption, utilization, and impacts of ride-hailing in the united states. *University of California, Davis, Institute of Transportation Studies, Davis, CA, Research Report UCD-ITS-RR-17-07*, 2017.

[42] Gregory Coldren, Frank Koppelman, Krishnan Kasturirangan, and Amit Mukherjee. Modeling aggregate air-travel itinerary shares: logit model development at a major us airline. *Journal of Air Transport Management*, 9(6):361–369, 2003.

[43] Robert Curley. NYC transit system information added to Uber app. Available at https://www.businesstraveller.com/business-travel/2019/11/17/nyc-transit-system-info-added-to-uber/, 2019.

[44] Carlos Daganzo. *Fundamentals of Transportation and Traffic Operations*. Emerald Group Publishing Limited, 1997.

[45] Carlos Daganzo. On the design of public infrastructure systems with elastic demand. *Transportation Research Part B: Methodological*, 46(9):1288 – 1293, 2012.

[46] Carlos Daganzo and Karen R. Smilowitz. Bounds and approximations for the transportation problem of linear programming and other scalable network problems. *Transportation Science*, 38(3):343–356, 2004.

[47] Guy Desaulniers and Mark Hickman. Public Transit. In Cynthia Barnhart and Gilbert Laporte, editors, *Handbooks in Operations Research and Management Science*, pages 69–127. Elsevier, 2007.

[48] Guy Desaulniers, Marius Solomon, and Jacques Desrosiers, editors. *Column Generation*. Springer, New York, 2005.

[49] Brian Donovan and Daniel Work. Empirically quantifying city-scale transportation system resilience to extreme events. *Transportation Research Part B: Emerging Technologies*, 79:333 – 346, 2017.

[50] Zhaoyang Duan and Lizhi Wang. Heuristic algorithms for the inverse mixed integer linear programming problem. *Journal of Global Optimization*, 51(3):463–471, 2011.

[51] Michelle Dunbar, Gary Froyland, and Chenglung Wu. Robust airline schedule planning: Minimizing propagated delay in an integrated routing and crewing framework. *Transportation Science*, 46(2):204–216, 2012.

[52] Gregory Erhardt, Sneha Roy, Drew Cooper, Bhargava Sana, Mei Chen, and Joe Castiglione. Do transportation network companies decrease or increase congestion? *Science advances*, 5(5), 2019.

[53] Daniel Fagnant, Kara Kockelman, and Prateek Bansal. Operations of shared autonomous vehicle fleet for austin, texas, market. *Transportation Research Record*, 2536:98–106, 2015.

[54] Wei Fan. *Optimal transit route network design problem: algorithms, implementations, and numerical results*. PhD thesis, University of Texas at Austin, 2004.

[55] Reza Zanjirani Farahani, Elnaz Miandoabchi, Waiyuen Szeto, and Hannaneh Rashidi. A review of urban transportation network design problems. *European Journal of Operational Research*, 229(2):281–302, 2013.

[56] Fehr & Peers. Estimated TNC Share of VMT in Six US Metropolitan Regions, 2019.

[57] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, 1956.

[58] Peter Furth and Nigel Wilson. Setting frequencies on bus routes: Theory and practice. *Transportation Research Record*, 818:1–7, 1981.

[59] Guillermo Gallego, Richard Ratliff, and Sergey Shebalov. A general attraction model and sales-based linear program for network revenue management under customer choice. *Operation Research*, 63(1):212–232, 2015.

[60] Chunhua Gao, Ellis Johnson, and Barry Smith. Integrated airline fleet and crew robust planning. *Transportation Science*, 43(1):2–16, 2009.

[61] Laurie Garrow, Gregory Coldren, and Frank Koppelman. MNL, NL, and OGEV models of itinerary choice. In Laurie Garrow, editor, *Discrete Choice Modelling and Air Travel Demand: Theory and Applications*, pages 203–251. Ashgate Publishing, Ltd, Burlington, VT, 2010.

[62] Fred Glover, Randy Glover, Joe Lorenzo, and Claude Mcmillan. The passenger-mix problem in the scheduled airlines. *Interfaces*, 12(3):73–80, 1982.

[63] Michael Graehler, Alex Mucci, and Gregory Derhardt. Understanding the Recent Transit Ridership Decline in Major US Cities: Service Cuts or Emerging Modes? In *Transportation Research Board*, 2019.

[64] Ian Greenwood, Rogers Dunn, and Robert Raine. Estimating the effects of traffic congestion on fuel consumption and vehicle emissions based on acceleration noise. *Journal of Transportation Engineering*, 133(2):96–104, 2007.

[65] Valerie Guihaire and Jin-Kao Hao. Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, 42(10):1251–1273, 2008.

[66] Xin Guo, Huijun Sun, Jianjun Wu, Jiangang Jin, Jin Zhou, and Ziyou Gao. Multiperiod-based timetable optimization for metro transit networks. *Transportation Research Part B: Methodological*, 96:46–67, 2017.

[67] Christopher Hane, Cynthia Barnhart, Ellis Johnson, Roy Marsten, George Nemhauser, and Gabriele Sigismondi. The fleet assignment problem: Solving a large-scale integer program. *Mathematical Programming*, 70(1):211–232, 1995.

[68] Alejandro Henao and Wesley Marshall. The impact of ridehailing on vehicle miles traveled. *Transportation*, 46:2173–2194, 2018.

[69] Evelien Hurk, Haris Koutsopoulos, Nigel Wilson, Leo Kroon, and Gabor Maroti. Shuttle planning for link closures in urban public transport networks. *Transportation Science*, 50(3):947–965, 2016.

[70] Omar Ibarra-Rojas and Yasmin Rios-Solis. Synchronization of bus timetabling. *Transportation Research Part B: Methodological*, 46(5):599 – 614, 2012.

[71] INRIX. The future economic and environmental costs of gridlock in 2030. `https://www.ibtta.org/sites/default/files/documents/MAF/Costs-of-Congestion-INRIX-Cebr-Report.pdf`, 2014.

[72] Stefan Irnich and Guy Desaulniers. Shortest path problems with resource constraints. In Guy Desaulniers, Jacques Desrosiers, and Marius Solomon, editors, *Column Generation*, pages 33–65, Boston, MA, 2005. Springer US. ISBN 978-0-387-25486-9.

[73] Timothy Jacobs, Barry Smith, and Ellis Johnson. Incorporating network flow effects into the airline fleet assignment process. *Transportation Science*, 42(4):514–529, 2008.

[74] Timothy Jacobs, Laurie Garrow, Manoj Lohatepanont, Frank Koppelman, Gregory Coldren, and Hadi Purnomo. Airline planning and schedule development. In Cynthia Barnhart and Barry Smith, editors, *Quantitative Problem Solving Methods in the Airline Industry: A Modeling Methodology Handbook*, pages 35–99. Springer, New York, NY, 2012.

[75] Alexandre Jacquillat and Amedeo Odoni. An integrated scheduling and operations approach to airport congestion mitigation. *Operations Research*, 63(6):1390–1410, 2015.

[76] Hai Jiang and Cynthia Barnhart. Robust airline schedule design in a dynamic scheduling environment. *Computers & Operations Research*, 40(3):831–840, 2013.

[77] Jiangang Jin, Kwongmeng Teo, and Amedeo Odoni. Optimizing bus bridging services in response to disruptions of urban transit rail networks. *Transportation Science*, 50(3):790–804, 2016.

[78] Atoosa Kasirzadeh, Mohammed Saddoune, and François Soumis. Airline crew scheduling: models, algorithms, and data sets. *EURO Journal on Transportation and Logistics*, 6(2):111–137, 2017.

[79] Konstantinos Kepaptsoglou and Matthew Karlaftis. Transit route network design problem: Review. *Journal of Transportation Engineering*, 135(8):491–505, 2009.

[80] Myungseob (Edward) Kim and Paul Schonfeld. Integrating bus services with mixed fleets. *Transportation Research Part B: Methodological*, 55:227 – 244, 2013.

[81] Timothy Kniker. *Itinerary-based airline fleet assignment*. PhD thesis, Massachusetts Institute of Technology, Cambridge, 1998.

[82] Frank Koppelman, Gregory Coldren, and Roger Parker. Schedule delay impacts on air-travel itinerary demand. *Transportation Research Part B: Methodological*, 42(3): 263–273, 2008.

[83] Nikita Korolko, Dawn Woodard, Chiwei Yan, and Helin Zhu. Dynamic pricing and matching in ride-hailing platforms. *Working Paper*, 2018. Available at SSRN: https://ssrn.com/abstract=3258234.

[84] Jourdain Lamperski and Andrew Schaefer. A polyhedral characterization of the inverse-feasible region of a mixed-integer program. *Operations Research Letters*, 43(6):575–578, 2015.

[85] Alan Lee and Martin Savelsbergh. An extended demand responsive connector. *EURO Journal on Transportation and Logistics*, 6(1):25–50, 2017.

[86] Alon Levy. Why are the NYC subway's operating costs so high? `https://ny.curbed.com/2017/10/13/16455880/new-york-city-subway-mta-operating-cost-analysis`, 2017. [Online; accessed 29-Jan-2019].

[87] Tao Liu and Avishai Ceder. Integrated public transport timetable synchronization and vehicle scheduling with demand assignment: A bi-objective bi-level model using deficit function approach. *Transportation Research Part B: Methodological*, 117: 935–955, 2018.

[88] Yang Liu, Prateek Bansal, Ricardo Daziano, and Samitha Samaranayake. A framework to integrate mode choice in the design of mobility-on-demand systems. *Transportation Research Part C: Emerging Technologies*, 105:648 – 665, 2019.

[89] Manoj Lohatepanont and Cynthia Barnhart. Airline schedule planning: Integrated models and algorithms for schedule design and fleet assignment. *Transportation Science*, 38(1):19–32, 2004.

[90] Da Lu and Fatma Gzara. The robust crew pairing problem: model and solution methodology. *Journal of Global Optimization*, 62(1):29–54, 2015.

[91] Virginie Lurkin. *Modeling in air transportation: cargo loading and itinerary choice*. PhD thesis, HEC-Liege,Belgium, 2016.

[92] Virginie Lurkin, Laurie Garrow, Matthew Higgins, Jeffrey Newman, and Michael Schyns. Accounting for price endogeneity in airline itinerary choice models: An application to continental u.s. markets. *Transportation Research Part A: Policy and Practice*, 100:228 – 246, 2017.

[93] Daniel Mangrum and Alejandro Molnar. The marginal congestion of a taxi in New York City. *Working paper, Vanderbilt University*, 2017.

[94] Adolf May. *Traffic flow fundamentals*. Prentice Hall, 1990.

[95] Daniel McFadden and Kenneth Train. Mixed mnl models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470, 2000.

[96] Jeffrey McGill and Garrett Ryzin. Revenue management: Research overview and prospects. *Transportation Science*, 33(2):233–256, 1999.

[97] McKinsey & Company. Cracks in the ridesharing market and how to fill them. *McKinsey Quarterly*, 2017.

[98] Marcos Medina, Ricardo Giesen, and Juan Carlos Munoz. Model for the optimal location of bus stops and its application to a public transport corridor in santiago, chile. *Transportation Research Record*, 2352(1):84–93, 2013.

[99] Anne Mercier, Jean-Francois Cordeau, and Francois Soumis. A computational study of benders decomposition for the integrated aircraft routing and crew scheduling problem. *Computers & Operations Research*, 32(6):1451–1476, 2005.

[100] Betty Mohler, William Thompson, Sarah Creem-Regehr, Herbert Pick, and William Warren. Visual flow influences gait transition speed and preferred walking speed. *Experimental brain research*, 181(2):221–228, 2007.

[101] Mitchell Moss, John Samuelsen, David Paterson, Sam Schwartz, Peter Ward, Darryl Towns, Scott Rechler, Tom Prendergast, James Molinaro, Kathy Wylde, Fernando Ferrer, Kevin Law, Bill Rudin, Floyd Flake, and Steve Bellone. Fix NYC Advisory Panel and Report. Technical report, Fix NYC, 2018. https://www.hntb.com/HNTB/Fix-NYC-Panel-Report.pdf.

[102] MTA. Metropolitan transportation authority (mta) data feeds. http://web.mta.info/developers/developer-data-terms.html#data>, 2019. [Online; accessed 29-Jan-2019].

[103] Paul Mudde and Parvez Sopariwala. U.S. airways merger: A strategic variance analysis of changes in post-merger performance. *Journal of Accounting Education*, 32(3):305–322, 2014.

[104] Seyed Mohammad Nourbakhsh and Yanfeng Ouyang. A structured flexible transit system for low demand areas. *Transportation Research Part B: Methodological*, 46 (1):204 – 216, 2012.

[105] Michael Patriksson. *The traffic assignment problem: models and methods*. Courier Dover Publications, 2015.

[106] Jon Petersen, Gustaf Solveling, John-Paul Clarke, Ellis Johnson, and Sergey Shebalov. An optimization approach to airline integrated recovery. *Transportation Science*, 46(4):482–500, 2012.

[107] Joao Pita, Cynthia Barnhart, and Antonio Antunes. Integrated flight scheduling and fleet assignment under airport congestion. *Transportation Science*, 47(4):477–492, 2013.

[108] Nikolas Pyrgiotis, Kerry Malone, and Amedeo Odoni. Modelling delay propagation within an airport network. *Transportation Research Part C: Emerging Technologies*, 27:60–75, 2013.

[109] Stefan Ropke and David Pisinger. An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. *Transportation Science*, 40 (4):455–472, 2006.

[110] Jay Rosenberger, Andrew Schaefer, David Goldsman, Ellis Johnson, Anton Kleywegt, and George Nemhauser. A stochastic model of airline operations. *Transportation Science*, 36(4):357–377, 2002.

[111] Andrew Schaefer, Ellis Johnson, Anton Kleywegt, and George Nemhauser. Airline crew scheduling under uncertainty. *Transportation Science*, 39(3):340–348, 2005.

[112] Bruce Schaller. The new automobility: Lyft, uber and the future of american cities. Technical report, Schaller Consulting, Brooklyn, NY., 2018.

[113] Marie Schmidt. Integrating routing decisions in network problems. In Stefan Helber, Michael Breitner, Daniel Rösch, Cornelia Schön, Johann-Matthias Graf von der Schulenburg, Philipp Sibbertsen, Marc Steinbach, Stefan Weber, and Anja Wolter, editors, *Operations Research Proceedings 2012*, pages 21–26, Cham, 2014. Springer International Publishing.

[114] Anita Schobel. Line planning in public transportation: models and methods. *OR Spectrum*, 34(3):491–510, 2012.

[115] David Schrank, Bill Eisele, Tim Lomax, and Jim Bak. 2015 Urban Mobility Scorecard. Technical report, Texas A&M Transportation Institute, 2015.

[116] Sergey Shebalov and Diego Klabjan. Robust airline crew pairing: Move-up crews. *Transportation Science*, 40(3):300–312, 2006.

[117] Hanif Sherali and Xiaomei Zhu. Two-stage fleet assignment model considering stochastic passenger demands. *Operation Research*, 56(2):383–399, 2008.

[118] Hanif Sherali, Ebru Bish, and Xiaomei Zhu. Polyhedral analysis and algorithms for a demand-driven refleeting model for aircraft assignment. *Transportation Science*, 39(3):349–366, 2005.

[119] Hanif Sherali, Ki-Hwan Bae, and Mohamed Haouari. Integrated airline schedule design and fleet assignment: Polyhedral analysis and benders' decomposition approach. *INFORMS Journal on Computing*, 22(4):500–513, 2010.

[120] Hanif Sherali, Ki-Hwan Bae, and Mohamed Haouari. A benders decomposition approach for an integrated airline schedule design and fleet assignment problem with flight retiming, schedule balance, and demand recapture. *Annals of Operations Research*, 210(1):213–244, 2013.

[121] Hanif Sherali, Ki-Hwan Bae, and Mohamed Haouari. An integrated approach for airline flight selection and timing, fleet assignment, and aircraft routing. *Transportation Science*, 47(4):455–476, 2013.

[122] Paul Sillers. How two boston-based startups are forecasting flight delays. `https://apex.aero/2019/01/23/fix-this-boston-startups-flight-delays`, 2019. [Online; accessed 29-April-2019].

[123] Alexander Skabardonis and Richard Dowling. Improved speed-flow relationships for planning applications. *Transportation Research Record*, 1572:18–23, 1997.

[124] Kenneth Small. Unnoticed lessons from london: Road pricing and public transit. *Access*, 26:10–15, 2005.

[125] Barry Smith and Ellis Johnson. Robust airline fleet assignment: Imposing station purity using station decomposition. *Transportation Science*, 40(4):497–516, 2006.

[126] Milind Sohoni, YuChing Lee, and Diego Klabjan. Robust airline scheduling under block-time uncertainty. *Transportation Science*, 45(4):451–464, 2011.

[127] Konrad Steiner and Stefan Irnich. Strategic Planning for Integrated Mobility-on-Demand and Urban Public Bus Networks. Working Papers 1819, Gutenberg School of Management and Economics, Johannes Gutenberg-Universitat Mainz, 2018.

[128] Mitja Stiglic, Niels Agatz, Martin Savelsbergh, and Mirko Gradisar. Enhancing urban mobility: Integrating ride-sharing and public transit. *Computers & Operations Research*, 90:12–21, 2018.

[129] StreetsBlog. Transit Ridership Falling Everywhere—But Not in Cities With Redesigned Bus Networks. Available at https://usa.streetsblog.org/2017/02/24/transit-ridership-falling-everywhere-but-not-in-cities-with-redesigned-bus-networks/, 2017.

[130] Lijun Sun, Jiangang Jin, Der-Horng Lee, Kay Axhausen, and Alexander Erath. Demand-driven timetable design for metro services. *Transportation Research Part B: Emerging Technologies*, 46:284–299, 2014.

[131] William Swan and Nicole Adler. Aircraft trip cost parameters: A function of stage length and seat capacity. *Transportation Research Part E: Logistics and Transportation Review*, 42(2):105–115, 2006.

[132] Prem Swaroop, Bo Zou, Michael Ball, and Mark Hansen. Do more us airports need slot controls? a welfare based approach to determine slot levels. *Transportation Research Part B: Methodological*, 46(9):1239–1259, 2012.

[133] Waiyuen Szeto and Yongzhong Wu. A simultaneous bus route design and frequency setting problem for tin shui wai, hong kong. *European Journal of Operational Research*, 209(2):141–155, 2011.

[134] Bassy Tam, Matthias Ehrgott, David Ryan, and Golbon Zakeri. A comparison of stochastic programming and bi-objective optimisation approaches to robust airline crew scheduling. *OR Spectrum*, 33(1):49–75, 2011.

[135] Terry Taylor. On-demand service platforms. *Manufacturing Service Operation Management*, 20(4):704–720, 2018.

[136] The Economist. Public transport is in decline in many wealthy cities. Available at https://www.economist.com/international/2018/06/21/public-transport-is-in-decline-in-many-wealthy-cities, 2018.

[137] TransitCenter. There's a reason transit ridership is rising in these 7 cities. Available at http://transitcenter.org/2019/02/27/theres-a-reason-transit-ridership-is-rising-in-these-7-cities/, 2019.

[138] Polly Trottenberg. New York city mobility report. Technical report, NYC Department of Transportation, 2018. Available at http://www.nyc.gov/html/dot/downloads/pdf/mobility-report-2018-print.pdf.

[139] UBER. Manhattan Uber prices. `https://www.uber.com/fare-estimate/`, 2019.

[140] U.S. Environmental Protection Agency. Greenhouse gas emissions from a typical passenger vehicle. `https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100U8YT.pdf`, 2018.

[141] Vikrant Vaze and Cynthia Barnhart. Modeling airline frequency competition for airport congestion mitigation. *Transportation Science*, 46(4):512–535, 2012.

[142] Di Wang, Diego Klabjan, and Sergey Shebalov. Attractiveness-based airline network models with embedded spill and recapture. *Journal of Airline Airport Management*, 4(1):1–25, 2014.

[143] Lizhi Wang. Branch-and-bound algorithms for the partial inverse mixed integer linear programming problem. *Journal of Global Optimization*, 55(3):491–506, 2013.

[144] Xiaoqing Wang, Maged Dessouky, and Fernando Ordonez. A pickup and delivery problem for ridesharing considering congestion. *Transportation Letters*, 8:259–269, 2016.

[145] John Glen Wardrop. Road paper. some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(3):325–362, 1952.

[146] Washington Post. Falling transit ridership poses an 'emergency' for cities, experts fear. Available at https://www.washingtonpost.com/local/trafficandcommuting/falling-transit-ridership-poses-an-emergency-for-cities-experts-fear/2018/03/20/ffb67c28-2865-11e8-874b-d517e912f125story.html, 2018.

[147] Oliver Weide, David Ryan, and Matthias Ehrgott. An iterative approach to robust and integrated aircraft routing and crew scheduling. *Computers & Operations Research*, 37(5):833–844, 2010.

[148] Clifford Winston and Vikram Maheshri. On the social desirability of urban rail transit systems. *Journal of Urban Economics*, 62(2):362–382, 2007.

[149] Rachel Wong, Tony Yuen, Kwok Fung, and Janny Leung. Optimizing timetable synchronization for rail mass transit. *Transportation Science*, 42(1):57–69, 2008.

[150] Jun Xie and Yu (Marco) Nie. A new algorithm for achieving proportionality in user equilibrium traffic assignment. *Transportation Science*, 53(2):566–584, 2019.

[151] Shangyao Yan, ChingHui Tang, and TsengChih Fu. An airline scheduling model and solution algorithms under stochastic demands. *European Journal of Operational Research*, 190(1):22–39, 2008.

[152] Joyce Yen and John Birge. A stochastic programming approach to the airline crew scheduling problem. *Transportation Science*, 40(1):3–14, 2006.

[153] Ying Zhou, Hongsok Kim, Paul Schonfeld, and Eungcheol Kim. Subsidies and welfare maximization tradeoffs in bus transit systems. *The Annals of Regional Science*, 42(3):643–660, 2008.

[154] Konstantinos Zografos, Yiannis Salouras, and Michael Madas. Dealing with the efficient allocation of scarce resources at congested airports. *Transportation Research Part B: Emerging Technologies*, 21(1):244–256, 2012.