ABSTRACT

Title of Dissertation: MODELS AND SOLUTION ALGORITHMS FOR EQUITABLE RESOURCE ALLOCATION IN AIR TRAFFIC FLOW MANAGEMENT

Ming Zhong, Ph.D., 2012

Directed By: Professor Michael O. Ball
Department of Decision, Operations and Information Technologies
Robert H. Smith School of Business

Population growth and economic development lead to increasing demand for travel and pose mobility challenges on capacity-limited air traffic networks. The U.S. National Airspace System (NAS) has been operated near the capacity, and air traffic congestion is expected to remain as a top concern for the related system operators, passengers and airlines. This dissertation develops a number of model reformulations and efficient solution algorithms to address resource allocation problems in air traffic flow management, while explicitly accounting for equitable objectives in order to encourage further collaborations by different stakeholders.

This dissertation first develops a bi-criteria optimization model to offload excess demand from different competing airlines in the congested airspace when the predicted traffic demand is higher than available capacity. Computationally efficient network flow models with side constraints are developed and extensively tested using

datasets obtained from the Enhanced Traffic Management System (ETMS) database (now known as the Traffic Flow Management System). Representative Pareto-optimal tradeoff frontiers are consequently generated to allow decision-makers to identify best-compromising solutions based on relative weights and systematical considerations of both efficiency and equity.

This dissertation further models and solves an integrated flight re-routing problem on an airspace network. Given a network of airspace sectors with a set of waypoint entries and a set of flights belonging to different air carriers, the optimization model aims to minimize the total flight travel time subject to a set of flight routing equity, operational and safety requirements. A time-dependent network flow programming formulation is proposed with stochastic sector capacities and rerouting equity for each air carrier as side constraints. A Lagrangian relaxation based method is used to dualize these constraints and decompose the original complex problem into a sequence of single flight rerouting/scheduling problems.

Finally, within a multi-objective utility maximization framework, the dissertation proposes several practically useful heuristic algorithms for the long-term airport slot assignment problem. Alternative models are constructed to decompose the complex model into a series of hourly assignment sub-problems. A new paired assignment heuristic algorithm is developed to adapt the round robin scheduling principle for improving fairness measures across different airlines. Computational results are presented to show the strength of each proposed modeling approach.

MODELS AND SOLUTION ALGORITHMS FOR EQUITABLE RESOURCE ALLOCATION IN AIR TRAFFIC FLOW MANAGEMENT

By

Ming Zhong

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor Michael O. Ball Chair
Professor David Lovell
Professor Lawrence Bodin
Professor Zhi-Long Chen
Professor S. Raghavan

Acknowledgements

I am indebted to my advisor and professor, Dr. Michael O. Ball, who provides me guidance and encouragement throughout the course of my Ph.D. study. His commitment to quality and excellence has been sources of inspiration making it a privilege to work with him.

Very special thanks are to my committee members, Professors David Lovell, Lawrence Bodin, S. Raghavan and Zhi-Long Chen for their support, invaluable insight, and constructive comments. I would also like to extend my gratitude to Dr. Robert Hoffman who supported and guided me during the beginning of my doctoral studies.

I wish to express my thanks to a great number of people who generously offered advice and friendship throughout my time at the University of Maryland and the University of Texas at Austin.

I would like to express my deep appreciation to my husband and my parents, who were very supportive during the final year of my studies.

Table of Contents

List of Tables

List of Figures

List of Major Abbreviations

Airspace Flow Program (AFP)

Air Traffic Control (ATC)

Air Traffic Control Systems Command Center (ATCSCC)

Air Traffic Flow Management (ATFM)

Bureau of Transportation Statistics (BTS)

Collaborative Decision Making (CDM)

Enhanced Traffic Management System (ETMS), now known as Traffic Flow Management System (TFMS)

Federal Aviation Administration (FAA)

Flow Constrained Area (FCA)

Ground-Delay Program (GDP)

Ground Delay Program Enhancement (GDPE)

National Airspace System (NAS)

Ration by Schedule (RBS)

Traffic Flow Management (TFM)

Chapter 1: Introduction

Population growth and economic development lead to increasing demand for travel and pose mobility challenges on capacity-limited air traffic networks. As air traffic demand continues to increase, the U.S. National Airspace System (NAS) operates near its capacity. Air traffic congestion is expected to remain as a top concern for the related public agencies and private industry. According to the Bureau of Transportation Statistics (BTS), a flight is classified as delayed if it arrives 15 minutes later than the published schedule. Reported by the airline on-time statistics website (BTS, 2011), from 2000 to 2010, the percentage of delayed flights has been remaining in a range of 17% to 22%, with a mean value of 20.70%. Specifically, in 2009, 17.21% of flights have late departures and 19.46% of lights have late arrivals, while 1.85% of flights were canceled and 0.26% of flights were diverted.

In a recently concluded Total Delay Impact Study (Ball et al. 2010), the total cost of US air traffic delays was estimated to be $32.9 billion dollars for calendar year 2007. The largest component is a $16.7 billion cost associated with the passenger time lost, in terms of schedule buffer, flight delays and cancellations, as well as missed connections. In a report prepared by Schumer and Maloney (2008) for the Senate Joint Economic Committee, the total direct costs to airlines and passengers were estimated to be $31 billion dollars. In 2009, the NAS delays were 30.6 percent of total delays, and 65.7% of NAS delays were due to weather.

1.1 Importance of Equitable Resource Allocation

To alleviate the current air traffic congestion problem, a number of initiatives have been developed to improve the overall air transportation system efficiency by providing additional infrastructure and facilities, such as runways at airports. Additionally, several planning and management initiatives also focus on the efficient use of the existing airport and airspace resources, such as flights separation, integrated weather prediction, as well as dynamic resource allocation. In particular, an air Traffic Flow Management (TFM) program aims to balance air traffic demand and ensure the maximum efficient utilization of the NAS within available capacity. The system-wide balance is accomplished by first predicting the impact of demand and capacity constraints and then responding as needed with flow management strategies.

Essentially, planning and operating in a collaborative environment requires mutual understanding and acceptance of respective roles and responsibilities among the NAS users. As a result, Collaborative Decision Making (CDM), one of the above key initiatives and a joint government/industry partnership, seeks to create common situational awareness of traffic congestion and constraints in the NAS. The first major thrust of CDM in the United States, Ground Delay Program Enhancement (GDPE), which targets airport arrival slot control, has been operated since 1998. When airport arrival capacity is reduced and may not meet the demand placed by arriving aircrafts, the Federal Aviation Administration (FAA) enacts a Ground-Delay Program (GDP) to delay flights before they depart from their origin airports, keeping traffic at an acceptable level for the affected arrival airport(s). Under GDPE, participating airlines send operational and schedule changes to the Air Traffic Control Systems Command Center (ATCSCC) on a continual basis. Using the flight schedule monitor (FSM) tool, the ATCSCC collects various types of information, monitors airport arrival demand and initiates GDPs at the major airports in the U.S.

As a resource allocation mechanism, Ration by Schedule (RBS), in conjunction with a slot exchange procedure, namely "compression", constitutes the arrival slot allocation process on which FSM is based. This procedure relies on the flight schedule to define an allocation standard to measure the degree of equity for each NAS user. Moreover, this procedure is implemented independent of flights' current status, encouraging all users to provide and exchange up-to-date information. RBS and "compression" strategies have been reported to significantly reduce delays and improve the efficiency of air traffic flow into airports. For example, during the period between January 20, 1998 and July 15, 1999, the planned (ATCSCC assigned) delay reduction (at airports with 10 or more compression cycles) ranged from 7.5% at Atlanta's Hartsfield Airport to 18.2% at Boston's Logan Airport (Ball et al. 2000), with an average reduction over all GDP airports being 12.7%. Compared to the previous system, these two strategies reduced assigned ground delay by over 3.1 million minutes.

The success of the CDM program has underscored the need for TFM to allocate resources in a fair-handed manner. As a service provider, the FAA has a strong commitment to an equitable allocation of the limited NAS resources. The equitable allocation is essentially an inherent responsibility of traffic flow management. More importantly, inequitable allocation could affect the active participation in the daily management of air traffic. Information being withhold or skewed could develop mistrust, and further jeopardize the quality and effectiveness of air traffic services, and the efficiency of the NAS as a whole. Thus, equitable allocation is critically needed to ensure successful deployment of existing or new TFM initiatives under the CDM paradigm, and it is also a key issue to be carefully examined and explored in this dissertation.

1.2 Research Motivations for Air Traffic Flow Management

To date, the potential benefits of utilizing advanced air traffic flow management strategies to consider both efficiency and equity objectives are still being explored. As shown below, many fundamental issues need to be addressed to fulfill the methodological capabilities required by the collaborative NAS decision making environment. These challenging questions place a greater need for systematic modeling methodologies for potentially competing objectives and efficient solution algorithms for real-world problems.

The first part of this dissertation is motivated by the efficiency and fairness concerns that arise from the resource allocation procedure in the airspace. Different from the resource allocation problem at the airport such as a Ground Delay Program (GDP) as stated earlier, it is challenging to assign airspace resources efficiently and equitably with respect to competing airlines or origin-destination pairs. In a GDP, all the flights have the same destination, and a flight schedule is used to define an allocation standard so as to measure the degree of equity for each NAS user. In the airspace congestion problem, however, there is no schedule which can be used to measure the degree of delay, leading to a number of modeling and operational difficulties. First, flights do not have a fixed flight plan to define the path of the flight within a set of sectors, fixes and jet routes. Usually, a flight is required to file the flight plan 45 minutes before its departure. Furthermore, different airlines have different patterns of filing the flight plan, with some airlines filing flight plans at the last minute while waiting for the final weather forecast. As a result, it is extremely difficult to accurately predict flight plans to be filed in the near-term future (say 2 hours ahead). Secondly, even assuming that a flight plan is fully predictable based on historical information, flights could intersect the congested airspace differently due to their own specific origin-destination geography.

This dissertation will consider the following theoretically important issues and provide new reformulations to improve both system-wide efficiency and equity when a Flow Constrained Area (FCA) is issued. Rerouting options will be provided so as to reduce the traffic to capacity level:

1. How to choose flights to be offloaded: Typically, flights between some specific city/center pairs are chosen to be rerouted. This practice might lead to a significant bias among airlines as some airlines might get exempted simply due to the fact that no flights were scheduled to arrive at the chosen destination airports.

2. More options than just rerouting: The ground holding option in which aircraft can be held at the departure airport is not provided as an alternative for rerouting in the current procedure.

3. Real-time operation adjustment under uncertainty: The offloaded flights are chosen statically two or more hours before the events occur without getting adjusted according to the evolving actual conditions. It is very difficult to predict the exact value of downgraded sector capacity, especially under evolving weather conditions.

1.3 Research Motivations for Airport Slot Assignment

Airports subject to slot controls have a restricted number of scheduled operations per day, in which “slots” are defined as a reservation for a flight to takeoff or land within an assigned time interval. In addition, airports have operational constraints determined by runway size, the number of terminals, and air traffic control facilities. As the demand for an airport approaches and in some cases exceeds capacity, significant flight delays could result. Therefore, it is important to assure slots at congested airports are allocated among airlines in an economically efficient manner.

One type of resource assignment problems arises in the long-term landing slot lease assignment practices, which aim to solve the demand/capacity imbalance by restricting schedules. Many slot control rules are designed and used to address increased congestion and delay that would likely occur in the absence of restrictions on the number of aircrafts scheduled to fly in and out of a major airport. Recently, the FAA and the DOT have also taken a number of steps to investigate market-based solutions in congestion management which aims to encourage competition and also allow the airports to operate at maximum efficiency and safety.

This dissertation plans to mathematically formulate the above proposed problems and develop efficient solution algorithms to assign scarce resources in terms long-term airport slots. The primary objective of this dissertation is to develop efficient algorithms to solve the slot assignment problems with systematic consideration of airlines' need and equity in the final assignment.

1.4 Dissertation Outline

The focus of this dissertation is on constructing theoretically rigorous models to effectively allocate scarce resources in the national airspace so as to balance the system capacity and airline economic tradeoff. We will formulate and develop mathematic models to describe different alternative approaches that address the flight offloading problem with special focuses on the problem complexity, capacity-demand interaction, and equity issues. The contents and contributions of each subsequent chapter are detailed below.

The dissertation includes six chapters. Chapter 2 provides a comprehensive review and discussions on air traffic management. Several air traffic management initiatives are briefly reviewed, two of which are discussed in details. The last section of Chapter 2 reviews important literature on models for air traffic flow management and airport slot assignment.

Chapter 3 develops a sector-level integer programming model to systematically formulate the flight offloading problem. A bi-criteria optimization model is proposed to divert excess demand from different competing airlines in the congested airspace when the predicted traffic demand is higher than available capacity. Computationally efficient network flow models with side constraints are developed and extensively tested using datasets obtained from the Enhanced Traffic Management System database.

Chapter 4 proposes an enhanced sector or space level model with ground holding and routing decisions. Given a network of airspace sectors with a set of waypoint entries and a set of flights belonging to different airlines, this study considers uncertain sector capacity using multiple scenarios. The proposed stochastic optimization model aims to minimize the total expected flight ground holding and rerouting cost subject to a set of flight routing equity, operational and safety requirements. A time-dependent network flow programming formulation is proposed with sector capacities and rerouting equity for each airline as side constraints. A Lagrangian relaxation based method is used to dualize these three side constraints and decompose the original complex problem into a sequence of single flight rerouting/scheduling subproblems.

Within a multi-objective utility maximization framework, Chapter 5 proposes several practically useful heuristic algorithms for the long-term airport slot assignment problem. Alternative models are constructed to decompose the complex model into a series of hourly assignment subproblems. A new paired assignment heuristic algorithm is developed to adapt the round robin scheduling principle for improving fairness measures across different airlines. Computational results are presented to show the strength of each proposed modeling approach.

The dissertation is concluded in Chapter 6 by a summary of research contributions and discussions of future research needs.

Chapter 2: Background Introduction and Literature Review

This chapter reviews several Air Traffic Flow Management (ATFM) initiatives, as well as critical literature on the specific problems under consideration in this dissertation. The following briefly introduces commonly used ATFM strategies such as Ground Delay Program and Collaborative Routing in Section 2.1 and Flow Constrained Area and Airspace Flow Program in Section 2.2. We offer detailed discussions on the current practice from the perspectives of system efficiency and air carrier equity, which will be studied further in Chapter 3 and 4. In Sections 2.3, we review critical optimization literature in the Ground Delay Program and Air Traffic Flow Management in general, with a focus on various formulations that are relevant to the equitable air space and slot resource allocation.

2.1 General Air Traffic Flow Management Strategies

This section first introduces the administrative structure for the control and coordination of aircrafts in the NAS, which is provided by Air Traffic Control (ATC) and Air Traffic Flow Management. Specifically, ATC is responsible for ensuring safe separations between aircraft, and ATFM is responsible for balancing demand and capacity to ensure the efficient use of the airspace. In general, ATC is a service provided by ground-based controllers who direct aircraft on the ground and in the air. A controller's primary task is to separate aircraft sufficiently with the use of lateral, vertical and longitudinal separations. Secondary tasks include ensuring safe, orderly, and expeditious flow of traffic and providing information to pilots, such as weather and navigation information. The Air Traffic Control system has limited capabilities due to many factors, such as the volume of traffic, frequency of congestion, quality of radar, controller workload, and higher priority duties. In contrast, ATFM focuses more on the system flow side of air traffic management by coordinating air traffic so that demands for various resources do not exceed capacities. ATFM is performed on a national level at the Air Traffic Control System Command Center (ATCSCC). The primary duty of the ATCSCC is to monitor the traffic situation in the NAS, and implement control measures when demand exceeding capacity. We will briefly review several ATFM strategies that are currently used in handling demand and capacity issues, and then focus on two major problems of Air Traffic Flow Management in section 2.2.

Essentially, a Ground Delay Program aims to solve airport arrival capacity shortfalls by applying ground delays to flights at their origin airports when they are bound for a common destination airport with reduced capacity. Interested readers are referred to Ball et al. (2007), Hoffman et al. (2011), and Libby et al. (2005). For allocation purposes, the time horizon of reduced capacity is divided into contiguous time intervals known as arrival slots. Prior to departure, each flight receives a discrete arrival slot based on availability at the destination. The Collaborative Decision Making (CDM) program has established a highly successful paradigm for allocation of airport arrival slots. The main allocation principle is “first-scheduled, first-served”, meaning that the earlier arrival slots are generally awarded to the flights that are scheduled to arrive earlier. The CDM experience has shown to be not only an equitable treatment of carriers advisable, but a necessary condition for efficient use of resources. Prior to CDM, effective GDP initiatives were based on dated flight data that unfortunately did not reflect the airline's intentions upon the day of operation.

Collaborative Routing is an approach to apply CDM technology and concepts to the management of en route traffic. In contrast to the highly refined algorithms employed in GDPs, the resource allocation problem for en route traffic has been less studied. A number of initial Collaborative Routing tools and procedures were prototyped in 1999. A collaborative routing coordination tool, CRCT, developed by MITRE, provides FAA traffic flow specialists with automated features that support the identification of flights affected by congestion and aids in the development of alternative routes. Other tools have been developed to support Collaborative Routing such as, Collaborative Convective Forecast Product (CCFP) which represents a consensus based on information from AOC and ARTCC weather units, Low Altitude Arrival and Departure Routes (LAADR) which contain a set of procedures for allowing the use of low altitude alternative routes to avoid congestion, and Coded Departure Routes (CDR) which providing a set of procedures and database for creating and storing alternative routes

Miles-in-Trail Restriction (MIT) aims to ensure that the traffic flow does not exceed the capacities in the en route sectors or congested regions in the NAS by imposing distance based metering or restrictions at different fixes. MIT restrictions keep the traffic flow below a certain level by specifying the minimum separation distance between two consecutive aircrafts flying across the same fix. MIT restrictions often cause airborne delay, but it is still less expensive and disruptive than airborne holding.

2.2 Flow Constrained Area and Airspace Flow Program

To improve overall airspace system performance, recent attention has shifted to the en route airspace, with the desire to most significantly improve airborne delays and throughput enhancement. It is well known that the primary factors causing congestion in the airspace include severe weather (especially in the summer), heavy traffic volume, and special use restrictions such as military activities and space rocket departures. In practice, all the above cases are classified as “lost space”. When any of the above events occurs, a constrained airspace problem would arise and would need to be resolved in real-time to maintain the safety for passing flights.

2.2.1 FCA in practice

Prior to 1998, the FAA dealt with the air traffic flow management problem in a centrally controlled manner with little airline involvement. In recent years, the Flow Constrained Area system was designed to evaluate and alleviate potential adverse effects to air traffic during periods when events may have a significant impact on the NAS. It provides a mechanism of automated data transfer and enables a common situational awareness to air traffic personnel and NAS users, who can receive advanced notifications of problem areas and have a chance to take proactive actions to prevent congestion.

Compared with the traditional air traffic flow management model, which takes into account a large piece of the airspace or even the entire NAS, the FCA strategy is a type of local and constraint-targeting solution. Instead of considering the entire airspace, an FCA targets the congested area and removes a portion of flights to ensure that the total number of flights does not exceed the reduced capacity.

The common NAS resources include airspace (sectors), fixes, and airports, each with limited capacity and subject to excess demand. In general, the goal of the FCA operational policy is to solve the congestion problem locally by offloading excess flights from the problem area and achieve the demand-capacity balance. The offloaded flights can be canceled or re-routed to the surrounding areas that have spare capacity. Another overarching goal of FCA is to solve the problem promptly and efficiently because typically, only a few hours are available to make a strategic decision once the event occurs. Although the current FCA operation still needs to be modified to improve system performance and fairness among airlines, the approach of focusing on constraints and generating problem based on the system constraint offers more flexibility in practical applications.

It should be remarked that, a Flow Evaluation Area (FEA) advisory is similar to FCA in that they both define the constrained area in a given time period with an attached flight list. More precisely however, the former only recommends actions for airlines, while the latter requires the airlines to comply with the issued advisory.

The current FCA operational procedure can be stated in details as the following with the corresponding flow chart is shown in Fig. 2.1 (Libby et al., 2005).

1) *Situation monitoring*. Traffic managers and NAS users monitor the situation to be aware of potential constraints and of potential responses to FAA Traffic Flow Management Unit personnel.

2) *Publishing FEA*. Traffic managers and NAS users send a request to the ATCSCC once they detect problems that might potentially become constraints.

3) *FEA reactions*. If a public FEA is published, traffic managers at ATRCC and other NAS users react to the potential constraints.

4) *Publishing FCA*. Once the ATCSCC recognizes that a constraint exists, it will declare it, e.g., by issuing an FCA advisory. At the same time or perhaps later, the ATCSCC will also provide route options around this FCA, but NAS users are left to implement them.

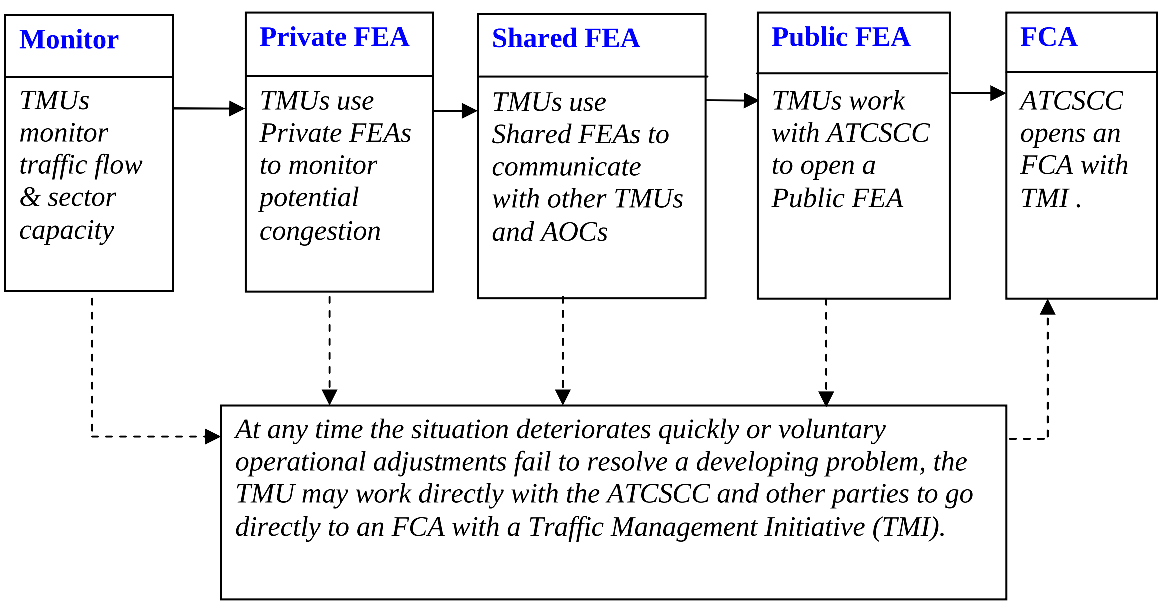


Figure 2-1 FCA Operation Flow Chart (Source: FCA Operational Concept document prepared by Libby et al., 2005)

Weather and traffic volume are currently the most significant reasons for defining FCAs. On a severe weather day, such as July 29th, 2003 shown in Figure 2.2, an FCA was defined as a 3-dimensional airspace filter with detailed coordinate information attached in a different file other than the advisory. Rather than being defined as the entire congestion area, where the true constraints exist, the FCA is currently used and defined as a small piece of airspace which is used to identify and filter out the flights that need to be offloaded or re-routed. As a general rule, a proportion of, or in some cases, all the flights scheduled to pass through this filter area during the FCA, need to be offloaded to maintain their en route safety.

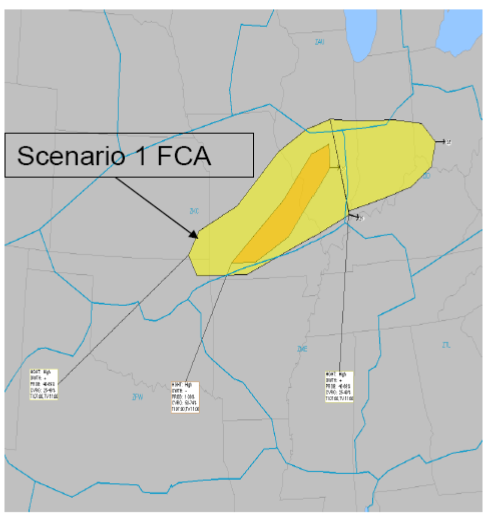


Figure 2-2 Illustration of Flow Constrained Area

Currently, flight departures from some ARTCCs to some major destinations are often chosen to implement rerouting policies due to tactical considerations. Since the majority of the flow usually goes to major airport, targeting those flights for rerouting can easily solve the problem. Of course, other considerations are also involved, such as wind direction and current time. For Example, most of the flights fly from east coast to west in the morning and from west coast to east in the afternoon. If the FCA is defined in the morning, offloading some east-west bound flights flow can solve the problem easily.

In some cases, excess volume is the principal cause of capacity-demand imbalance. For these cases, FCA is defined as a 3-dimensional airspace filter with an associated time interval. Some of the flights will be restricted to pass through using certain routes, e.g. chokepoint routes. Similar to the case discussed earlier on weather FCAs, the flights are usually selected by taking departures from certain ARTCCs to some major airports to maintain the sector capacity/demand balance.

The above FCA scenarios summarize how the FCA is currently being used (as is envisioned to be used). Some important issues, however, remain unsolved in the current ad-hoc type of operations.

(1) The current operational procedures typically use a filter area to show the set of flights to be offloaded or rerouted. However, the true congested or constrained area is not revealed in the published advisory. That is, only the solution determined by ATCSCC is shown to the traffic unit personnel. The airline operational control center personnel and the ARTCC personnel are unable to recognize the true constraints and may not respond or cooperate with the published advisory. This may affect the compliance rate and may introduce more workload to ensure the problem is solved.

(2) Typically, flights between some specific city/center pairs are chosen to be rerouted. The current operational procedure introduces offload/reroute bias among the carriers. For instance, large carriers may have more flights going to a major airport,

while some smaller carriers may not have any flights heading to that airport. This may lead to a significant bias among airlines as some airlines might get exempted simply due to the fact that no flights were scheduled to arrive at the chosen destination airports.

(3) The ground holding option in which aircraft can be held at the departure airport is not provided as an alternative for rerouting in the current procedure. For those flights that can be easily held on the ground for a short time to avoid the traffic jam, ground holding might be a much better option than being rerouted around the constraint area resulting in more fuel consumption and flying time.

(4) The offloaded flights are chosen statically two or more hours before the events occur with no adjustment according to the real-time situation development. The issue here is that more or less flights than actually needed can be selected to be offloaded. If flights are not offloaded enough, should FAA adopt a more dynamic strategy to consider and implement real-time offloading and rerouting? If more flights are offloaded, and there is still remaining capacity left, what criterion can be taken to assign the rest of the resource to the airlines?

In summary, the essential question is that how FCA strategy could be used to solve the demand/capacity imbalance more efficiently and fairly compared to the current implementation. In this dissertation, Chapters 3 and 4 proposed two models to address the above 4 issues in more details, aiming to offer possible improvement of the current process.

2.2.2 Airspace Flow Program

The FAA has developed a number of tools to deal with different Traffic Flow Management problems. When convective weather reduces capacity somewhere in the airspace, the FAA can define a portion of the airspace to be a Flow Constrained Area. TFM tools can then identify flights expected to pass through the FCA so some portion of the flights can be routed around the problem. Often though, rerouting flights is not sufficient to address extended capacity reductions in the airspace in an FCA advisory and the need for additional tools has long been recognized. Essentially, the Airspace Flow Program combines the power of GDPs and FCAs to allow more efficient, effective, equitable, and predictable management of airborne traffic in congested airspace, and it could be viewed as an extended FCA program with ground delay as major offloading option.

When TFM specialists at the ATCSCC decide that the weather conditions are appropriate, they can plan and deploy an AFP. The first step is to use a tool, e.g. traffic situation display, to examine predicted weather and traffic patterns and identify the problem area by creating an FCA. Secondly, the Enhanced Traffic Management System takes the FCA description and produces a list of the flights that are expected to pass through the FCA and the time they are expected to enter. This list, updated with fresh information every five minutes, is sent to the flight schedule monitor, which displays the projected demand in a number of formats designed to support effective planning. The TFM specialists at the ATCSCC can enter the capacity of the FCA, expressed as the number of flights that can be managed per hour or per certain time interval, and FSM will then assign each flight a controlled departure time that will provide a smooth managed flow of traffic to the FCA. These departure times are sent to the customers for their planning and to the towers at the departure airports for enforcement.

The principal goal for the initial deployment of the AFP program is to better manage en route traffic during severe weather events. Compared to current approaches such as GDP or FCA in certain scenarios, AFP will reduce unnecessary delays while providing better control of demand and more flexibility for customers. Furthermore, the AFP gives more flexible solutions than an FCA program. Although AFP improves FCA with more flexible options, there are certain limitations to the current procedure. Moreover, as the 'slot' in AFP is not the same as the regular slots in GDP, equity issues in allocating ground delay would arise. How to assign ground delay as well as reroute among the flights equitably is a big challenge in AFP. Chapter 4 will discuss some proposed models to resolve the two issues discussed here.

2.3. Literature Review

2.3.1 Deterministic and Stochastic Ground Holding Problems

The airport congestion problem, which is caused by too many flights attempting to take off or land relative to airport capacity, has been extensively studied by many researchers in the last few decades. Odoni's (1987) systematically defined a Ground Holding Problem (GHP) in ATC which marked the start of a significant research effort on single-airport and multi-airport versions of the problem. Andreatta and Romanin-Jacur (1987) addressed the one-airport congestion problem for a single time period, and their model aimed to optimize the total expected delay cost and a polynomial solution algorithm was derived. Terrab and Odoni (1993) presented an exact solution algorithm for problems involved with one airport, multiple time period and deterministic capacity. The static multi-airport ground-holding problem was studied by Vranas et al. (1994) through introducing generic integer programming models, which assign optimal ground holding delays in a general network of airports to minimize the both ground and airborne delay cost of all flights. Navazio and Romanin-Jacur (1998) presented integer programming models for a set of airports, taking into account the operations dependency of the star-shaped airport networks.

Most of optimization models for GDP involve the construction of <u>space-time networks</u>. That is, the time horizon of interest is decomposed into a discrete set of time intervals, and various spatial components (such as airports, sectors, and waypoints in general) are modeled using a time-expanded structure. Typically, the basic flow variables x_{fte} in a standard space-time model represents flight *f* occupying spatial element *e* during time interval *t*, and these variables are subject to the fundamental flow balance and capacity constraints in the form of $\sum_{f} x_{fte} \le CAP(t,e)$, where $CAP(t,e)$ is the flow capacity of element *e* at time *t*. Alternatively, Bertsimas and Stock's (1998) model introduce cumulative flow count variables w_{fte} through a simple linear transformation $w_{fte} = \sum_{\tau=1}^{t} x_{f\tau e}$, where w_{fte} represents if flight *f* arrives at spatial element *e* by time *t*. This cumulative flow count representation enables many additional modeling features, such as propagating travel times along routes. Hoffman and Ball (2000) constructed several models of the single-airport ground holding problem with banking constraints, accommodating the hubbing operations of major airlines. In particular, by examining the strength of different formulations, they offered the following important remarks. Both representations of using variables x_{fte} vs. w_{fte} have the equivalent Linear Programming (LP) strength, but the cumulative flow count w_{fte}-based representation requires an additional set of non-negative flow constraints in terms of $x_{fte} = w_{fte} - w_{f(t-1)e} \geq 0$, which requires more iterations in solving the LP relaxation than the flow variable x_{fte} based representation.

Focusing on GDP planning under uncertainty, Richetta and Odoni (1993) provided a linear programming reformulation to problems with one airport, multiple time periods and stochastic airport capacities. A set of coupling constraints are needed in this case to ensure unique flow assignment solutions across different scenarios, that is,

$$x_{fte}(q=1) = x_{fte}(q=2) = \cdots = x_{fte}(Q)\,, \tag{2.1}$$

where scenario index $q = 1, 2, ..., Q$. This set of coupling constraints can be viewed as a special case of nonanticipativity (NAC) constraints for constructing deterministic equivalents to the scenario-based stochastic optimization models.

Ball et al. (2003) developed a stochastic integer program with dual network structure and showed its application to the ground-holding problem. The dual network structure can be viewed as a special case of the compact representation approach for modeling NAC, while the resulting coefficient matrix in the dual network was shown to have a desirable total unimodularity feature that leads to efficient network flow algorithms. To further effectively deploy stochastic programming methods in practice, Ball and Lulli (2004) proposed a simple exemption policy to help mitigate uncertainty for ground delay programs. Mukherjee and Hansen (2007) introduced a scenario tree-based stochastic optimization model for the GDP, and each scenario corresponding different capacity conditions based on weather forecasts at sequential decision stages.

Aiming to provide alternative resource allocation methods to widely accepted RBS method, Pourtaklo (2009) studied the problem of fair allocation of limited resources in the context of an Airspace Flow Program. To determine a fair share of available airspace resources among flight operators, a preference based proportional random allocation method is developed to ensure the slot assignment to each is close to their fair shares and expectations. In addition, she also presented new resource rationing principles to improve resource assignment fairness and efficiency, through considerations of slot values and dual pricing. Churchill and Lovell (2012) developed a two-stage stochastic integer programming model for coordinated aviation network resource allocation under capacity uncertainty, and two types of consistency constraints were proposed to ensure the feasibility and compatibility between the first and second stages decisions of resource allocation.

2.3.2 Air Flow Management and Flow Constrained Area

Optimization in Air Traffic Flow Management has received significant attention in the past 30 years. Classifying by applications, there are two major categories in this area: 1) optimization models that account for airport take-off and landing capacities only, 2) models that account for both airport arrival/departure and en route capacity constraints. Most research on Ground Holding Problem fall into the first category. In contrast, research in airspace congestion with arrival fix constraints will generally need to consider both airport and airspace/en route capacities, as both resources are subject to certain capacity reduction. In this section, an overview of the published literature on optimization models in ATFM will be discussed.

To improve overall airspace system performance and reduce the congestion affecting en route airspace, optimization models involving en route capacity were developed by Lindsay et al. (1993) and Tosic et al. (1995). Deterministic optimization models considering both airport and en route capacity constraints were formulated as multi-commodity network flow problem by Helme (1992). Using cumulative flow count variables w_{fte} to represent if flight *f* arrives at spatial element *e* by time *t*, Bertsimas and Stock (1998) formulated disaggregate deterministic integer programming models for deciding the departure time and route of individual flights. Using space-time flow variables x_{fte} to represent if flight *f* occupies spatial element *e* during time interval *t*, Bertsimas and Stock (2000) proposed a dynamic multi-commodity network flow model to consider both routing and scheduling decisions, but it produces non-integer solutions for even small scale problems. Therefore, they suggested a number of heuristics (such as random rounding and solving an integer packing problem) to obtain integer solutions. Although both formulations produce non-integer solutions from LP relaxation, the latter model achieves integrality in many more instances compared to the former.

A number of studies have been conducted to extend the model proposed by Bertsimas and Stock (1998). Alonso, Escudero and Ortuno (2003) proposed a stochastic 0-1 program to consider random capacity at different scenarios, and a minmax function was introduced to reformulate the nonanticipativity constraint in the objective function. Along this line, Agustín et al. (2012) further developed a deterministic equivalent of the stochastic mixed 0-1 program with full recourse for the multi-stage ATFM problem, and a compact representation approach was used to handle NAC. A recent dissertation by Chang (2010) proposed a Lagrangian relaxation approach to dualize a number of equalities corresponding to NAC, such as

$$x_{fte}(q=1) = x_{fte}(q) \qquad \forall q = 2,...,Q\,. \tag{2.2}$$

A subgradient method is used to adjust the Lagrangian multipliers associated with (2.2), and a very tight solution quality gap was reported between the Lagrangian-based lower bounds and upper bounds generated from a rolling horizon method.

Lulli and Odoni (2007) presented a deterministic multi-commodity optimization model for the European ATFM problem in a space-time network with en route sector and airport capacity constraints.

Rios and Ross (2010) applied a parallel Dantzig–Wolfe decomposition technique to relax the capacity constraints in the Bertsimas and Stock (1998)'s model where flight trajectory-based subproblems were constructed and solved simultaneously. Motivated by the hydrodynamic theory for highway traffic flow, a large-capacity Cell Transmission Model was proposed by Sun and Bayen (2008) in order to model high altitude air traffic flow. Sun et al. (2011) recently developed a dual decomposition method to relax the sector capacity constraints in their aggregated traffic flow model, which leads to relaxed linear programming problems with better computational efficiency.

Bertsimas, Lulli and Odoni (2011) extended the Bertsimas and Stock's (1998) model to consider additional re-routing options, speed control and airborne holding options, and three classes of valid inequalities were presented to strengthen the polyhedral structure of the underlying relaxation. Using a scenario-tree and cumulative flow count based representation, Mukherjee and Hansen (2009) developed a general stochastic programming model to allow dynamic flight rerouting decisions under stochastic capacity. Liu et al. (2008) examined several methods to classify capacity profiles into a small number of nominal scenarios for constructing representative scenario trees. Ganji et al. (2009) presented a two-stage stochastic program that aims to optimize the first-stage flight rerouting plan in a FCA while considering the time of capacity windfall as a random variable.

2.3.3 General Scheduling Methods and Equity-related Models

In airline industries, commercial airlines need to present their services to passengers through published schedules between select city-pairs, and each underlying flight schedules is comprised of flight legs between airport locations. Ball et al. (2007) provided a detailed survey on air transportation under irregular operations and related control strategies. Beatty et al. (1998) analyze delay propagation, as a perturbation in the timing of one flight leg can have significant “downstream" effects leading to delays on several other legs. Considering recurring and nonrecurring delay conditions, a number of research tools and commercial packages (e.g. Niznik, 2001 and Yu, 1997) have been developed in response to the strong economic incentives to further improve the overall system performance. Over the past 15 years both in the U.S. and Europe, a growing range of CDM-based decision support systems have been prototyped and deployed to “optimize" the relationship between an air navigation service provider and the flight operator, e.g., Ball et al. (2001), Chang et al. (2001), Richetta and Odoni (1993).

In addition to the studies focusing on network-wide scheduling and routing options in the previous section, there are a number of optimization models that consider the air traffic flow management problem at a more microscopic level. ATC needs to ensure that flights crossing a sector are safely separated and the flights arriving or departing the runway of a certain airport also satisfy the separation standard. Bianco and Bielli (1993) proposed different network models for ATC that determine traffic flow measures for both before and after flight’s departure, including ground delays, queue at holding points, etc. Barbosa-Povoa et al. (2001) proposed a bipartite directed network model to address the grouping and scheduling of ATC sectors. Their model takes into account controller availability and sector capacities so as to minimize delay cost. Vranas et al. (1994) proposed optimization models to allocate tactical ground delays for flights crossing different congested airspaces in Europe. Goodhart (2000) developed disaggregate deterministic models for ATFM, in which airline’s priorities on various flights are accommodated. Churchill, Lovell, and Ball (2010) studied the impact of flight delay propagation (due to degraded airport and airspace capacity) on strategic air traffic flow management. In order to further characterize the sensitivity of ATFM models to uncertainty in various capacity parameters, Churchill and Lovell (2011) proposed a modified Monte Carlo framework to assess the impact of stochastic capacity variation on coordinated air traffic flow management.

The notion of equity has been examined in a number of ATFM contexts. Vossen (2003) (See also Vossen (2002)) proposed an optimization model for mitigating bias of flight exemptions during a GDP and showed that it could reduce systematic biases that exist under current procedures. To comprehensively consider and reduce the effect of uncertainty in weather forecasts, Ball, Hoffman and Mukherjee (2010) recently developed methods that tradeoff the efficiency benefits and the loss in equity.

Vossen and Ball (2006) analyzed the ration-by-schedule (RBS) method, and showed it embodied certain fair allocation principles. Ball, Donohue and Hoffman (2005) provided a systematic discussion of different aviation-related market mechanisms, which allows better modeling of safe, efficient and equitable allocation of limited airspace system resources. Focusing on a real-time version of compression, Ball et al. (2005) presented various response mechanisms for dynamic air traffic flow management. Recently, a second transaction-oriented version of compression called adaptive compression has been implemented by Federal Aviation Administration in 2008. Specifically, each slot credit substitution (SCS) transaction is initiated by an airline, adaptive compression transactions are initiated by FAA's Enhanced Traffic Management System.

Through extensive experimental results for the European ATFM model, Lulli and Odoni (2007) also highlighted an important trade-off between efficiency and fairness in the network case, where the solution with maximum efficiency might be disadvantageous to certain classes of users while solutions focusing on fairness may lead to system-wide inefficiency.

By considering flight safety, air traffic control, and airline equity constraints, Sherali, Staats and Trani (2003) and (2006) developed a large-scale airspace-planning and collaborative decision-making (APCDM) model that aims to select a set of flight plans in an airspace region. Their equity measure is expressed as a relative performance ratio. Sherali et al. (2011) further incorporated slot exchange mechanisms in their model and extended an on-time performance equity measure (for each flight plan) from Vossen and Ball (2006) to construct collaboration efficiency functions.

Chapter 3: Flight Offloading Problem in Congested Airspace

The common National Airspace System (NAS) resources include airspace (sectors), fixes, and airports, each with limited capacity and subject to excess demand. This chapter focuses on a fundamental problem of off-loading volume of airspace subject to capacity constraints. In a particular example of airspace flow management, once a Flow Constrained Area (FCA) is issued, the decision makers need to solve the congestion problem locally by offloading excessive flights from the problem area so as to achieve the demand-capacity balance. The offloaded flights can be canceled or rerouted to the surrounding areas with available spare capacity.

In a typical centralized management procedure, the FAA sends out increasingly severe warnings and/or advisories, starting from recommended movements and ending with required offloads and reroutes. The recommended actions can be issued to the traffic managers in the Air Route Traffic Control Center (ARTCC) and Airline Operational Centers (AOC) using a Flow Evaluation Area (FEA) advisory. As previously mentioned in Chapter 2, the current FCA approach does not consider the entire congested area nor equity among air carriers when choosing offloaded flights. In this chapter, a model is proposed to address these two issues.

This chapter is organized as follows. Section 3.1 presents several general integer programming formulations for the airspace flight offloading problem, followed by alternative network programming-based reformulations in sections 3.2

and 3.3. The related computational comparisons between different models are discussed in Section 3.4.

3.1 Integer Programming Model for Airspace Offloading Problem

The airspace offloading problem can be described as follows. When an FCA is issued, the 3-dimensional volume of congested airspace needs to be specified accordingly, along with an impacted time interval (defined in terms of start time T^S and end time T^E), in order to identify a list of flights subject to capacity constraints. For each airspace sector *j*, a flight *f* is defined as an involved flight when it has an entering time $e_{f,j}$ and a leaving time $l_{f,j}$ with $T^S \le e_{f,j} < T^E$ or/and $T^S < l_{f,j} \le T^E$. By definition, the involved flight set contains all such flights. The following model would take the involved flights as input and produce a list of flights to be rerouted and listed as attached flights for the FCA.

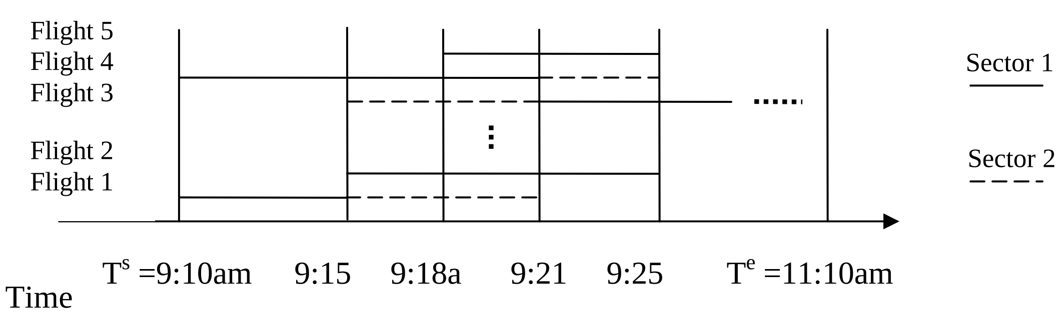


Figure 3-1 Example of impacted flights in a congested airspace

As shown in Figure 3.1, the entire impacted time period in a congested airspace is divided into small time intervals according to the flights' entering/leaving times for each sector. Take flight 4 for instance, it enters sector 1 at 9:10am and transfers to sector 2 at 9:21am. At 9:25am, it leaves this entire area. Time stamps of 9:10am, 9:21am and 9:25am are recorded, corresponding to the events that involve entrance or exit of flights. For each time interval between these time instances, a constraint is imposed to guarantee that the total (simultaneous) number of flights does not exceed the capacity. It should be noted that the actual airspace sector capacity is quite complex in its own right, as it also depends on a number of highly dynamic factors, such as the route structure and controller's capability. Interested readers are referred to discussions of the "dynamic density" in Masalonis et al. (2003) and Davison et al. (2003). Without loss of generality, this chapter considers sector capacity constraints only on the instantaneous number of flights.

Another important consideration is equity among air carriers. To balance the demand capacity in the congested airspace, each air carrier needs to remove some flights. To avoid delay cost due to flight rerouting, an air carrier obviously wants to keep as many flights on their original routes as possible. Since these airlines have different numbers of flights to be considered for rerouting, it is desirable to allocate the rerouting requests evenly among those air carriers. Mathematically, each agent (i.e. airlines) in the collaborative decision-making problem likes to experience a similar offloading percentage. As a result, an equity constraint is introduced in this research for each air carrier, so as to control or minimize the deviation of each air carrier's offloading percentage from the overall percentage for the entire congested airspace.

The notation, decision variables, and objective function of the proposed model are described as follows.

Notation:

T^S starting time of impact time period;

T^E ending time of impact time period;

I total number of involved air carriers;

J total number of involved sectors;

F total number of involved flights;

i air carrier index, where i =1, 2, …,I;

t time interval index, t = 1, 2, …T, where $T = T^E$-T^S;

j sector index, j = 1, 2, ..J;

f flight index, f = 1, 2, …F;

A set of involved flights;

A_i set of involved flights for air carrier i;

$e_{f,j}$ originally scheduled entering time of flight f on sector j;

$l_{f,j}$ originally scheduled leaving time of flight f on sector j;

$U(j,t)$ set of impacted flights at sector j at time t, where a flight belongs to $U(j,t)$ when

$T^S \le e_{f,j} < T^E$ or $T^S < l_{f,j} \le T^E$;

c_j reduced capacity of sector j during the FCA time interval;

q_f total flying/travel time in the congested airspace during specified time interval for each flight f;

δ_f extra distance for flight f if it is rerouted;

Decision variables:

x_f binary variable, $x_f = 1$, if flight f passes through the FCA using original schedule, 0 otherwise.

r_i rerouted flight percentage for air carrier i, $r_i = \dfrac{\sum_{f \in A_i} (1 - x_f)}{|A_i|}$

$\overline{r}$ average offloading percentage across all airlines, $\overline{r} = \dfrac{\sum_{f \in A} (1 - x_f)}{F}$

With decision variable x_f representing the routing decision for each flight, variable r_i is introduced as an air carrier-based index for capturing its overall offloaded percentage. Below are a number of possible objective functional forms available to take into account the equity consideration.

Possible objective functions:

$$\text{Max} \sum_f \left(w_f x_f \right) \tag{3.1}$$

$$\text{Min} \sum_{i=1}^{I} \left| r_i - \overline{r} \right| \tag{3.2}$$

$$\text{Min} \sum_i (r_i - \overline{r})^2 \tag{3.3}$$

$$\text{Min Max}_i \left(r_i - \overline{r}\right) \tag{3.4}$$

The first objective function (3.1) focuses on the system-wide efficiency, and the weights w_f in Eq. (3.1) can be determined according to specific traffic management goals. For example, by setting an equal weight of $w_f = 1$, the resulting objective function is intended to maximize the total number of flights that go through the FCA using their original schedules. Alternatively, an objective function of Max $\sum_f \left(q_f x_f\right)$ aims to maximize the planned amount of flight time left undisturbed. In comparison, a function of Min $\sum_f \left[(1 - x_f) \times \delta_f\right]$ can minimize the total rerouting delay for the entire congested airspace, where δ_f is the extra distance for flight f if it is rerouted and $(1 - x_f) = 1$ when flight f uses alternative schedule. Nonetheless, there are two major practical issues when implementing the objective function (3.1). First, the rerouting delay can be difficult to estimate *a priori*, because air carriers may not provide multiple route options when filing the flight plans. Additionally, this efficiency-oriented objective function does not take equity issues into consideration, which might lead to significant offloading imbalance among different air carriers.

The equity-oriented objectives shown in functions (3.2), (3.3) and (3.4) are intended to distribute the rescheduling and rerouting workload among air carriers as evenly as possible. Specifically, Objective function (3.2), Min $\sum_i \left|r_i - \overline{r}\right|$, aims to minimize the absolute deviation of offloading percentage from the average value $\overline{r}$

among the air carriers. Objective function (3.3), on the other hand, seeks to minimize the squared deviation of the offloaded ratio among all the carriers. Comparing functions (3.2) and (3.3), the latter places greater penalties on large deviations from the mean value $\bar{r}$. Focusing on the worst-case scenario for individual carriers, Objective function (3.4) aims to minimize the maximum deviation across different air carriers.

Based on the above discussions, the following mathematical program can be constructed so as to (1) balance achieving equity among air carriers and (2) reducing the number of rerouted flights. Essentially, the final goals are to efficiently utilize the FCA and to distribute offloaded flights fairly among airlines. One natural way of dealing with this problem is to adopt the following two-objective optimization formulation.

Model 1: Multi-objective Integer Programming Model

$$z_1 = \text{Max} \sum_f x_f \tag{3.5}$$

$$z_2 = \text{Min Max}_i\, r \tag{3.6}$$

Subject to:

$$\sum_{f \in U(j,t)} x_f \le c_j \qquad \forall j, t \tag{3.7}$$

In this formulation, the defined FCA involves F flights and J sectors. At most, $J \times T$ constraints are required to keep the sector capacity/demand balance, where T is the number of time periods in the study horizon. There are two different types of objective functions, the 1st maximizing the number of non-offloaded flights and the 2nd minimizing the maximum positive deviation from the average offloaded ratio for all air carriers. To investigate the tradeoff between these two criteria, one can formulate a single objective function as the weighted summation of the two objectives to generate a set of Pareto optimal solutions. An alternative approach incorporates the equity measure into the constraint set and then uses the ε-constraint method, which generates multiple solutions by varying the value of the parameter ε:

Model 2: ε-constraint model:

$$z_1 = \text{Max} \sum_f x_f \tag{3.8}$$

Subject to:

$$\sum_{f \in U(j,t)} x_f \le c_j \qquad \forall j,t$$

$$r_i \le \bar{r} \times (1+\varepsilon) \qquad \forall i \tag{3.9}$$

According to the definitions of r_i and $\bar{r}$, inequality (3.9) can be rewritten as the following function in terms of decision variable x_f.

$$\sum_{f \in A_i} (1-x_f) \le \sum_{f \in A} (1-x_f) \times \frac{|A_i|}{F} \times (1+\varepsilon) \qquad \forall i \tag{3.10}$$

$1+\varepsilon$ is a coefficient to control the percentage of allowable offloading deviation from the proposed flight routing ratio $\bar{r}$. Another way of controlling the overall allowable deviation is to add an upper bound, and the constraint (3.10) can be expressed as

$$\sum_{f \in A_i} (1 - x_f) \le \sum_{f \in A} (1 - x_f) \times \frac{|A_i|}{F} + \varepsilon \qquad \forall i \tag{3.11}$$

where ε is the overall deviation upper bound.

Goodhart (2002) discussed a similar equity constraint with an upper bound for a related traffic flow management problem. Her formulation used the amount of weighted delays as a measure of deviation and did not consider the impact of the carrier sizes. A general discussion on equity-related reformulations can be also found in Young (1994). Computational comparisons between the above bi-objective model and the ε-constraint model with two types of the equity constraints will be discussed in Section 3.6.

3.2 Alternative Network with Side Constraints Models: Circulation Model

Section 3.1 discusses the details of the general integer programming models, which typically require computationally intensive branch-and-bound search techniques to implicitly enumerate binary variables of x_f, especially for a large-sized problem involving many flights. In practice, initial FCA advisories are usually declared about 2 to 5 hours before the events occur, and the rerouting decisions can be revised as the event situation changes. As a result, the flight offloading problem needs to be solved in a timely manner, and a computationally efficient solution algorithm is critically needed in real-world applications as it would allow the involved air carriers to rapidly respond and request alternative actions.

In order to achieve better computational performance and utilize the special structure of this problem, this section is focused on developing alternative formulations using a network flow optimization model with side constraints. In particular, two network flow formulations with side constraints are examined and possible variants of the problem are discussed accordingly.

3.2.1 Single-sector case

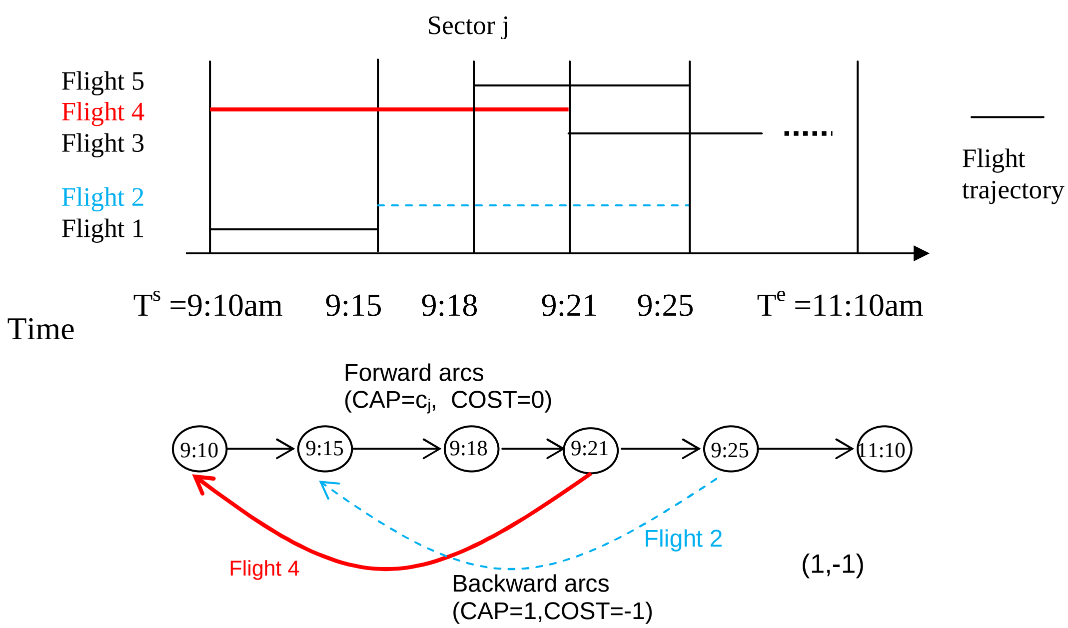


Figure 3-2 Network flow model for single-sector case

Let us consider a time-expanded network over node set $N = \{1,2,\ldots,T\}$ and arc set V. Each node represents a time instance when there are flights entering or leaving the area. The arc set consists of two classes of arcs: (1) a forward arc $(t, t+1)$ represents the time interval between time instance t and $t+1$; (2) a backward arc, (l_f, e_f), corresponding to a flight circulation arc that moves from time index of l_f to e_f, where e_f is the index of the flight's entry time and l_f is the index of the flight's exit time from the area.

For each arc (t, t'), let us denote $CAP_{t,t'}$ as the arc capacity and $COST_{t,t'}$ as the unit flow cost. In particular, on each forward arc ($t,t+1$), $COST_{t,\, t+1} = 0$ and $CAP_{t,t+1}$ = sector capacity c_j for sector j, which is the maximum simultaneous number of flights that a controller can handle during that time period. For simplicity, c_j is considered as a constant over time. For each backward arc, ($t=l_f$, $t'=e_f$), $CAP_{t,\, t'} = 1$ and $COST_{t,\, t'} =$ -1. As shown in Figure 3.2, the label on each arc ($CAP_{t,t+1}$, $COST_{t,t+1}$) represents (capacity, cost). In a solution to the min-cost flow problem, if there is one unit flow on the backward arc, there is a corresponding flow on the forward arcs, which forms a closed flow-conserving cycle. Note that, in practice and as assumed in this model, usually a flight can enter and leave a sector at most once. For the case where a flight enters and leaves the same sector more than once, additional side constraints are needed.

Through the above circulation network model reformulation in Figure 3.2, it is easy to show the corresponding node-arc incidence coefficient matrix is totally unimodular. A general discussion on Total Unimodularity (TU) for network matrices can be found in Wolsey (1998), and the corresponding linear programming relaxation (if feasible and finite) always has an integral optimal solution. Based on this unique structural property, the problem can be reformulated as a min-cost network flow model. This basic circulation network model is well-known; an early reference is Segal (1974), where it was widely applied to telephone operator scheduling.

3.2.2. Multiple sector case

In this section, we further extend the model in the single-sector case to the multi-sector case, where multiple layers of time-staged sector sub-networks are used to represent multiple sectors. As shown in Figure 3.3, each node in a layer for sector *j* corresponds to a time instance *t*, *t*=1, 2, …, *T*, the congested airspace time period covers from T^s to T^e. Accordingly, each sector includes a sequence of forward arcs that flow from node *t* to *t*+1 on sector *j*. Recall that, for the single-sector case, a backward arc can be used to represent each flight. In this complex multi-sector case, for flights that pass through more than one sector, the backward arcs need to cross at different layers. Thus, a new variable of $x_{f,(j,t),(j',t')}$ is introduced to represent a backward arc flow for each flight, from the ending sector *j'* at FCA leaving time $t = l_{f,j'}$, to starting sector *j* at FCA entering time $t = e_{f,j'}$. In addition, as illustrated in Figure 3.3, vertical transition arcs are used to represent the transition of a flight from one sector *j* to another sector *j'* at time *t*. We need the following additional notation for the multi-sector problem.

Notation:

o_f originally scheduled starting sector of flight *f* in FCA;

d_f originally scheduled ending sector index of flight *f* in FCA;

e_f originally scheduled entering time of flight *f* in FCA;

l_f originally scheduled leaving time of flight *f* in FCA;

$x_{f,(j,t),(j',t')}$: backward arc flow for each flight, from sector *j* at time *t* to sector *j'* at time *t'*

$w_{j,j',t}$: transition arc flow from sector j to j' at time t

$y_{j,t,t+1}$: forward arc flow from node t to t+1 on sector j

$\Phi(j, j', t)$: set of flights transfer from sector j to sector j' at time t.

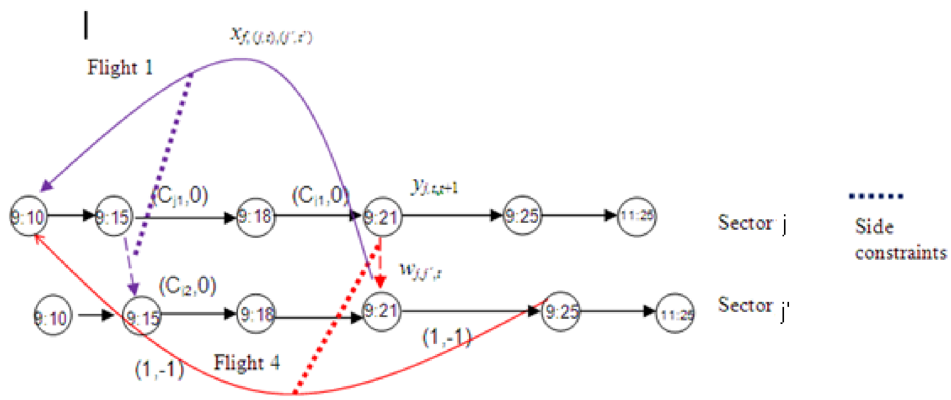


Figure 3-3 Network flow model for multiple-sector case

Side constraints are needed here to avoid the problem of non-unique tours since multiple vertical transition arcs might form multiple flow cycles for a particular flight. In Figure 3.3, both flight 4 and flight 1 enter sector *j'* from sector j but with different time stamps, 9:21 and 9:15, through the dashed arcs. If no side constraints are given, flight 4 can use the dashed arc at time 9:15 to enter sector *j'*, and flight 1 can also use the dashed arc to transfer sector *j* to *j'* at time 9:21 (without following the original schedule). To restrict those possible non-unique cycles in the network model, a side constraint is proposed as the following:

$w_{j,j',e_f} = x_{f,(j',l_f),(j,e_f)}$ where flight f transfer from sector j to sector j' at time t.

If more than one flight transfers from one sector to another sector at the same time, the flow of vertical arcs can be set equal to the total flow of backward arcs, which use the same vertical arc for transition.

$w_{j,j',e_f} = \sum_{f} x_{f,(j',l_f),(j,e_f)}$, where flight f transfer from origin sector j to destination sector j' at time e_f .

With the above formulation, the model can be extended from the single-sector case with extra side constraints, which is shown as follows.

Model 3: Network circulation formulation with side constraints

$$\text{Max} \sum_{f} x_{f,(j',l_f),(j,e_f)} \tag{3.12}$$

Subject to:

Flow balance constraint at each sector-time node (j,t),

$$\sum_{j',t'} x_{f,(j',t'),(j,t)} - \sum_{j',t'} x_{f,(j,t),(j',t')} + y_{j,t-1,t} - y_{j,t,t+1} + \sum_{j'} w_{j',j,t} - \sum_{j'} w_{j,j',t} = 0 \qquad \forall j,t \tag{3.13}$$

Sector capacity constraint:

$$y_{j,t,t+1} \le c_j \ \ \forall j,t \tag{3.14}$$

Side constraints for each flight transition arc (j,j',t),

$$w_{j,j',t} = \sum_{f \in \Phi(j,j',t)} x_{f,(j',l_f),(j,e_f=t)} \ , \ \forall (j,j',t) \tag{3.15}$$

Equity constants or fairness objective.

$$r_i = \frac{\sum_{f \in A_i} (1 - x_{f,(j',l_f),(j,e_f)})}{|A_i|} \quad \forall i \tag{3.16}$$

$$\sum_{f \in A_i} (1 - x_f) \le \sum_{f \in A} (1 - x_f) \times \frac{|A_i|}{F} \times (1 + \varepsilon) \qquad \forall i \tag{3.17}$$

Nonnegative and integer constraints for variables *x*, *y* and *w*.

In the above model, constraint set (3.13) represents the flow balance constraint for arc-based variables *x*, *y* and *w* imposed at each sector-time node (*j*,*t*), where the variable *x* corresponds to the flow on the backward arcs for each flight. Variable *w* corresponds to the flight transition between two sectors, and variable *y* carries the flow on sector j from time *t* to time *t*+1. For the constraint matrix *B* corresponding to this set of flow balance constraint, each column of *B* (that is, each variable, *x*, *y*, or *w*) only contains exactly two non-zero entries, +1 and -1, respectively, from the upstream sector-time node, or to the downstream sector-time node. The rest of the coefficients are zero for each variable. As a result, we can show that the incidence matrix A is totally unimodular.

Constraint set (3.14) imposes the sector capacity constraint on variable *y* at each time interval. Side constraints (3.15) are needed to ensure each flight transition arc *w* carries flow only if the corresponding flights are allowed to use the sector through backward arcs *x*.

3.3 "Flight on the Node" Model

The above network flow model uses a circulation network structure where nodes correspond to time instances and arcs correspond to flights. On the other hand, based on the special problem characteristics, an alternative formulation can be obtained using an "activity on node" network, where the arcs imply the precedence relations between any two flights' trip within a sector. A similar tanker scheduling problem was first developed by Dantzig and Fulkerson (1954). The problem of determining the minimum number of oil tankers required to meet a fixed transportation schedule was formulated as a linear programming problem and solved with the simplex algorithm. The same scheduling problem is discussed by Ahuja et al. (1993) with a different solution approach by constructing an equivalent network structure that can be solved by efficient maximum flow algorithms. In this study, a similar network structure is adapted, but the sector capacity is given and the objective is to keep as many flights as possible.

3.3.1 Single-sector case

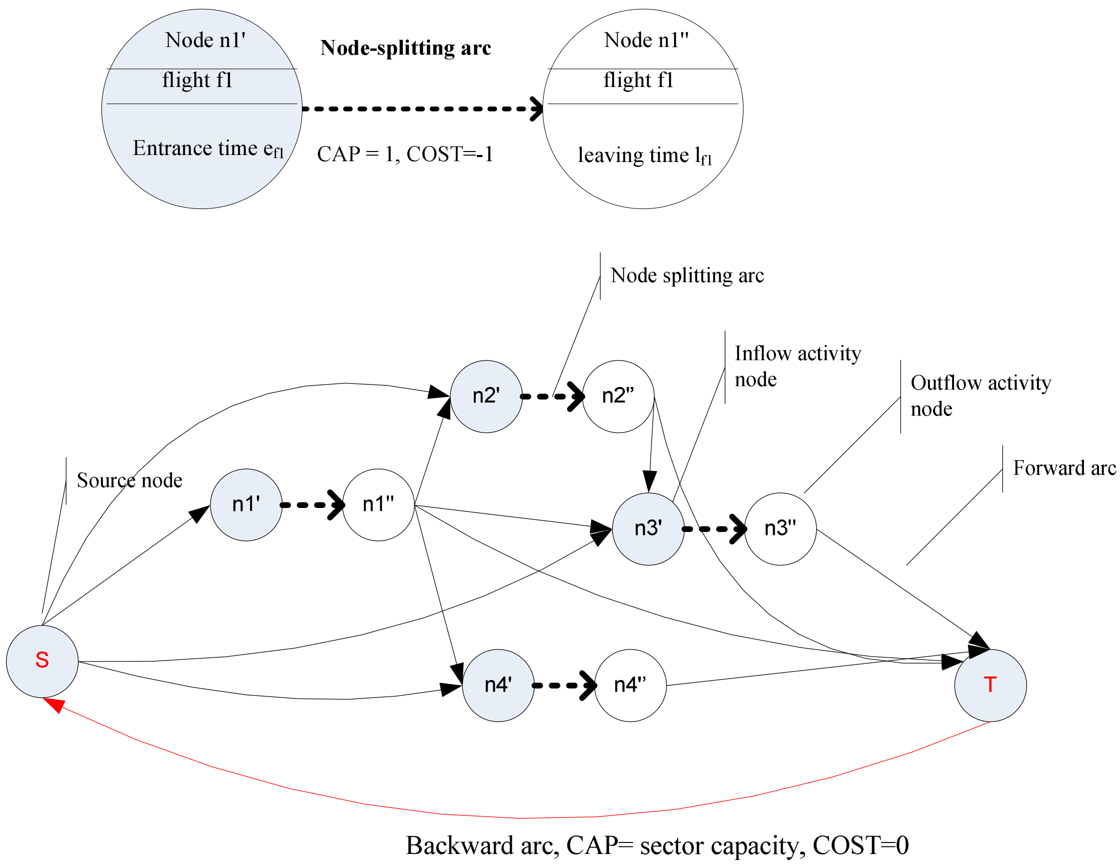
First, let us consider a network over node set $N = \{S, 1, 2,\ldots,n, T\}$ and arc set V. Each node, $n \in N\backslash\{S, T\}$, represents a flight's trip activity in the sector. Nodes S and T represent the source node and sink node, respectively, in the network. For each arc (n_1, n_2), $CAP(n_1, n_2)$ represents the capacity of arc(n_1, n_2), $COST(n_1, n_2)$ is the unit flow cost of (n_1, n_2) .

Each activity node is associated with flight index $f(n)$, and its entering time and exit time $e_{f(n)}$ and $l_{f(n)}$. In order to restrict flow through a node, this study uses a node-splitting technique, which replaces a node n with two nodes n' and n''. Each inflow node n' accepts all the inflow to the standard activity node n, and outflow node n'' handles all the outflow from the standard activity node n. A single node-splitting arc connects split nodes n' and n''. A capacity of 1 is imposed on the node-splitting arc in order to limit the flow through activity node n.

There are two additional sets of arcs. For $n_1 \in N\backslash\{S\}$, $n_2 \in N\backslash\{T\}$, forward arc, $(n_1, n_2) \in V$ if and only if the sector exit time of n_1 is less than or equal to the sector entry time of n_2, that is, $l_{f(n1'')} \leqslant e_{f(n2')}$. This means, these two activities n_1 and n_2, can be scheduled sequentially if the sector capacity allows. The backward arc (T, S) has flow capacity that corresponds to the sector capacity. The cost -1 is assigned to this arc to ensure that the objective value changes by -1 if a trip is chosen to remain.

In summary, for each node-splitting arc of node between node (n_1', n_1''), $CAP()$= 1, corresponding to a single flight trip, and $COST()$ = -1. For each forward arc (n_1'', n_2'), $CAP()$ = 1 corresponding to a feasible flight-to-flight connection, $COST()$ = 0 . For backward arc (T,S), $CAP()$ = sector capacity in terms of the maximum number of flights can be handled simultaneously in the sector, and $COST()$ = 0. If the sector capacity varies by time, the study time period can be divided into several intervals.

As shown in the example in Figure 3.4, there are 4 flights in a sector, and each flight has a standard activity node. Each standard activity node is split into an inflow and an outflow activity node, say $n1'$ and $n1''$. In this example, it is feasible to schedule flight 2, flight 3 or flight 4 after flight 1, as the leaving time of flight 1 (9:10) is earlier than the other flights' entering times (9:12, 9:15 or 9:12,). The occupancy time duration of flight 2 is very short (9:12-9:15), so it is also possible to schedule flight 2 before flight 3 (entering the sector at 9:15). For the node-splitting arc from node n1' to node n1" (corresponding to the activity of flight 1), we need to place a capacity constraint of 1 so that the arc is used only once, although this node $n1$ has three outgoing forward arcs to nodes $n2$, $n3$ and $n4$.



Flight number	Entering time e_f	Leaving time l_f
1	9:00	9:10
2	9:12	9:15
3	9:15	9:20
4	9:12	9:17

Figure 3-4 Network model for single-sector case

The solution algorithm needs to find the minimum-cost flow path through this network. From the optimum path passes through the nodes, we can identify the flight trips associated with these nodes can remain in the original route with original schedule. Note that, usually the flight could enter and leave the sector (on the scheduled route/sector list) once. If the flight needs to enter and leave the same sector more than once, a side constraint is needed.

3.3.2 Multiple-sector case

In the multi-sector case, multiple layers of sub-networks are needed to construct for the "flight on the node" model, while each sub-network corresponds to one sector. It is worth noting that, there are no arcs across the different sector networks, and additional side constraints are introduced to keep the consistency of the solution for the same flight due to the existence of the multiple trips in different sector networks. One way of adding the side constraints to the corresponding network structure is to set costs on the split arcs to be -1/*m*, where *m* is the number of sectors one flight traverses. As a result, when the flight is chosen in the solution, the total cost from all the flow in arcs corresponding to this flight becomes $-1/m \times m = -1$. The final solution of this model yields a number of parallel flow strings, while the flow through each string indicates which flights can stay with their original route and schedule.

The corresponding min-cost flow model is depicted below.

Additional Notation:

$x_{n1', n1''}$: flow on the node-splitting arcs;

$w_{T,S}$: backward arc flow from node T to S

$y_{n1'',n2}$: forward arc flow from node n_1'' to n_2'

A_{NS}: set of node-splitting arcs

N_I: set of inflow activity nodes

N_O: set of outflow activity nodes

$N_u(n)$: set of upstream nodes to node n

$N_d(n)$: set of downstream nodes to node n

Model 4: Flight on the node model without arc elimination

$$\text{Min} \sum_{(n1',n1'')\in A_{NS}} \left(c_{n1',n1''} x_{n1',n1''} \right) \tag{3.18}$$

Subject to:

Flow conservation constraint at each node:

For each inflow activity node:

$$\sum_{n1'\in N_u(n2')} y_{n1'',n2'} = x_{n2',n2''=N_d(n2')} \qquad \forall n2' \in N_I \tag{3.19}$$

For each outflow activity node

$$x_{n1'=N_u(n1''),n1''} = \sum_{n2'\in N_d(n1'')} y_{n1'',n2'} \qquad \forall n1'' \in N_O \tag{3.20}$$

Side constraints for all nodes n_k's that correspond to the same flight f

$$x_{n_1'(f),n_1(f)''} = \cdots = x_{n_k(f)',n_k(f)''} \qquad \forall f \tag{3.21}$$

Equity constraint:

$$r_i = \frac{\sum_{f\in A_i} (1 - x_{n'(f),n''(f)})}{|A_i|} \quad \forall i \tag{3.22}$$

$$\sum_{f \in A_i} (1-x_f) \leq \left[\sum_{f \in A} (1-x_f) \right] \times \frac{|A_i|}{F} \times (1+\varepsilon) \qquad \forall i \tag{3.23}$$

Nonnegative and integer constraints for variables *x* and *y*.

3.4 Analysis of Alternative Formulations

The preceding sections present four formulations, and we now want to prove that the feasible region of the integer programming Model 1 is in fact a projection to the other models. We will further evaluate the actual computational time of the different models in the previous sections.

Lemma 1: The feasible region of Model 1(integer programming model) P_1 is a projection of the feasible region P_3 of Model 3 (circulation network with side constraints model).

Proof:

We want to show that

i) when solution variables *x* is feasible to P_1, there exists *w* and *z* which make (*x*,*y*,*w*) feasible to P_3 ,

ii) ii) when solution (*x*,*y*,*w*) is feasible to P_3, *x* is feasible to P_1.

The outline of proof is given as follows.

i) Consider $x_f \in P_1$, which represents if flight *f* passes through the FCA using its original schedule. Clearly, x_f corresponds to the flow $x_{f,(j,t),(j',t')}$ on the backward arcs for flight *f*, from sector *j* at time *t* to sector *j'* at time *t'*, in the circulation network in Model 3. Through the flow conservation constraints (3.13) at sector-time node (*j*,*t*), one can express variables on the forward arc $y_{j,t,t+1}$ (from node *t* to *t*+1 on sector *j*) in terms of a summation of incoming flow *x* to time *t* , that is, $y_{j,t,t+1} = \sum_{j',t'} x_{f,(j',t'),(j,t)}$, as there is flow on the corresponding transition arc and $w_{j,j',t}$ =0. Similarly, side constraints for each flight transition arc (*j*,*j'*,*t*) at (3.15) allows us to construct $w_{j,j',t} = \sum_{f \in \Phi(j,j',t)} x_{f,(j',l_f),(j,e_f)}$. If the capacity constraint (3.7) holds, that is, $\sum_{f \in U(j,t)} x_f \le c_j$, in Model 1, we can also establish constraint (3.14) $y_{j,t,t+1} \le c_j$ in Model 3. Therefore, if solution *x* is feasible to P_1, then we can always construct another set of feasible solution (x,y,w) to P_3.

ii) Let v=(x,y,w) $\in P_3$, because v satisfies every constraints in Model 3. We can first express x_f in Model 1 in terms of the value $x_{f,(j,t),(j',t')}$ on the corresponding arc in the circulation network. If the capacity constraint in Model 3 holds, $y_{j,t,t+1} \le c_j$. Because the flow on each forward arc $y_{j,t,t+1}$ is the total number of flights simultaneously occupying the specific sector during time period (*t*,*t*+1), then the corresponding x_f satisfies the constraints $\sum_{f \in U(j,t)} x_f \le c_j$ in Model 1. □

<u>Lemma 2</u>: The feasible region of Model 1(integer programming model) P_1 is a projection of the feasible region P_4 of Model 4 (flight on the node with side constraints model).

Proof:

Similar to the proof of Lemma 1, it is easy to substitute the side constraints to the flow conservation constraints and easily express *y*, *w* variables in terms of *x*. As a result, it can be shown that i) when *x* is feasible to P_1, there exist *w* and *z* which make (*x, y, w*) feasible to P_4 ii) when (*x, y, w*) is feasible to P_4, *x* is feasible to P_1. □

3.5 Computational Results

This section applies the proposed models in two scenarios generated using real-world data. This section is organized as follows: the four models are first evaluated, and Model 1 is used to conduct the multiple-objective case study. The bi-criteria formulations in Model 2 are then tested with respect to different weights. Specifically, two different approaches of handling the multiple objectives, linear weighting and ε-constraint method are discussed in detail. Finally, we evaluate different models and approaches for solving the multiple-objective formulation.

3.5.1 Alternative formulation comparison

The following computational experiments are conducted based on datasets obtained from the Enhanced Traffic Management System (ETMS) database. The datasets are chosen from good weather days. Severe weather scenarios are created so that the demand could exceed the reduced capacity due to the weather. Specifically, the dataset consists of 33 and 19 sectors, 283 and 859 flights respectively.

Table 3-1 Problem size comparison of alternative formulations

	Model 1: Multi-objective Integer programming model	**Model 3: Circulation model**	**Model 4: Flight on the node model**
# of variables	859	2289 (corresponding to arcs)	>5000 (corresponding to arcs)
# of constraints	1573	2314 (flow balance constraints + sector capacity constraints+ equity side constraints)	2578 (flow balance constraints + sector capacity constraints+ equity side constraints)

The experiment uses 4 hours as an FCA time period. All experiments are performed on a Pentium IV 1.6GHz PC with 482 MB RAM. The program is coded in C with Callable Library, and CPLEXMIP 8.1 is used as the solver. Taking a case of 859 flights, Table 3.1 compares the size of the problem in alternative formulations. The basic integer programming model has the least number of variables and constraints. In the “flight on the node” model, for each pair of flight trip, which have a strict sequential order of using the airspace, there is an arc between the corresponding nodes. Thus, Model 4 has the largest number of arcs and variables.

In terms of computational efficiency, Model 1 (IP) and Model 3 (circulation) have similar computational running time, and in particular the LP relaxation problem for Model 1 with the single-efficiency-oriented objective problem is solved within a few seconds. In our experiments, only the 5 closest arcs are kept for Model 4. The computational study indicates that, with a limited number of sequencing arcs, Model 4 has similar performance compared to the other two alternative network models.

3.5.2 Computational comparisons between different multi-objective approaches

Multi-objective programming generally involves conflicting objectives, which cannot simultaneously arrive at the corresponding optimal levels. If there is an assumed utility function that could combine different objectives, one can accordingly choose appropriate (compromising) solutions by constructing a single objective maximizing solution. However, that is not the case here as it is generally difficult to predetermine the weights on the two types of metrics. As a result, the following section employs methods for generating representative Pareto optimal solutions.

By systematically changing the weights for different objective functions, one can obtain a set of solutions with different tradeoffs among the objective functions. As a result, considerable running time is required in order to obtain a subset of the frontier of Pareto optimal solutions. To obtain the tradeoffs of the multiple criteria and fully utilize the simple network structure of the problem under consideration, this study uses the following modified approximate ε-constraint method. Specifically, in the equity constraint (3.11), the average offloading ratio $\bar{r}$ is replaced by Z_1/P, where Z_1 is a constant number which can be estimated by solving the single objective model involving the system efficiency objective function only. Secondly, the equity constraints are added in the model and the average offloading ratio is replaced by Z_1/P from the first iteration. As a result, the equity constraints are simplified from general coefficient constraints to generalized upper bound constraints, and varying the value of ε yields a set of non-inferior solutions. The new equity constraints become:

$$\sum_{f \in A_i} (1 - x_f) \le Z_1 \times \frac{|A_i|}{P} \times (1 + \varepsilon) \qquad \forall i \tag{3.24}$$

where Z_1 is the objective value obtained from the first iteration.

Table 3.2 summarizes running time statistics for different models, where the results shown are obtained based on a dataset with 859 flights. The approximate ε-constraint method solves the problem using the least time compared to the other two methods, requiring around 14 seconds for 2 iterations that include the computation time of solving the Z_1 model and the second approximate ε-constraint model. The linear weighting method and the ε-constraint method have similar levels of performance efficiency. Both methods solve the problem relatively fast when the weights or the bounds are set to close to the extreme limit.

To further investigate how the proposed methods handle multiple objectives, the following analysis compares the tradeoff curves and the shading areas formed by the non-inferior points. Figures 3.5, 3.6 and 3.7 show the tradeoff curves from different methods.

Overall, the linear weighting method obtained 4 solution points, the ε-constraint method and approximate method obtained 6 points each. Theoretically, the approximate ε-constraint method can obtain the solution point which has first objective function value of 745. It should be noted that, the ε-coefficient needs to be chosen carefully. If it is set too big or too small, the method might not obtain a different solution or can make the problem infeasible. In this study, a standard step size rule is used for both ε-constraint methods. Comparing the solution points from the first two methods and combining all the points into a new figure, Figure 3.7

shows that some of the solution points obtained from ε-constraint methods coincide with or are dominated by the linear weighting method.

Table 3-2 Running time comparison among linear weighting and ε-constraint methods

	Obj. 1 Value (number of offloaded flights)	Obj. 2 Value	Running time (sec)
Linear weighting	765 (94)	0.2239	12.4
	759 (100)	0.0503	3.0
	756 (103)	0.0134	>20,000
	745 (114)	0.00003	718.6
ε-constraint	765 (94)	0.5	1.5
	765 (94)	0.4	1.8
	765 (94)	0.25	1.3
	761 (98)	0.2	1.1
	758 (101)	0.0491	>20,000
	757 (102)	0.0192	>20,000
	748 (111)	0.005	>20,000
	745 (114)	0.001	>20,000
Approximate e-constraint	765 (94)	0.2239	14
	761 (98)	0.1359	14
	759 (100)	0.0503	13.9
	757 (102)	0.0228	14
	756 (103)	0.0128	14
	745 (114)	0.001	14

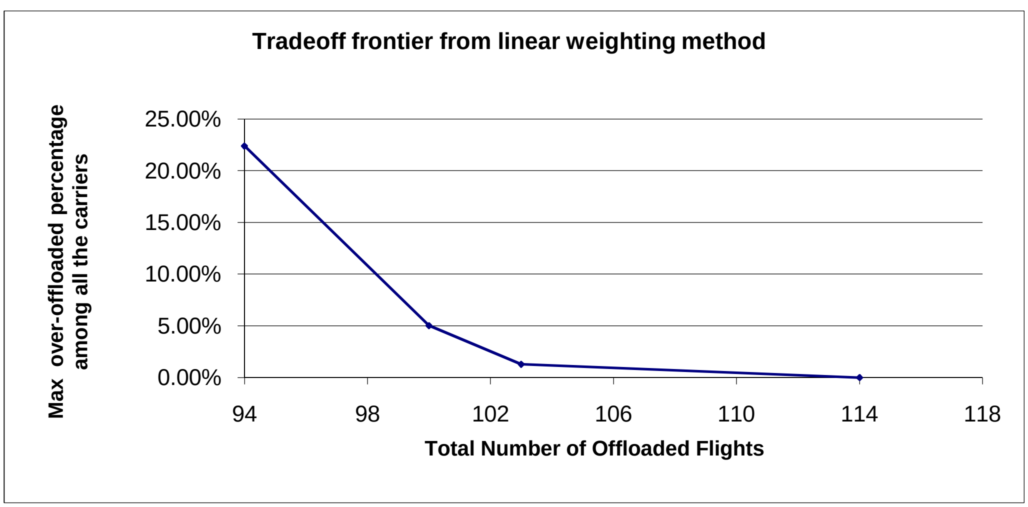


Figure 3-5 Tradeoff curve from linear weighting method

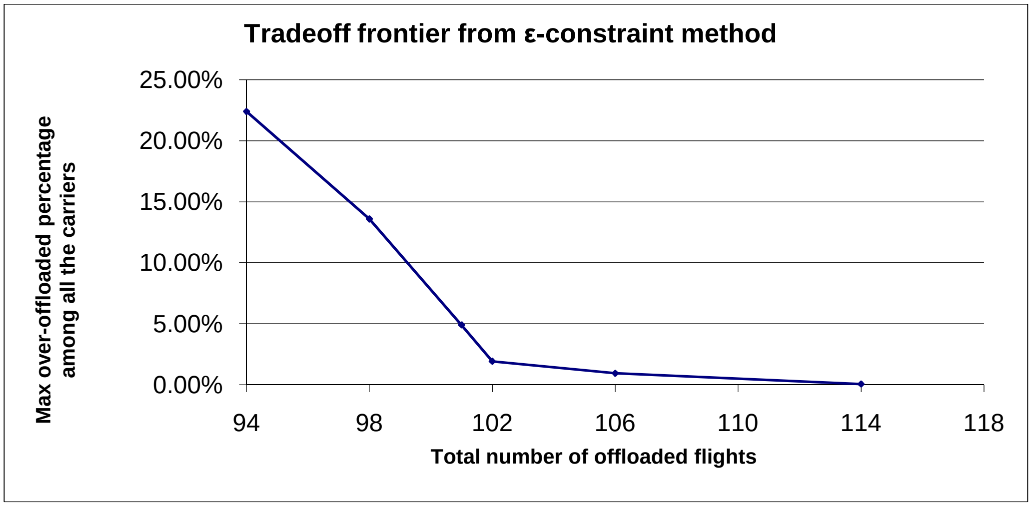


Figure 3-6 Tradeoff curve from ε-constraint method

Warburton (1987) proposed an ε-approximate algorithm to quantify the degree of accuracy in approximating trade-off curves and surfaces in a multiple criteria space. To compare the three methods' performance in terms of covering the non-inferior solutions, we use the following multi-objective solution quality measure in terms of the possible area of non-dominance solution space formed by the existing non-dominated solutions.

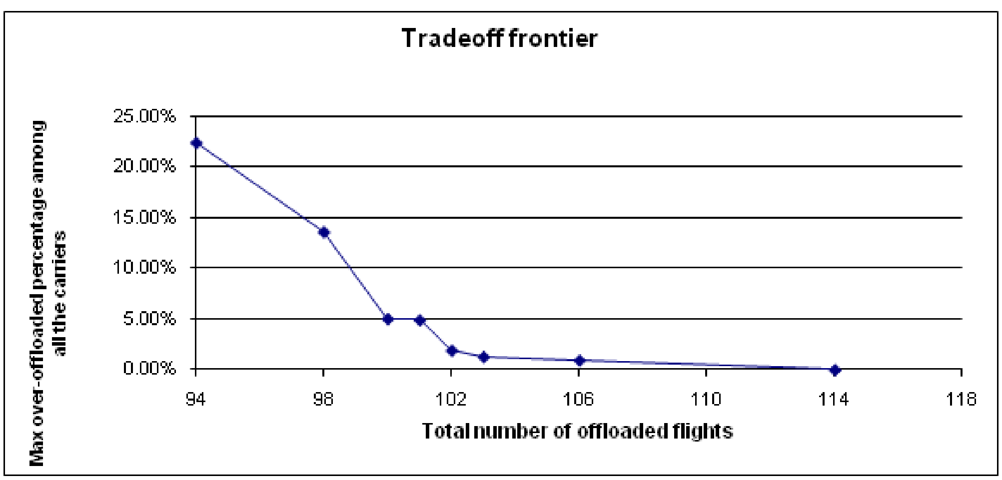


Figure 3-7 Combined tradeoff curve from all three methods

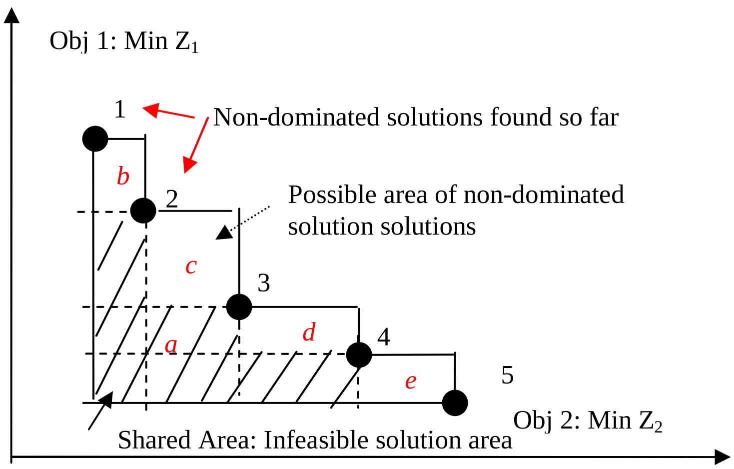


Figure 3-8 Example tradeoff curve showing non-dominated domain and approximation error

As illustrated in Figure 3.8, there are currently 5 existing non-dominated solutions on the tradeoff curve for two minimization objective functions. The area formed by the solid lines represents the domain that is not dominated by the existing non-inferior points. Within this area, no feasible solutions can be found inside the shaded area. We can simply verify the above statement by contradiction. If there exists a solution (say, *a*, with better objective function values Z_1 and Z_2) in the shaded area, then exiting solution 3 will be dominated, which contradicts with the fact that those five points correspond to non-dominated solutions. On the other hand, it is still possible to have solutions *b*, *c*, *d* and *e*, as none of them are dominated by existing solutions 1 to 5. As a result, the blank blocks sounding points *b*, *c*, *d* and *e* can be viewed as the region with possible non-dominated solutions. In this study, the total area of blank blocks can be viewed as an approximation measure of solution errors. To quantify the multi-objective solution quality, we want a smaller blank area or, equivalently, a larger shaded area. In the following discussion, let us compare the blank areas formed by points obtained from different solution approaches.

First, both efficiency and equity objective functions are converted to minimization functions. Particularly, the blank areas of possible non-dominated solutions obtained from the linear weighting, exact ε-constraint and approximate ε-constraint methods are 1.299, 0.822, and 0.718, respectively. Not surprisingly, the linear weighting method obtains the least number of non-dominated solution points, corresponding to a larger approximation error. For the other two methods, although the exact ε-constraint method obtained 6 points, the area containing possible non-dominated solutions is still bigger than that of the approximate method. Overall, this limited experiment shows that the approximate ε-constraint method produced a good quality tradeoff curve with the least computational effort. Moreover, if all the solutions points are combined together, the area of blank blocks, i.e. solution approximation error, is further reduced to 0.642.

Figure 3.9 further shows air carriers' absolute offloaded percentage difference compared with the average ratio across all the airlines. Four series of solutions, generated from the linear weighting method, are selected to illustrate the changes made by different weights of the coefficients. The horizontal axis sorts the air carriers by the total number of involved flights in this FCA advisory. The vertical axis shows the absolute offloaded percentage difference for each carrier. For this example, the total number of involved flights is 859.

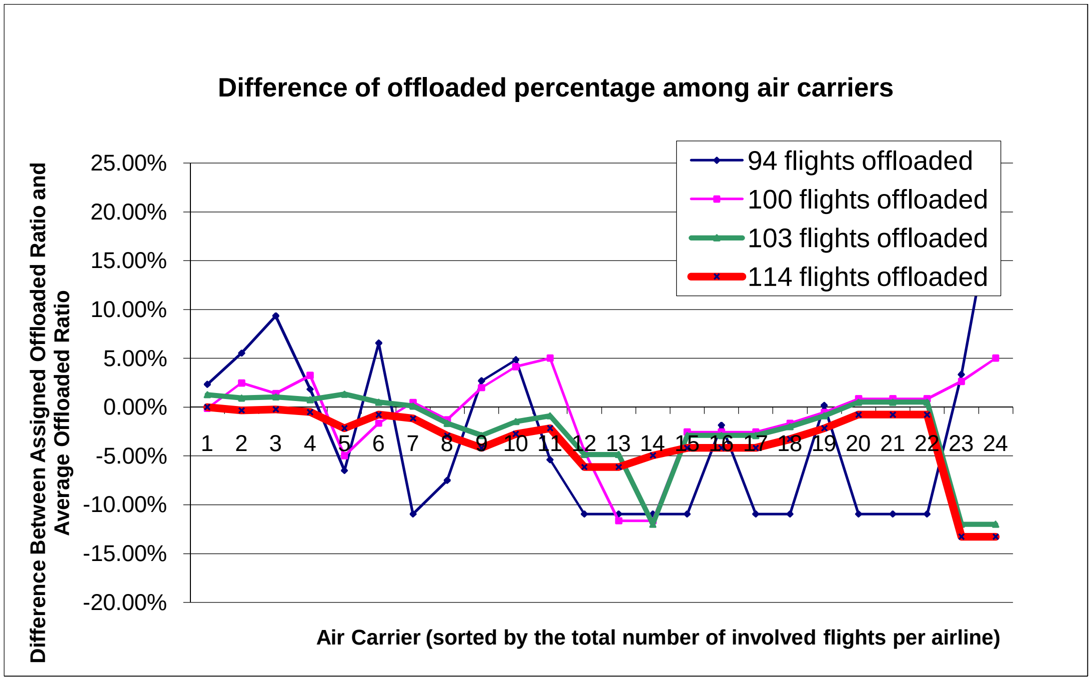


Figure 3-9 Difference of offloaded percentage among air carriers

In the different scenarios shown in Figure 3.9, the average offloading percentage is about 11%-13%. Overall, an increase of the weight on the equity objective function reduces individual deviations from the average offloading percentage. The total number of fights offloaded increases from 94 to 114 when the weight is set to be a very large value. As a result, the positive deviation is approaching to 0 for all air carriers.

It should be remarked that, some individual flights or airlines with very few flights are not considered well in this percentage equity function in terms of $r_i = \frac{\sum_{f \in A_i} (1 - x_f)}{|A_i|}$. It is possible that the offloaded flights are mostly chosen from these airlines with fewer flights when the penalty coefficients for equity increase dramatically. On the other hand, ignoring airlines with few flights does not violate the overall fairness standard, although the change of a single flight for a single-involved-flight airline will only have two possible values for the offloading percentage calculation: 0% or 100%. In order to minimize deviations for all airlines, including small airlines, the flights (belonging to the small airlines) have to be left in the FCA, which could degrade the overall system efficiency due to the limitation of this particular percentage equity functional form.

Chapter 4 Equitable Stochastic Airspace Routing Models and Algorithms

This chapter will develop models and algorithms to support efficient and equitable resource allocation of Airspace Flow Programs (AFP). When an AFP is issued, a 2-dimensional or 3-dimensional volume of airspace is specified with a time interval, corresponding to a period of reduced capacity. Under the AFP procedure, similar to holding an aircraft at the departure airport in a GDP program, the air traffic control center can adjust and optimize flight arrival times to the congested area, e.g., through a rationing algorithm that aims to offload the excess demand. In order to support real-time operation adjustment, the proposed model adopts a dynamic and stochastic optimization approach, where the offloaded flights are chosen and notified a few hours before the events occur, based on predicted adverse weather conditions. To balance the equity considerations for offloaded flights across different airlines, we incorporate an additional criterion to assign the airspace resources to the airlines in a fair manner.

This chapter first formulates a multi-commodity network flow with side constraints model for the AFP planning problem. The proposed model will consider the following two important modeling requirements: 1) time-dependent and stochastic airspace capacity and 2) equity considerations for ground-holding and rerouting flights across airlines. To jointly evaluate two major airspace congestion mitigation options, namely ground holding or rerouting excess flights from the problem area, we construct our model based on a space-time network that includes a set of airspace waypoint entries and airports.

To enable the equitable ground holding and rerouting decisions, this research considers multiple airline companies that own different numbers of flights, and the objective is to minimize the total weighted flight delay while enforcing equity of allocation, operational consistency and safety requirements. As a result, a time-dependent multi-commodity network flow formulation is developed, with airspace capacity and rerouting equity for each airline as side constraints. Moreover, to consider stochastic airspace capacity under severe weather conditions, we use multiple scenarios to represent random realizations of predicted capacities, and further integrate non-anticipatory constraints to ensure the first-stage solutions across different scenarios have the same values. A Lagrangian relaxation based method is used to dualize these three sets of side constraints so that the original complex problem can be decomposed into a sequence of linear programming problems with total unimodularity properties. Under a special case of deterministic capacity conditions, the original problem can be further decoupled into a sequence of space-time shortest path problems with very efficient solution algorithms.

Recall that, Chapter 3 proposes a model to use the entire congested/affected airspace and offload flights equitably among air carriers when minimizing total delay cost. The offloaded flights could be assigned to some alternative routes. This chapter further develops the model from Chapter 3 to include ground holding options, as well as stochastic capacity.

4.1 Space-Time Network Flow Model

This section aims to formulate the network flow optimization problem with a time-space expanded network structure. The notation of parameters and variables are shown below.

Index:

f	flight index, $f \in F$, F is the set of flights
u	airline index, $u \in U$, U is the set of airlines
$F(u)$	Set of flights belonging to airline u
t	index of scheduling time interval, t= 1,…, T, T is the length of planning horizon
k	index of stochastic scenarios, k=1, …, K, K is the number of scenarios
i,j	node index $i, j \in V$ in airspace routing network
A	arc index $a \in A$ in space-time network

Input Parameters:

$o(f)$	origin node of flight f
$d(f)$	destination node of flight f
$EDT(f)$	Earliest departure time of flight f from its origin airport
$LDT(f)$	Latest departure time of flight f from its origin airport
$PDT(f)$	Planned departure time of flight f from its origin airport
$EAT(f)$	Earliest arrival time of flight f at its destination airport
$PAT(f)$	Planned arrival time of flight f at its destination airport
$LAT(f)$	Latest arrival time of flight f at its destination airport
α	Cost of holding one flight in the origin airport for one time interval
β	Cost associated with one time period of delay at the destination airport, compared to planned arrival time
$S_{i,j}^{f}$	sector travel time of flight f on link (i,j)
$CAP_{i,j}(k,t)$	capacity constraint on link (i,j) at time t under scenario k.
$\theta(u)$	Threshold for average routing and ground holding cost per flight for airline company u
$\bar{T}$	The ending time of first planning stage that requires the unique solution across different scenarios.

Variables:

$x_{i,j}^{f}(k,t,t')$	= 1 represents flight f uses link from node i *to* node j with departure time t and arrival time t' under scenario k, *=0 otherwise*

A set of flights $f \in F$ belongs to different air carriers and an air carrier $u \in U$ has a set of flights $F(u)$. Each flight f is assumed to have a planned departure time $PDT(f)$ at origin airport $o(f)$ and a planned arrival time $PAT(f)$ at destination airport $d(f)$. The flight f needs to leave from the origin airport between a feasible range of Earliest Departure Time $EDT(f)$ and Latest departure time $LDT(f)$, and arrives at the destination airport before latest arrival time of flight $LAT(f)$.

Consider an airspace sector network G=(N,L), where node set N includes a set of waypoints and airports $n \in N$, and the set of links L corresponding to airspace sectors, $l \in L$. Without loss of generality, for a network N under consideration, a flight might originate from a boundary waypoint, rather than its origin airport. In this case, we will model delaying options at the inbound waypoint at the boundary of the subarea to allow the ground holding decisions. That is, according to the delayed arrival time at the inbound waypoint, the actual departure time at the originating airport of a flight can be consequently adjusted to avoid adverse weather conditions. Each link l can be denoted as a directed link (i,j) with upstream node i and downstream node j. The deterministic travel time for flight f on link (i,j) is $S_{i,j}^{f}$.

We then construct a space-time network to further develop a dynamic network flow model formulation. Let $STG(\boldsymbol{V}, \boldsymbol{A})$ represent the space-time network, where $\boldsymbol{V}$ is the set of vertices and $\boldsymbol{A}$ is the set of arcs (including sector traveling arcs and airport waiting arcs). A node n is extended to a set of vertices (n,t) at each time interval t in the study horizon, t=1,2, …,T, where T is the length of the optimization horizon. In the proposed space-time network representation, there are two types of nodes: airspace waypoints and airports. We also consider three types of arcs as follows.

(1) Sector traveling arcs are extended from a link (i,j) and each arc traverses from vertex (i,t) to vertex (j,t+$S_{i,j}^{f}$).

(2) Ground holding (i.e. airport waiting) arcs from ($o(f)$,t) to ($o(f)$,t+1) at the origin airport/waypoint. The feasible time window at the node $o(f)$ covers from the earliest departure time $EDT(f)$ to the latest departure time $LDT(f)$. By introducing the ground holding arcs, we can construct a single source vertex at the origin airport $o(f)$ and at the time instance of $EDT(f)$.

(3) Dummy waiting arcs from ($d(f)$, t) to ($d(f)$,t+1) at the destination airport, from the earliest arrival time EAT(f), to the latest arrival time $LAT(f)$. By introducing the dummy waiting arcs, we can construct a single sink vertex at the destination airport $d(f)$ and at the time instance of $LAT(f)$.

This special single-origin, single destination network structure (for each flight) allows us to establish the totally-unimodular coefficient matrix for all the flow balance constraints around vertices V in the time-expanded network.

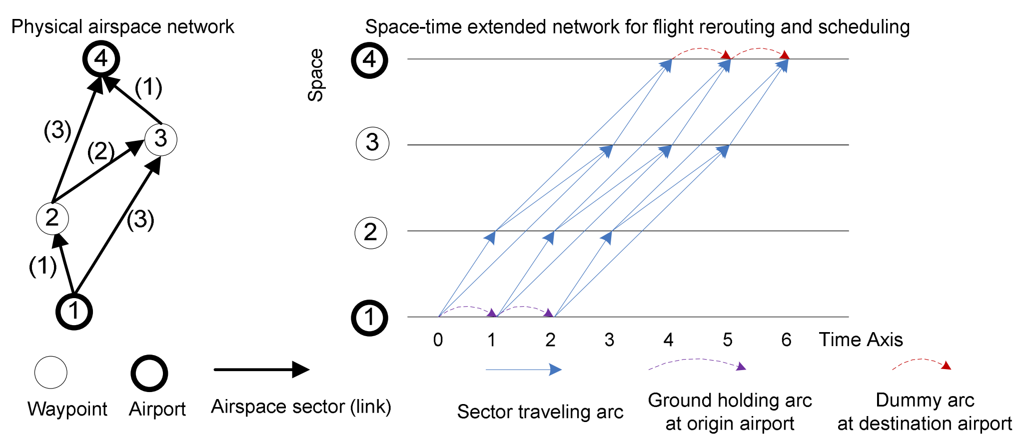


Figure 4-1 Physical airspace network and space-time extended network

The left portion of Figure 4.1 shows a simple airspace network with two airports 1 and 4, and two waypoints 2 and 3 that connect airspace sectors, and the labels on arcs correspond to link travel times for the associated flight. The right-hand-side of Figure 4.1 illustrates the space-time expanded network with possible fight trajectories starting from departure time t=0, 1, and 2 along ground holding arcs and sector traveling arcs.

Based on a weather prediction model or historical capacity reduction profiles, we can obtain predicted time-dependent capacity $CAP_{i,j}(k,t)$ under different scenarios k. Let us consider binary variable $x_{i,j}^{f}(k,t,t')$ that indicates the selection of link (i,j) in the space-time network. Within a two-stage stochastic optimization framework, the air traffic controller needs to make the re-routing and ground holding decisions for flight schedule variables $x_{i,j}^{f}(k,t,t')$ before time $\bar{T}$. Each airline u has a threshold $\theta(u)$ for the maximum average routing and ground holding cost per flight for a set of flights $f \in F(u)$. The subsequent multi-commodity network flow model is formulated to minimize the total expected weighted cost over the entire planning horizon, subject to the given sector capacity, airline total routing cost constraints, and non-anticipatory constraints. The stochastic integer programming formulation for the dynamic and equitable airspace routing and ground holding model can be described as follows.

Problem P4.1: Dynamic airspace routing and ground holding model

$$Z = min \sum_k \left\{ \sum_f \sum_{t>PDT(f)} \left[\alpha \times \left(x^f_{o(f),o(f)}(k,t,t+1) \right) \right] + \sum_f \sum_{t>PAT(f)} \sum_i \left[\beta \times \right. \right.$$

t×xi,d(f)fk,t−Si,d(f)f,t−PAT(f) (4.1)

Subject to

Flow balance constraints at origin airport vertex (at earliest departure time):

$$\sum_{j:ij\in A} x^f_{i,j}\left(k,t,t+S^f_{i,j}\right) + x^f_{i,i}(k,t,t+1) = 1, \qquad \forall k,f,i = o(f) \; and \; t = EDT(f) \tag{4.2}$$

Flow balance constraints at origin airport vertex after earliest departure time:

$$-x^f_{i,i}(k,t-1,t) + x^f_{i,i}(k,t,t+1) + \sum_{j:ij\in A} x^f_{i,j}\left(k,t,t+S^f_{i,j}\right) = 0, \qquad \forall k,f,i = o(f), EDT(f) < t < LDT(f) \tag{4.3}$$

Flow balance constraints at airspace waypoints:

$$\sum_{j:ij\in A} x^f_{i,j}\left(k,t,t-S^f_{i,j}\right) - \sum_{i:ij\in A} x^f_{i,j}\left(k,t,t+S^f_{i,j}\right) = 0, \qquad \forall k,f,t,i \in N - \{o(f),d(f)\} \tag{4.4}$$

Flow balance constraints at destination airport vertex

$$x^f_{j,j}(k,t-1,t) + \sum_{i:ij\in A} x^f_{i,j}\left(k,t,t-S^f_{i,j}\right) - x^f_{j,j}(k,t,t+1) = 0, \qquad \forall k,f,j = d(f), EAT(f) < t < LAT(f) \tag{4.5}$$

Flow balance constraints at destination airport vertex and at the last time stamp T

$$\sum_{i:ij\in A} x^f_{i,j}\left(k,t,t-S^f_{i,j}\right) + x^f_{j,j}(k,t-1,t) = 1, \qquad \forall k,f,j = d(f) \; and \; t = LAT(f) \tag{4.6}$$

Sector capacity constraints on link (i,j):

$$\sum_f \left\{ \sum_{i:ij\in A} x^f_{i,j}\left(k,t,t+S^f_{i,j}\right) \right\} \leq CAP_{i,j}(k,t) \quad \forall k,i,j,t = 1, 2, ..., T \tag{4.7}$$

Total routing and ground holding cost for airline company u:

$$\sum_k \left\{ \sum_{f \in F(u)} \sum_{t > PDT(f)} \left[\alpha \times \left(x^f_{o(f),o(f)}(k, t, t+1) \right) \right] + \sum_{f \in F(u)} \sum_{t > PAT(f)} \sum_i \left[\beta \times \right. \right.$$

$$t \times xi,d(f)fk,t-Si,d(f)f,t-PAT(f)-F(u)\times K\times\theta(u)\leq 0 \;\forall u \qquad (4.8)$$

Nonanticipativity constraints:

$$x^f_{i,j}(k, t, t') = x^f_{i,j}(1, t, t') \qquad \forall k > 1, f, i, j, t < \bar{T}, t' \qquad (4.9)$$

Binary constraints for $x^f_{i,j}(k, t, t')$={0,1}

The objective function in Eq. (4.1) aims to minimize a weighted combination of expected cost of ground-holding delay and arrival delay at the destination airport. The first term scans through the ground-holding arcs starting from *PDT*(*f*). $x^f_{o(f),o(f)}(k, t, t+1)$ =1 represents that the ground holding arc (*t*,*t*+1) is used at the origin airport *o*(*f*) under scenario *k*. α is the cost of holding one flight in the origin airport for one time period, thus the term of $\sum_{t > PDT(f)} \left[\alpha \times \left(x^f_{o(f),o(f)}(k, t, t+1) \right) \right]$ in Eq. (4.1) captures the total ground holding cost for a flight *f*.

The second term in Eq. (4.1) scans through all incoming nodes *i* to the destination airport *d*(*f*), and the link flow selection variable $x^f_{i,d(f)}\left(k, t - S^f_{i,d(f)}, t\right)$ is set to 1 if flight *f* arrives at the destination airport before time t. Therefore, $t \times x^f_{i,d(f)}\left(k, t - S^f_{i,d(f)}, t\right)$ represents the actual arrival time of flight *f* at the destination. Without loss of generality, the other cost factors for route adjustment can be also included in the objective function, such as fuel usage, en-route turbulence as well as safety considerations.

Eqs. (4.2)-(4.6) represent the flow balance constraints at the single source and single sink vertices in the space-time network, respectively. Eq. (4.3) and (4.5) maintain the flow conservation relations at each time instance t at the origin and destination airports, and those time-expanded nodes can be viewed as intermediate vertices in the space-time network. The flow balance constraints around the airspace waypoints at each time stamp are ensured by Eq. (4.4). Because the flow balance constraints (4.2-4.6) are established for each space-time vertex in the time-expanded network, it is clear that, a linear programming problem with this group of flow balance constraints is totally unimodular and its relaxation leads to integer solutions.

The time-dependent airspace sector capacity constraint is enforced in (4.7) under each scenario *k.* Air carrier-specific equity constraint (4.8) incorporates a ground delay and routing cost term per flight, which is similar to the cost in objective function (4.1).

Nonanticipativity constraints NAC (4.9) are used to construct deterministic equivalents to the scenario-based stochastic optimization models. In this two-stage problem, this set of NAC constraints implies that, the scenario-based variables $x_{i,j}^{f}(k,t,t')$ have the same values across different scenario k in the first stage $t < \bar{T}$.

In the case of a two-stage problem, this set of NAC constraints implies that, the scenario-based variables $x_{i,j}^{f}(k,t,t')$ have the same values in the first stage. Two modeling approaches have been developed in the literature to consider NAC: a splitting variable approach and a compact representation approach. The first method requires adding explicit coupling constraints, such as (4.9), across different scenarios in the first stage. The NAC constraints can be dualized into the objective function through a Lagrangian relaxation technique so that the scenario-based subproblems can be solved in parallel. The second method uses the single set of variables, say $\overline{x_{i,j}^{f}}(t,t')$, directly for all scenario-based subproblems in the first stage, which leads to a tight model with fewer variables and constraints.

Ball et al. (2003) developed a stochastic integer program with dual network structure and showed its application to the ground-holding problem. The dual network structure can be viewed as a special case of the compact representation approach for modeling NAC, while the resulting coefficient matrix in the dual network was shown to have a desirable total unimodularity feature that leads to efficient network flow algorithms. In our proposed space-time network-based formulation, if only the ground holding arcs are considered (i.e., without considering the airspace sectors), then the corresponding model can be further simplified to the dual network structure investigated by Ball et al. (2003).

Regarding scenario-based stochastic optimization methods, interested readers are referred to papers by Wets (1974), Birge (1995) and Shapiro et al. (2009) for more modeling details. Several techniques have been proposed to reformulate NAC constraints, including progressive hedging by Rockafellar and Wets (1991), augmented Lagrangian decomposition by Ruszczynski (1989) and the diagonal quadratic approximation algorithm of Mulvey and Ruszczynski (1992) to name a few.

4.2 Lagrangian Relaxation-based Solution Algorithm

With three sets of side constraints, the proposed multi-commodity network flow model is still very complex to solve by standard integer programming solvers, especially for real-world problems with a large number of flights and a long planning horizon. The proposed solution algorithmic framework addresses the following two questions: (1) how to find a valid Lagrangian relaxation procedure to provide tight lower bounds; (2) how to construct decomposed subproblems with efficient solution algorithm.

The constraints in the above airspace flight scheduling formulation can be classified as two groups. The first group directly relates to the flow balance constraints (4.2-4.6), which are all embedded in the space-time network characterized by a general inequality $MX \leq b$, and the network matrix M is totally unimodular. The second group includes link capacity constraints, air carrier equity constraints and the NAC constraints. Those three sets of coupling constraints cover either a set of different flights belonging to the same carrier u, a set of flights passing through the same arc from vertex (i,t) to vertex (j,t+$S_{i,j}^{f}$), or equations across different scenarios k. In this research, we plan to relax those complicating constraints, and accordingly decompose the large-scale airspace flight rerouting problem into multi-commodity network flow subproblems that are easier to solve. In general, network flow subproblems are desirable because they can be solved by many algorithms that are more computationally efficient than the standard simplex algorithm for linear programming problems.

By introducing a set of nonnegative Lagrangian multipliers π_u, $\rho_{i,j}(k,t)$ and $\lambda_{i,j}^{f}(k,t,t')$, we incorporate the coupling capacity, carrier equity and NAC constraints in the following objective function with penalty term.

Problem P4.2: Dualized dynamic airspace routing and ground holding model

$min \sum_{f,k} c^f(k) + \sum_{i,j,t,k}\{\rho_{i,j}(k,t) \times$

$fi{:}ij{\in}Axi,jfk,t,t{+}Si,jf{-}CAPi,jk,t{+}$

$\sum_u\{\pi_u \times [\sum_{k,f \in F(u)}\{c^f(k) - |F(u)| \times K \times \theta(u)\}]\} + \sum_{f,i,j,k,t,t'<\bar{T}}\{\lambda_{i,j}^f(k,t,t') \times$

$xi,jfk,t,t'{-}xi,jf1,t,t'$ (4.10)

where the generalized cost for each flight f under scenario k is denoted as

$c^f(k) = \sum_{t>PDT(f)}\left[\alpha \times \left(x_{o(f),o(f)}^f(k,t,t+1)\right)\right] + \sum_{t>PAT(f)}\sum_i\left[\beta \times\right.$

$t{\times}xi,d(f)fk,t{-}Si,d(f)f,t{-}PAT(f)$ (4.11)

Subject to the flow balance constraints (4.2-4.6) and nonnegativity constraints of $x_{i,j}^f(k,t,t') \geq 0$.

The positive multiplier vector ρ can be interpreted as the cost of $\rho_{i,j}(k,t)$ charged for utilizing a link resource (i,j) at arrival time t under scenario k with the sector capacity constraint $CAP_{i,j}(k,t)$. The multiplier π_u represents the penalty for exceeding average fight routing and ground holding cost threshold $\theta(u)$ for each individual airline, and $\lambda_{i,j}^f(k,t,t')$ corresponds to the penalty for not having the unique solution in the first stage. Essentially, the major goal of the Lagrangian function is to balance the total flight routing and ground holding cost, and the cost for utilizing limited facility resources through choosing appropriate resource prices. To obtain the largest possible bound values, we need to solve the following Lagrangian dual problem for variable $x_{i,j}^f()$, given multipliers $\rho_{i,j}(k,t)$, π_u and $\lambda_{i,j}^f(k,t,t')$.

Clearly, the dualized problem with only flow balance constraints is totally unimodular, so its linear relaxation produces integer solutions directly.

Since the dual cost function (4.10) is not differentiable everywhere, we solve the dual problem by updating $\{\pi, \rho, \lambda\}$ using the following subgradient method, which is intended to iteratively adjust the resource prices.

$$\pi_u^{q+1} = max\left\{0, \pi_u^q + \gamma^q \times \left[\sum_{k,f\in F(u)}\{c^f(k) - |F(u)| \times K \times \theta(u)\}\right]\right\}, \tag{4.12}$$

$$\rho_{i,j}^{q+1}(k,t) = max\left\{0, \rho_{l,t}^q + \gamma^q \times \left[\sum_f \sum_{i:ij\in A} x_{i,j}^f\left(k,t,t+S_{i,j}^f\right) - CAP_{i,j}(k,t)\right]\right\} \tag{4.13}$$

$$\lambda_{i,j}^{f,q+1} = \lambda_{i,j}^{f,q} + \gamma^q \times \left[x_{i,j}^f(k,t,t') - x_{i,j}^f(1,t,t')\right] \tag{4.14}$$

where superscript q is the iteration index used in the dual updating procedure, and π^q, ρ^q, λ^q , γ^q denote iteration-specific multiplier values, and step size at iteration q, respectively. To overcome "zip-zag" courses in the optimum search process, the step size parameter is updated as

$$\theta^q = \mu^q \frac{(\bar{L} - L^q)}{\left\|\Delta^q\right\|}, \tag{4.15}$$

where $\bar{L}$ is the objective function value of the optimal solution, which can be approximated by a feasible solution generated from the heuristic method, L^q is the value of Lagrangian relaxation at iteration q, Δ is the deviation vector associated with $\left[\sum_{k,f\in F(u)}\{c^f(k) - |F(u)| \times K \times \theta(u)\}\right]$, $\left[\sum_f \sum_{i:ij\in A} x_{i,j}^f\left(k,t,t+S_{i,j}^f\right) - CAP_{i,j}(k,t)\right]$ and $\left[x_{i,j}^f(k,t,t') - x_{i,j}^f(1,t,t')\right]$. Note that, $0<\mu^q<2$ is required to ensure theoretical convergence. Another modeling issue associated with Equation (4.15) is that there are a large number of constraints to be dualized, which leads to a large value of $\left\|\Delta^q\right\|$, a negligible step size and a potentially slow convergence rate. For simplicity, this research uses a step size updating rule of $\gamma^q = \dfrac{1}{q+1}$ in our numerical study. Recognizing that most of the capacity constraints are non-binding in the optimal flight re-routing solution, the relax-and-cut logic described in Caprara et al. (2002) is adapted here to dynamically relax resource capacity constraints by only dualizing a subset of constraints at every iteration. Specifically, if a sector has not reached its capacity within several recent iterations, the algorithm automatically resets the multiplier $\rho_{i,j}(k,t)$ for the under-utilized link capacity resource on link (*i*,*j*) at time *t* back to zero. With this dynamic constraint generation scheme, the set of Lagrangian multipliers $\rho_{i,j}(k,t)$ varies along the iterative process.

The uncapacitated multi-commodity flow optimization problem ***P4.2*** can be further separated into a set of subproblems, and each problem corresponds to a time-dependent flight-based network programming problem for flight *f* under scenario *k*, with the objective function associated with the weighted cost for $x_{i,j}^{f}\left(k,t,t+S_{i,j}^{f}\right)$. That is, given a set of resource prices $\rho_{i,j}(k,t)$ associated with arcs (*i*,*j*) from time *t* to $t+S_{i,j}^{f}$, we can now compute the optimal cost of a flight *f* for each possible entering/departure time at its inbound waypoint or origin airport, and possible routes in the airspace. The flight-based subproblems are then formulated as a sequence of time-dependent shortest path problems in the space-time network *STG*(***V***, ***A***), for given values of Lagrangian multipliers. For a comprehensive description of the shortest path algorithm in a space-time expanded network, we refer the readers to Ahuja et al. (1993).

4.3 Numerical Experiment

This section aims to test the computational efficiency and effectiveness of the proposed Lagrangian relaxation algorithm, as well as the impact of incorporating equity constraints under stochastic capacity conditions.

The first case study uses a network similar to the hypothetical subarea around Dallas-Fort Worth International Airport (DFW). This network has a total of 7 major airports. We consider 15 minute intervals and 20 time periods for a planning horizon of 5 hours. There are 6 major origin airports, DEN, LAS, RZC, LIT, EIC and AEX, and one destination DFW. We consider 144 flights belonging to 4 airlines, and those airlines own about 40%, 30%, 20% and 10% of total flights in the area. Those four air carriers operate on all origin-destination pairs. Two capacity scenarios are considered. The first scenario functions under a normal capacity of 12 aircraft per 15 minutes, which allows all flights to use primary routes. The second scenario has FCA in the shaded area in Figure 4.2, with a reduced capacity of 6 aircraft per 15 minutes, so that some flights need to take alternative routes. The cost of ground hold is set as 50% of arrival cost, that is, $\alpha=0.5\beta$ in Eq. (4.1). Related to the equity constraint (4.8), the air carrier-specific threshold of average routing and ground holding cost per flight $\theta(u)$ is set to 105% of the overall average value for all air flights.

The proposed two models are implemented in GAMS (Rosenthal, 2008), which is a high-level modeling system for mathematical programming and optimization. An open-source COINGLPK solver is used to solve the binary integer problem and the relaxed problem. In particular, the integer programming problem is solved through linear programming relaxation and branch and bound algorithm. Table

4.1 shows the size of the problem for different formulations. Each model has been solved to the optimal solution.

The problem instances in Mukherjee and Hansen's study (2009) have a relatively large number of flights and time intervals, and they use an AMPL/CPLEX solver to obtain integer solutions from LP relaxation of their proposed model. In their study, the reported computational time required by the LP relaxation for all cases was within 5 seconds, on a computer with 1.2 GHz processor speed and 16 GB RAM, while they also acknowledged that it was possible the LP relaxation might not yield integer solutions in some instances. In our study, it takes about 122 seconds to obtain the integer solutions through a complex branch-and-bound search process. The following discussion is not intended to compare the computational efficiency for different solution algorithms of the IP model (e.g., branch-and-bound vs. a simple round-off heuristics). Instead, we will focus on the problem size and solution quality associated with different model reformulations and relaxation techniques discussed in this dissertation.

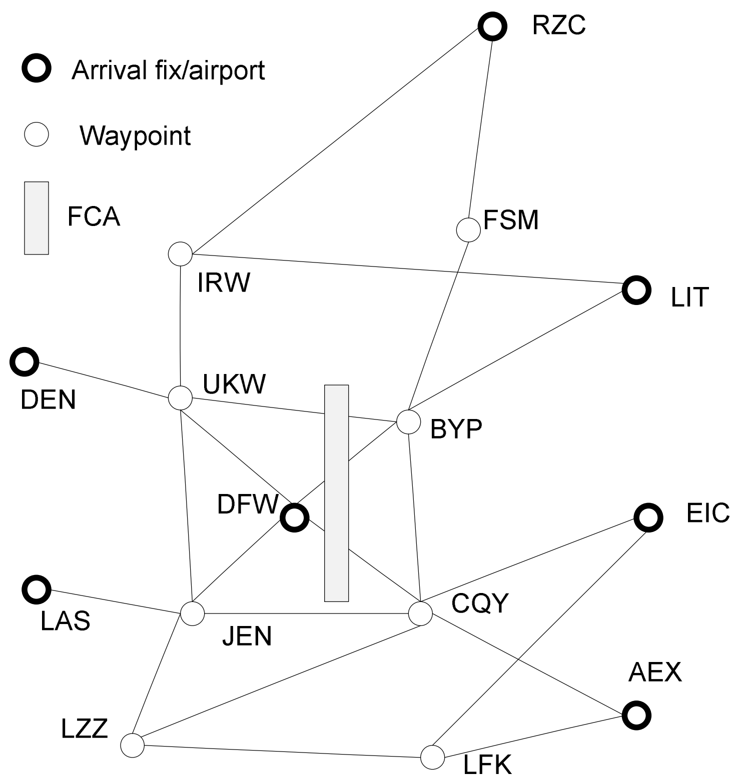


Figure 4-2 Hypothetical airspace network around DFW, extended from the test network from Mukherjee and Hansen (2009)

For different first-stage length, obviously, the relaxed formulation has the same number of variables as the original IP model. While the relaxed problems still keep about 80%-90% of the original (flow balance constraints), the overall solution times per iteration are reduced to about 2%-5% compared to the original IP model P4.1. For the relaxed problem P4.2, different lengths of the first stage (45 min vs. 15 min) have the same number of variables and constraints, as the NAC constraints (for the first stage variables) in both models are dualized. The length of the first stage decision does not significantly change the solution time, as both problem instances have similar computational times of 2-3 seconds.

Table 4-1 Problem size and solution time of reformulated problems under different first-stage lengths

	Complete Model 4-1 (first stage 45 min)	Relaxed subproblem model 4-2, (first stage=45 min)	Complete Model 4-1 (first stage 15 min)	Relaxed subproblem model 4-2 (first stage=15 min)
# of variables	165,601	165,601	165,601	165,601
# of equations	85,971	68,833	75,315	68,833
# of non-zero elements	859,393	496,225	3,451	496,225
Solution time (seconds)	122.507	2.844 (per iteration)	16.094	2.438 (per iteration)

Table 4-2 Size of side constraints

	Size of constraints	Value in the test problem with 45 min- first stage interval
stochastic sector capacity	# of scenarios K× # of links (i,j)× # of time intervals T	2×46×20 = 1840
airline total routing cost constraints	# of airlines $\lvert U \rvert$	4
non-anticipatory constraints	# of flights $\lvert F \rvert$× # of time intervals in the first time interval $\bar{T}$	144×3 = 432

Table 4-3 Computing time and solution quality for different types of relaxation (first stage 45 min)

Model	Computing time, unit: second (percentage of time compared to original IP formulation)	Solution quality in terms of percentage of $Z^{LB}(x)/Z(x^*)$
A. Original IP formation	122.507 (100%)	100%
B. IP with NAC + Equity constraint	6.234 (5.09%)	90.73%
C. IP with Capacity + Equity constraint	14.343 (11.71%)	93.43%
D. IP with Capacity+ NAC constraint	13.063 (10.66%)	95.05%
E. IP with all three-side constraints being relaxed	3.000 (2.45%)	92.68%
F. LP with all three-side constraints being relaxed	2.844 (2.32%)	90.73%

Table 4.2 shows the number of constraints for each set of side constraints. In general, the airline-specific routing cost equity requirement corresponds to the smallest number of constraints. On the other hand, the size of stochastic capacity constraints is relatively large, but it can be dramatically increased when more scenarios are needed to enable a realistic stochastic capacity representation. The size of nonanticipativity constraints is highly dependent on the number of total flights and the length of the first-stage decision time horizon.

Table 4.3 aims to systematically examine the computing time and solution quality of different relaxation models. The quality of lower bounds or relaxations, in this research, is measured by the percentage gap between a lower bound estimate $Z^{LB}(x)$ and the corresponding optimal value $Z(x^*)$ for the total system-wide cost, where x^* is the optimal solution. With the flow balance constraints, all the relaxations only need about 2% to 10% of the execution time for the original IP model (Model A), but provide approximation solutions within 10% of the optimality gap.

For three sets of side constraints, an interesting question is which set of constraints is the “hard” constraint to be dualized in order to enable effective model reformulation. By relaxing each group of side constraints individually, relaxed models B, C and D in Table 4.3 take less time to find their own optimal solutions. Overall, the capacity constraints are still dominating “hard” constraints, as Model B (with the capacity constraint being relaxed) uses the least time to solve. For nonanticipativity constraints and equity constraints (considered respectively in Models C and D), both models take about similar CPU time and generate similar solution quality gaps around 93%-95%. In particular, Model D with capacity and NAC constraints provides the tightest lower bound estimator (95.05%), while taking about 10% of the solution time compared to the original IP Model A. To investigate why the limited 4 equity constraints in Mode D still lead to a significant solution gap, we vary the values of $\theta(u)$, which is an air carrier-specific threshold of average routing and ground holding cost per flight. The current setting of 105% of the system-wide average routing and ground holding cost is indeed very difficult to satisfy, which requires seamless corrodination and reassignment among different airlines. When increasing $\theta(u)$ to 120% of the system-wide average, these 4 constraints become much easier to solve, and the relaxed Model D also reaches less than 2% of solution quality gaps.

The linear program relaxation in Model F shows a marginal advantage over integer programming Model E, as both models reach similar solution gaps with comparable computational time. This might be explained by the total unimodularity properties associated with the network flow balance constraint, which leads to many integral solutions from the linear program reformulation and limited steps for the branch and bound algorithm to solve the fractional solutions to meet the integer constraints.

Among all the possible relaxation models, we select Model D: IP with Capacity+ NAC constraint, and Model F: LP with all three-side constraints being relaxed, for further examination in an iterative Lagrangian relaxation solution process. Figure 4.3 shows the estimation quality of our proposed lower bounds compared to the optimal solution obtained by solving the IP model P4.1. As expected, the Lagrangian relaxation-based lower bound rule iteratively increases the estimation value, and, in general, marginal improvements become insignificant after 4 iterations for Model D as the relaxed subproblem, and 12 iterations for Model F as the relaxed subproblem. The maximum achievable Lagrangian lower bounds from these two reformulations are within 5% of the optimal objective function value. By considering the average computational time per iteration, using computationally efficient Model F as the relaxed subproblem is more beneficial overall, as it takes about only 65% of total CPU time to reach the same level of solution quality.

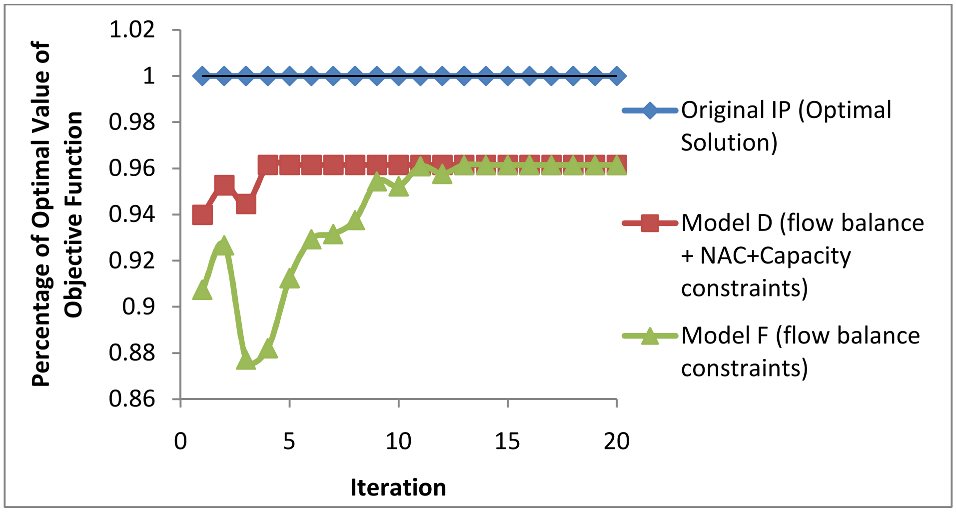


Figure 4-3 Estimation quality of proposed lower bounds compared to the optimal solution obtained by solving the model P4.1

4.4 Feasible Solution to Original IP

Another practical issue is how to generate feasible solutions based on the final solution from the relaxed model. For example, in Model F with all three side constraints being relaxed, there are still a few time slots where the number of flights exceeds the reduced capacity on FCA, the first-stage solutions at two different scenarios could have different values, and some airlines have routing and ground holding costs which are significantly larger than the system-wide average.

To quickly generate a solution that satisfies the relaxed capacity constraints, we start with the resulting Lagrangian multipliers $\rho_{i,j}^{q}(k,t)$ from the last iteration *q*, and further increase the penalty for using over-capacity sectors. To construct the unique solution for the first stage decision, we use the first-stage solution from the worst case scenario directly for each flight so that NAC constraints can be automatically satisfied. By fixing the first-stage solution to the worst case decision, we can solve optimally for each scenario and obtain the feasible solutions for each flight. It should be noticed, that worst case solution in the first stage might lead to too much freedom in terms of sector space satisfaction in most likely cases (or in non-worst case scenarios), and this heuristics might be ineffective, especially when the time horizon in the first stage is extremely long.

When Model B, C or D can be used for a given problem size, the issue becomes easier as the solution can be only infeasible to one set of constraints. In Model D, the resulting solution from relaxation model offers a feasible solution with respect to the NAC and capacity constraints. By increasing penalty on equity constrains, an upper bound of the optimal total system-wide cost can be obtained with a very small relative solution quality of 2.49%.

Overall, under tight equity constraints, it can be challenging to use a heuristic method to construct a feasible solution, as the tight equity constraints themselves might not ensure the existence of feasible solutions. We should also recognize that the equity constraints are soft constraints, and it is relatively easy to modify or relax the equity constraints to correct the infeasible solutions. In practice, one can first measure the degree of equity constraint violation in the current solution, and accordingly relax the equity constraints iteratively to obtain compromising solutions. In this chapter, we do not consider complicated heuristics to enforce the equity constraints, and various related heuristic rules will be discussed in Chapter 5.

In practice, it is not straightforward to consider the equity constraint for airlines with a single flight, especially when the proposed optimization model is applied for a single day or short optimization horizon. In this case, one can apply a randomization scheme or provide an exemption policy to reduce possible adjustment bias. On the other hand, an extended optimization model can be developed to take into account the rerouting and ground holding decisions from individual days, and evaluate the equity measure across multiple days in a cumulative fashion. By doing so, we can also smooth the penalty function on airlines with limited number of flights and avoid the adjustment bias.

4.5. Conclusions

This chapter models and solves an integrated flight re-routing and scheduling problem on an airspace network. Specifically, the optimization problem is concerned with a network of airspace sectors with a set of waypoint entries and a set of flights belonging to different airline companies. The goal of the optimization model is to minimize the total flight travel time subject to a set of flight routing equity, stochastic and assignment equity requirements. A time-dependent network flow programming formulation is proposed with airspace capacities and rerouting equity for each airline company as side constraints. A Lagrangian relaxation based method is used to dualize these side constraints and decompose the original complex problem into a sequence of flight rerouting/scheduling problems.

Chapter 5: Assignment Problem in Long-term Airport Slot Allocation

This chapter aims to study the long-term airport slot assignment problem, with a special focus on the equity issue across different air carriers. We first review the background information using an example at LaGuardia Airport, and then present a multi-objective integer programming model in sections 5.2 and 5.3 to optimize both system efficiency and air carrier equity. Alternative models and heuristic algorithms are developed in sections 5.4 and 5.5. The chapter is concluded with computational results.

5.1 Background Information on Slot Allocation

How to allocate airport landing slots to competing airlines has historically been a controversial issue at many airports. For example, at LaGuardia Airport (LGA), a High Density Rule (HDR) first went into effect in 1968, in which the incumbent operators are the “owners” of slots. On the other hand, a “use it or lose it” rule was established so that returned slots can be put into a pool for reallocation if the slots were not used 80% of the time. Since 1985, under the HDR, airlines have been able to trade slots in a secondary market, but such activity has declined over the years (Gleimer 1996). Between 2000 and 2010, the “use it or lose it” rule ceased to apply to the N.Y. area airport, and it was replaced with single caps on operations.

In 2000, the U.S. Congress enacted the Wendell H. Ford Aviation Investment and Reform Act of the 21st Century (AIR–21). Prior to AIR-21, LGA handled about 1,050 operations per day (spread over about 16 hours). Within seven months after

AIR-21, the number of scheduled operations had climbed to about 1,350 per day (i.e. a 28.5% increase). Because of the resulting levels of delays and cancellations, the FAA limited the number of slot exemptions using the "slot lottery" mechanism to bring the number of scheduled operations per hour down to 75. When planning a phase-out of the HDR, different authorities, including the FAA and the Office of the Secretary of Transportation (OST), recognized the possibility that there could be an increase in congestion and delay at the affected airports. Over the past several years, various market-based mechanisms have been proposed to allocate limited capacity at LaGuardia. Several new ideas had been suggested to alleviate the current problem in particular, expansion of airport infrastructure, confiscating a percentage of each airline's slots, mandatory use of larger aircraft, a 'congestion fee' for arrivals or departures during high traffic times, as well as slot auctions. However, LaGuardia cannot realistically expand its runway infrastructure because it borders on Bowery Bay and Flushing Bay.

In November 2004, the National Center of Excellence (NEXTOR) conducted a 2-day strategic simulation experiment to measure airline responses to a variety of congestion pricing fees and administrative rules. In February 2005, NEXTOR conducted a second strategic simulation to examine how an auction model could be used to allocate capacity. There are many issues to be addressed prior to implementing an auction of take-off or landing authorizations at LaGuardia. To name a few, the notion of incumbency; associated property rights and their duration, if any; the impact that auctions may have on airport revenues; predictability of the auction outcome; the impact on small communities; and the financial impact on the airlines and their customers. On the other hand, several advantages to the auction mechanism were also explicitly recognized in the discussion. For example, auctions rely on markets, which are more robust and responsive to industry changes than administrative regulations and seem to allocate scarce resources less arbitrarily than allocating slots under an administrative solution (such as a lottery).

Recently, under a proposed rulemaking, the FAA proposed to attach finite lifetimes to existing slots authorized. Additionally, expired landing slots would be subject to reassignment, using a marketing mechanism, such as auction. Interested readers are referred to a recent report prepared by Ball et al. (2005) on an overview of auction use and auction design, as well as various options for controlling congestion at LaGuardia Airport after December of 2006. The expiration and reallocation of slots should drive airlines to put slots they hold to the best possible use because the slots would no longer represent an indefinite investment interest. The revolving allocation process also would provide new entrant airlines and incumbent airlines wishing to expand service at a particular airport the opportunity to acquire landing slots.

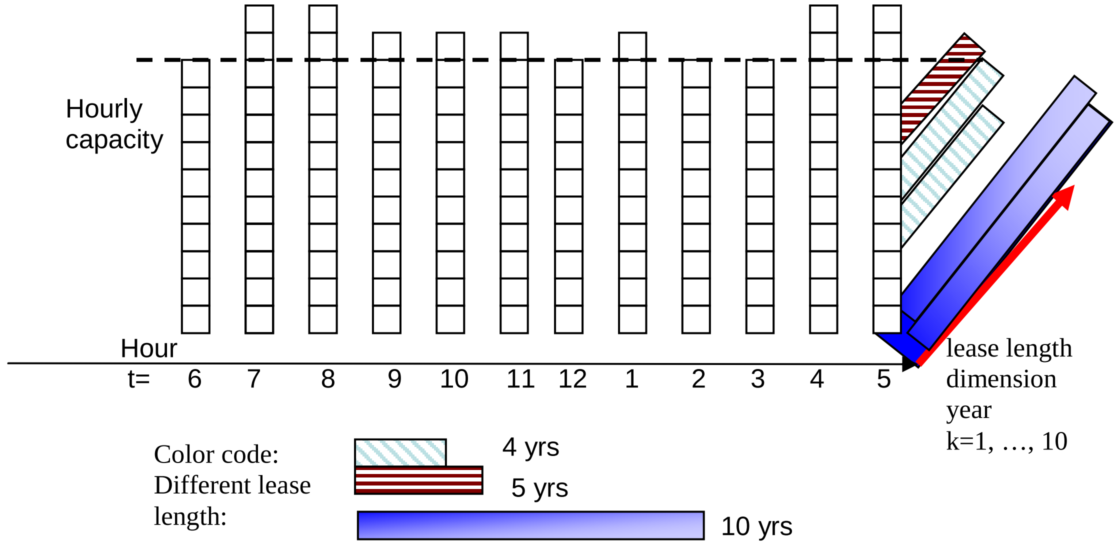
5.2 Problem Statement for Long-term slot allocation

We first start with the formal problem statement and key assumptions of the long-term landing slot assignment problem. In this special case of the resource reallocation problem, the FAA would limit the number of scheduled flight arrivals and departures at a major airport. For instance, Monday through Friday from 6:30 a.m. to 9:59 p.m. (peak hour) and Sunday from noon to 9:59 pm would have a ceiling on hourly operations. In general, slots are created according to the hourly limit on operations in terms of the number of scheduled arrivals and departures, so that these slots would be allocated to carriers at the airport based on historic usage. A few slots can also be reserved for general aviation.

Assumption 1: The assignment problem under consideration only addresses arrival slot allocation. ,If an airline obtains an arrival slot, say at 8:30am, that airline can schedule a paired departure without any restriction. However, this pairing assumption has certain limitations and can be studied in the future research. For example, flights are most likely to be scheduled compactly to maximize equipment and crew efficiency, and a departure slot from the flight originating airport is heavily dependent on its demand for an arrival slot at the flight’s destination airport.

Assumption 2: If the departure arrangement by an airline causes any potential capacity issue in a certain hour, a departure time window can be assigned to each landing slot to ensure the balance. The above example can be modified to allow the carrier to schedule a departure between 9:15am and 10:15am.

The problem could be illustrated in the following three-dimensional assignment plot in Figure 5.1. During each hour (along the x time axis), the current number of slots that are operating needs to be controlled under the capacity level (in the vertical z axis), and each slot should be assigned to an airline with a determined slot lease term (along the y axis, e.g., 1 yr-10 yrs, only counting from the fourth year when reallocation takes effect).



X axis: hour assignment: t=1, …,
Y axis: length of lease: years 1-10
Z axis: total number of slots

Figure 5-1 Illustration of long-term airport slot allocation problem as a three-dimensional assignment problem.

Overall, in the proposed assignment, each carrier's holdings of slots would satisfy two conditions/constraints: (1) the average "life" and value of the slots would be approximately the same for all airlines; and (2) expiration of slots would be staggered so that no airline would lose a disproportionate number of slots in a given time period.

It should be noted that, landing slots in different hours have different values. In general, landing slots in early morning and late afternoon, i.e., peak hours, have higher values than the slots in the middle of the day. In other words, the expiration dates of the regular authorizations in each hour would be assigned as follows.

(1) Capacity limitation: the number of slots is equal to the average number of "slot" operations held under the HDR or subject to a predetermined capacity in each hour time period;

(2) Equity among carriers: the average remaining life for all slots is roughly 5.5 years or a similar value if a time discount factor is applied;

(3) Minimal service interruption (evenly distributed reallocation annually): the total years of the remaining life among all slots would be distributed so that 10 percent of the total slots at the airport expire each year.

5.3 Integer Programming Formulation

In this section, we present several integer programming models that assign "life" or lease to landing slots at a major airport to achieve the system optimal capacity utilization, equity among air carriers and minimal service interruption as much as possible. Generally, multi-objective programming involves conflicting objectives, so it is possible that not all objectives can simultaneously reach their optimal levels. Recall that multiple solutions are generated in Chapter 3 so as to construct Pareto optimal tradeoff curves. Alternatively, within a single utility maximizing framework, this section focuses on how to combine different objective functions together, and then compare different resulting models and possible heuristic algorithms.

The notation and decision variables are defined as the following. The time of day is divided into a finite set of time periods of equal duration (for example, 1 hour, denoted by T). This formulation considers the slot lease assignment time window from 7AM to 10PM. The time interval can be one hour without loss of generality.

Subscripts:

i index of air carrier,

k slot lease length, k=1…10 years,

t time period of day, t = 1,2, ...,T (e.g. 7am-10pm during a day).

Parameters:

S_{it} number of slots carrier i owns in time period t,

H_t target average lease (slot years/slot) for time period t,

C_{kt} number of available slots with lease length k in time period t,

N_i number of slots carrier i owns in all time periods,

V_t value of a slot in time t,

TA_i target average slot value for carrier i,

TT_i target total value for carrier i,

α weight coefficient for max slot lease percentage deviation,

β weight coefficient for air carrier slot value deviation percentage.

Decision Variables:

x_{ikt} number of *k-year* slots assigned to airline i in time period t,

p_{it} average slot lease percentage deviation for airline i in time period t,

y_t max average slot lease percentage deviation in time period t,

γ maximum allowable deviations from target.

As described earlier in this section, there are several possible objectives in this problem. We start by definition for each airline i.

Average Slot Value (ASV_i) is $\frac{\sum_{t,k}\left(k \times V_t x_{ikt}\right)}{N_i}$, we can now define several equity metrics.

1. Overall average slot value deviation from target:

$$ADG_i = ASV_i - TA_i = \frac{\sum_{t,k}\left(k \times V_t x_{ikt}\right)}{N_i} - TA_i$$

2. Overall total slot value deviation from target:

$$TDG_i = \sum_{k,t}\left(k \times V_t x_{ikt}\right) - \frac{N_i \sum_{i',k,t}\left(k \times V_t x_{i'kt}\right)}{\sum_i N_i}$$

Air carrier performance equity metrics:

1. Equity metric weighting air carriers equally

$$EMA = \sum_i \max(0, ADG_i)$$

2. Equity metric weighting by air carrier size

$$EMB = \frac{\sum_i \left[N_i \times \max(0, ADG_i)\right]}{\sum_i N_i}$$

3. Equity metric weighting by *squared root* of air carrier size:

$$EMC = \frac{\sum_i \left[\sqrt{N_i} \times \max(0, ADG_i)\right]}{\sum_i \sqrt{N_i}}$$

To assign slot lease equitably in terms of both slot lifetime length and total value for each airline, one objective could be defined as the weighted sum of an hourly metric and EMA, as described in the following model.

Model: Slot Assignment Problem (SAP):

$$\min \quad \alpha\sum_t y_t + \beta\times EMA$$

$$= \alpha\sum_t y_t + \beta\times\sum_i \max(0, ASV_i - TA_i) \tag{5.1}$$

$$= \alpha\sum_t y_t + \beta\sum_i \max(0, \frac{\sum_{k,t}\left(k\times V_t x_{ikt}\right)}{N_i} - TA_i)$$

Subject to:

Capacity constraint:

$$\sum_i x_{ikt} \le C_{kt} \quad \forall k,t \tag{5.2}$$

Supply constraint:

$$\sum_k x_{ikt} \le S_{it} \quad \forall i,t \tag{5.3}$$

Min-max definitional constraint:

$$y_t \ge p_{it} \quad \forall i,t \tag{5.4}$$

$$p_{it} \ge 1\text{-}\frac{\sum_k\left(k\times x_{ikt}\right)}{S_{it}H_t} = \frac{S_{it}H_t - \sum_k\left(k\times x_{ikt}\right)}{S_{it}H_t} \quad \forall i,t \tag{5.5}$$

In objective function (5.1), the first term $\alpha\sum_t y_t$ is the summation of the maximum slot lease percentage deviation over different time periods. Essentially, this single-hour metric aims to ensure slot life equitably among air carriers in each time period. The second term $\beta\sum_i \max(0, \frac{\sum_{k,t}\left(k\times V_t x_{ikt}\right)}{N_i} - TA_i)$ is the summation of overall airline slot value deviation percentage (i.e., equity metric EMA). For a typical multi-objective optimization program, coefficients α and β are the weights that can be adjusted depending on the importance of the two metrics. Depending on the decision makers' specific consideration and the other available alternative metrics, the problem could focus on air carrier performance, e.g., air carrier slot lease percentage deviation and air carrier slot value percentage deviation, or on hourly metric. Without loss of generality, the following section will illustrate algorithms and model improvement based on Eq. (5.1).

There are several challenging issues in solving the above model:

1. Symmetry in time values: One symmetry problem comes from the coefficients if no time discount is applied for the lease type. For example, if airline *a* has 2 slots in hour 7, given the target average lease term = 5.5, leases of 5 and 6 years and leases of 4 and 7 years have equal values. To find the optimal solutions, a typical integer programming solver needs to maintain symmetric nodes in the branch-and-bound tree, leading to large solution times. To break the symmetry in the model, a time discount factor could be applied. This will be further discussed in the following section.
2. Complicated constraints Eqs. (5.4) and (5.5): With only constraints (5.2) and (5.3), the problem is a simple transportation problem, however, with constraints (5.4) and (5.5), the problem becomes more general (profoundly difficult IP).

The above issues make the problem difficult to solve optimally within reasonable computational time. In the following section, an alternative way of solving the problem to near optimality will be described.

5.4 Sequential Optimization Models

Considering the complexity of solving the above model with multiple time periods, this section proposes to decompose the problem into multiple single-hour optimization subproblems, as shown in Figure 5.2. Essentially, this sequential model iteratively applies the single period model to each time window and then adjusts targets after each iteration to achieve global balance (2nd objective function). The single-hour optimization problem considers an hour at a time, so the decision variables x_{ikt} are reduced from a three-dimensional vector (air carrier *i*, lease length *k* and time period *t*) to a two-dimensional vector (air carrier *i*, lease length *k*) for a specific time *t*. The assignment results from the previous time periods to the current time *t* is provided as a result of solving the same subproblems at the previous time periods, say, $\tau=7, 8, \ldots, t-1$. That is, for the single-hour subproblem at subject hour *t*, assignment results $\overline{x_{ik\tau}}$ are given for $\tau=1, 2, \ldots, t-1$, and we can use a new variable vector x_{ik} to represent the number of *k*-year slots assigned to airline *i* in time period *t*.

Single-Hour Slot Assignment Problem $SHSAP_t$:

$$\min \quad y_t \tag{5.6}$$

Subject to:

Capacity constraint:

$$\sum_i x_{ik} \le C_{kt} \quad \forall k \tag{5.7}$$

Supply constraint:

$$\sum_k x_{ik} \le S_{it} \quad \forall i \tag{5.8}$$

MinMax definitional constraint:

$$y_t \ge p_{it} \quad \forall i \tag{5.9}$$

$$S_{it}H_t \times p_{it} + \sum_k \left(k \times x_{ik}\right) \geq S_{it}H_t \quad \forall i \tag{5.10}$$

Constraints representing maximum allowable deviations from target:

$$\sum_i \left(\frac{\sum_k \left(k \times v_t x_{ikt}\right) + \sum_{\tau<t} \sum_k v_\tau x_{ik\tau}}{N_i} - TA_i \right) \leq \gamma \tag{5.11}$$

It is easy to observe that the new constraints (5.7) and (5.8) can be viewed as the supply and demand constraints in a standard assignment problem, and the min-max definitional constraint can be handled in the post-checking stage. The second and third terms in Eq. (5.11) are constants since the other sub-problem solutions are fixed except for the subject hour that is solved. In Eq. (5.11), γ is the threshold used in the airline metric, e.g., 5% meaning the maximum allowed air carrier deviation percentage. After determining this threshold and obtaining the list of airlines with the metric exceeding the threshold, Eq. (5.11) is dynamically added to the improvement problem. For air carriers with more deficits from their targets, the threshold could be slightly altered so that the air carrier could be compensated more in the improvement procedure. The scheme used here is similar to the ε-constraint method discussed in Chapter 3. The new single-hour sub-problem becomes much smaller in size, e.g, when the time period includes 16 hrs, the sub-problem is 1/16 size of the original problem, which dramatically reduces the computational time.

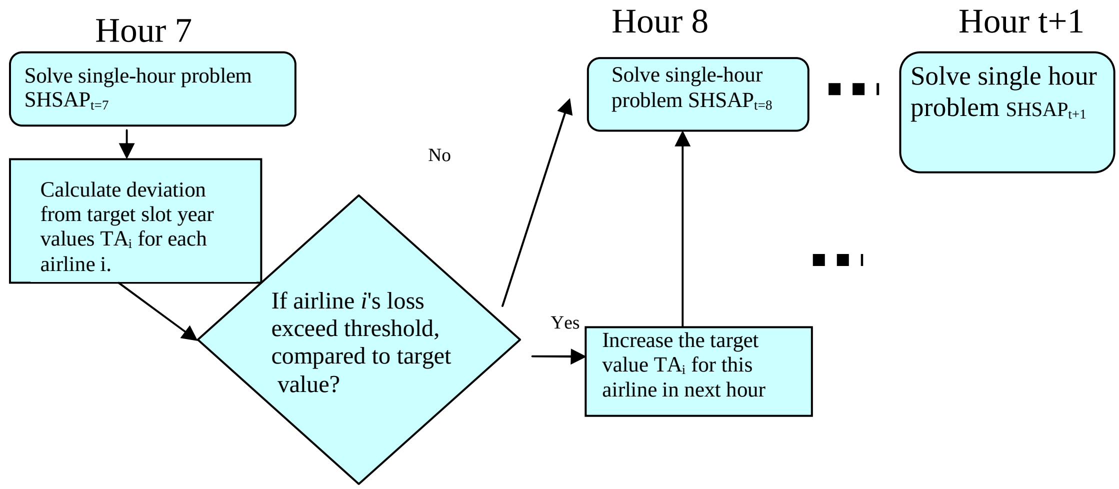


Figure 5-2 Sequential optimization models illustration

Essentially, there are two ways of implementing the Sequential optimization procedure.

1. Solve each hour chronologically using equity measure 1 as the objective and adjust air carriers' target after solving each single hour problem.

2. Obtain an initial feasible solution, and update chronologically by adjusting airlines' target hour by hour.

Accordingly, we propose the following two heuristic algorithms to the slot assignment problem.

Algorithm 1:

For iteration n=1 to M

Step 1: Solve a single-hour problem SHSAP$_t$ with the hourly metric objective (5.6) subject to the air carrier metric threshold (maximum allowable deviations from target, i.e. γ) constraints (5.11), and then calculate the deviation from the target slot year value and other equity metrics for each air carrier *i*.

Step 2: Compare each air carrier's loss or deviation with respect to the threshold (a constant set at the beginning, e.g., 5%). Adjust air carrier's target TA_i in the next hour if the threshold is reached.

Step 3: Repeat step 1 for the following hour until all hours have been calculated at least once and no air carrier's loss or deviation exceeds the threshold.

End for

Stop and output final solution.

Another way of utilizing the characteristics of the hourly problem can be described in the following algorithm.

<u>Algorithm 2:</u>

Step 1: Obtain an initial feasible solution by solving a single-objective (hourly metric) problem *SAP* without considering the air carrier performance objective function.

Step 2: Starting from the first hour *t*, calculate air carrier deviations ADG_i or TDG_{i} in the current solution.

Step 3: Choose the max air carrier deviation as $\max_{\mathrm{i}}\{ADG_i\}$, and compare it to the deviation threshold γ. If no air carrier metric exceeds the threshold, continue for the next hour *t*+1. Otherwise, hold the solutions in the other hours to be constant, add air carrier deviations exceeding threshold as additional constraints and solve the resulting single-hour problem $SHSAP_{\mathrm{t}}$ to obtain improved solutions.

5.5 Priority-Based Paired Allocation Heuristic Algorithms

To achieve reasonable fair slot assignment results, this section aims to explore alternative heuristic methods by enhancing the commonly used round-robin method in the field of computer resource scheduling. In general, the round-robin procedure alternately allows claimants to choose among resources left, and it is considered to embody the fundamental fairness principle. In our study, we view the round-robin method as a heuristic for solving Integer Programming, and further enhance it to allow interactive participation by air carrier representatives.

There are a wide range of scheduling algorithms available for allocating scarce resources, such as first-come first-served scheduling, shortest job first scheduling, priority scheduling and round-robin scheduling. In particular, the round-robin scheduling method has been widely used in time sharing CPU systems, in which a small unit of CPU time resource, called "*time quantum*", is defined. Each process/user only obtains a small slice of time quantum (typically 10-100 million seconds), and time slices are assigned to each process/user in equal portions and in circular order.

Let us consider the slot selection for a particular subject hour, where air carriers can be viewed as slot users and the assets to be assigned are slots with different lease lengths. We can assign each air carrier to one slot at a time from the pool of available slots. By doing so, all air carriers are handled without priority in this round-robin scheduling method, which can lead to the following key properties related to max-min fairness, as illustrated in Figure 5.3.

(1) Slot resources are allocated to air carriers in order of increasing demand, such that no users receive more than requested.

(2) Users with low demand will receive all of their requests, and users with high demands will not have all their demand satisfied but will evenly split the remaining slot resources.

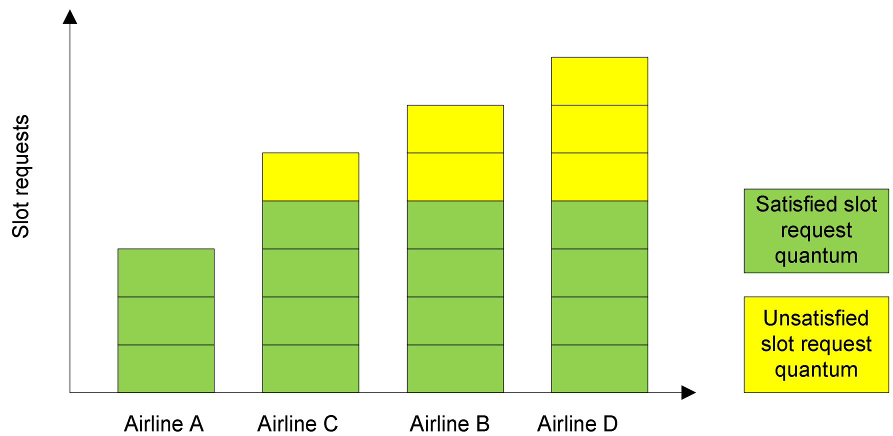


Figure 5-3 Illustration of round-robin scheduling method

If a simple round-robin mechanism is used, each air carrier has a chance to select one preferred slot in a cyclic order, but it may not be desirable if the value of slots (with different lease lengths) varies widely from one to another. The remaining challenge is how to create more balanced "*slot request quantum*" to ensure fair assignment across different air carriers, because the long-term slot assignment problem under consideration also involves an additional dimension of slot lease lengths (e.g. 1 year vs. 10 years). For example, a long-term lease (e.g., 10-year lease) would be favored over other short-term leases (1 or 2 year lease). In this case, the air carriers which can select slots early will always select favored slots with higher value first, and the leftover slots will have low values. To further ensure fairness, the proposed enhanced round-robin algorithm first creates a number of bundles (i.e. resource quantum), and each bundle includes one or two slots and the corresponding average lease length is closer to the average lease length for all of the slots available. By doing so, each bundle will have similar or equal value. In another words, each time an air carrier A chooses, if A is "owed" 2 or more slots then A chooses a pair of slots from among a specific list of "balanced" pairs, otherwise a single slot is chosen. For example, if an air carrier needs to request 7 slots, then it has 4 time quantum which can select 2, 2, 2 and 1 slot(s), respectively. If the available slot leases are 1-10 years, without time discount factor, the average target slot lease should be close to 5.5 years. Depending on the size of request in a quantum (one slot vs. two slots), a single-slot request quantum will be assigned a slot with a lease length of 5 years or 6 years, while a two-slot request quantum will be assigned to a two-slot combination of (1 year + 10 year) or (2 year + 9 year).

Algorithm 3: Enhanced round-robin assignment algorithm

Step 1: (Initialization) For subject hour t, for each air carrier i, calculate the target slot value TA_i and actual average slot values $AA_i = \frac{\sum_k (k \times v_t x_{ikt}) + \sum_{\tau<t} \sum_k v_t x_{ikt}}{N_i}$ for the beginning of assignment t=1 to the current hour t. Calculate the difference between TA_i and AA_i as max$\{0, AA_i - TA_i\}$, according to Eq. (5.11).

Step 2: (Creating time quantum for slot selection) Divide air carriers' requests into selection time quantum. At each time quantum, an air carrier can select at most 2 slots as a bundle.

Step 3: (pre-sorting time quantum) Use one of the following priority rules to sort the selection time quantum in the queues.

1. If an air carrier only has one slot request, the associated time quantum will be placed first.
2. Else if an air carrier has a large slot value deviation as abs(AA_i-TA_i), then this air carrier will be placed in the beginning of quantum queue.

Step 4: (assign slots to quantum) For each time quantum in the queue, select available slot(s) sequentially. If two slots are requested from a time quantum (as a bundle), then a combination of two slots with the average target lease length close to the overall average target lease length will be selected.

Step 5: (Update) Update the actual assigned slot values in each slot and total assigned slot values. If $t = T$, stop, otherwise advance time clock $t = t+1$ and go back to step 1.

To further improve the performance of heuristic methods, the following rules are proposed. The growing deviation metrics will be introduced first.

The growing deviation value (GDV) metric is defined as:

$$GDV_{i,t} = \frac{\sum_{\tau=1}^{t-1} \sum_{k} \left(k \times v_{\tau} \times x_{ik\tau} \right)}{N_i} - TA_i$$

The growing deviation slot year metric is defined as if 5.5 is the target slot lease

$$GDY_{i,t} = \frac{\sum_{\tau=1}^{t-1} \sum_{k} \left(k \times x_{ik\tau} \right)}{N_i} - 5.5$$

Heuristic Rule 1: After running each hour *t* sequentially, calculate target value per slot based on past assignment for each air carrier; use this target value per slot in the following hour, deviation metrics are based on the GDV, eg, if one air carrier got lower than target assignment for the already assigned hours, the new calculated target value per slot will be adjusted higher to make up for the deficit.

Heuristic Rule 2: Slightly different from Rule 1, adjust target value after each subgroup assignment instead of each hour.

Heuristic Rule 3: Slightly different from Rule 1, use GDY instead of GDV

Heuristic Rule 4: Slightly different from Rule 2, use GDY instead of GDV

5.6 Computational Comparisons among Different Models and Approaches

We first explain the experimental and computational settings, followed by comparisons of the performance of different models. Since daily slot numbers vary slightly from day to day, the test data set in this study is constructed from a single day (March 15^{th} 2005) of LGA's Aviation System Performance Metrics Official Airline Guide data, which contains departure/arrival slot information for each airline. As discussed earlier, departure slots and arrival slots are assumed to be properly paired, so the following experiments only focus on assignment of arrival slots, for simplicity. It is assumed that available slot leases are 1-10 years in length and every year the discount factor is 0.97, i.e. a 2-year slot lease will be worth $2^{0.97} = 1.9592$.

All the optimization experiments are conducted on CPLEX 9.1 from SUN workstations, and the paired-assignment heuristic algorithm (algorithm 3) is coded in

C++. CPLEX cannot solve the complete formulation of the Slot Assignment Problem (SAP) with an analysis horizon of 16 hours, 21 airlines and 35 – 40 slots per time window (hour), and an “out of memory” error message was encountered after several days of execution. The sequential optimization model, presented in section 5.4, obtains results within a reasonable amount of time, and the paired-assignment heuristic algorithm finishes in a few seconds.

Figure 5.4 shows the number of slots in each hour for the given day, when the reduced capacity is set to 36 per hour.

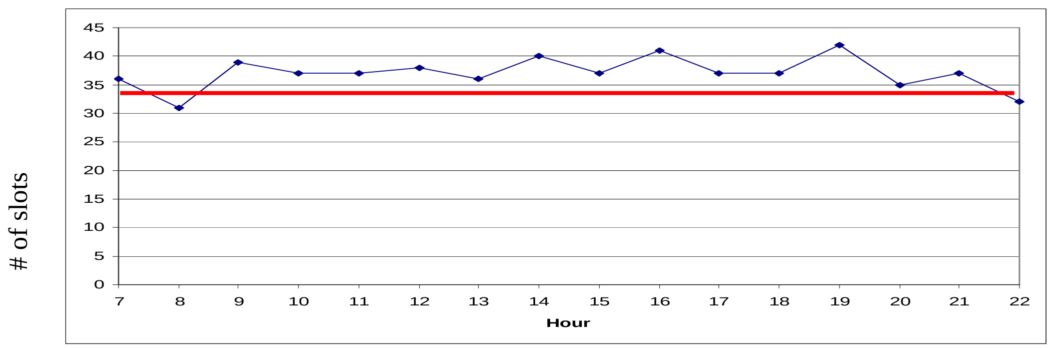


Figure 5-4 Slot schedule in each hour and reduced capacity of 36 flights per hour

<u>Results of single-objective (hourly metric) optimization model</u>

The single-objective optimization model aims to minimize the maximum air carrier slot year percentage deviation in each hour: min $\sum_t y_t$. Figure 5.5 shows the percentage of deviation for hour *t*: $y_t = \max_i \left[1 - \frac{\sum_k (k \times x_{ikt})}{S_{it} H_t} \right]$, corresponding to constraints (5.4) and (5.5), is less than 7% from hour 7 to hour 22. If all airlines’ performance is measured in terms of $ADG_i = ASV_i - TA_i = \frac{\sum_{t,k}\left(k \times V_t x_{ikt}\right)}{N_i} - TA_i$, that is, the ratio of each airline's actual slot year values ASV_i, and target slot year values TA_i, Figure 5.5 shows that air carrier slot year value deviation from the target could be as large as 15%.

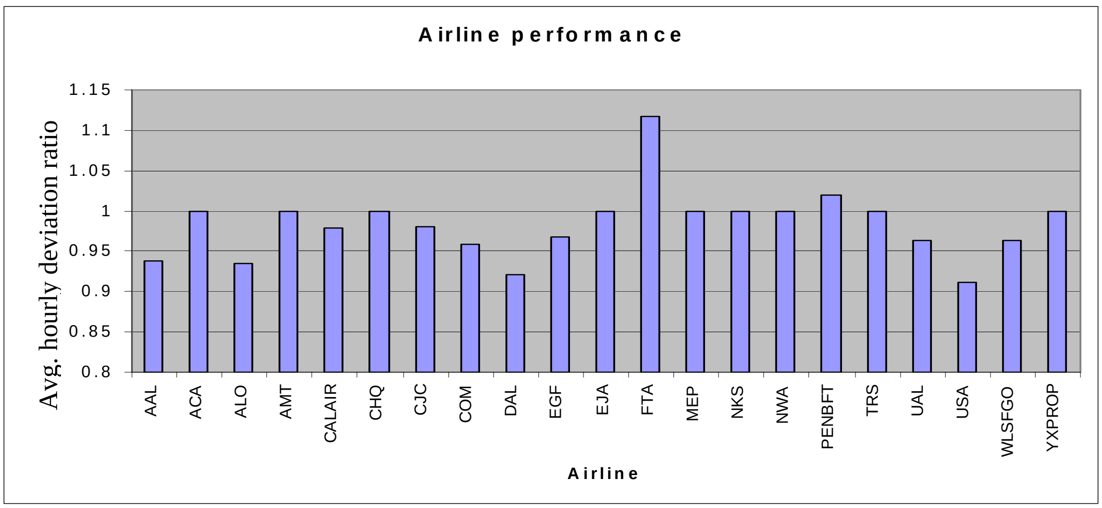


Figure 5-5 Air carrier specific fairness measure from single-objective optimization model

We then examine the results from the sequential solution procedure. Starting from the feasible solution obtained by solving the individual hourly problem to solve the min-max problem, the sequential optimization procedure aims to improve the initial feasible solution and adjust each airline's goal slightly. The results in figures 5.6 and 5.7 show the hourly metric change and air carrier performance change after a single round of improvement.

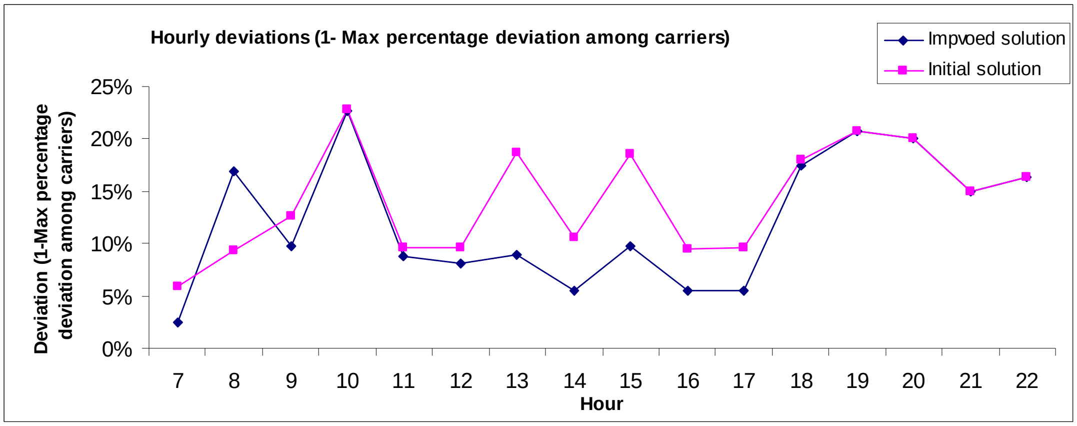


Figure 5-6 Hourly deviation changes from initial feasible solution to improved solution

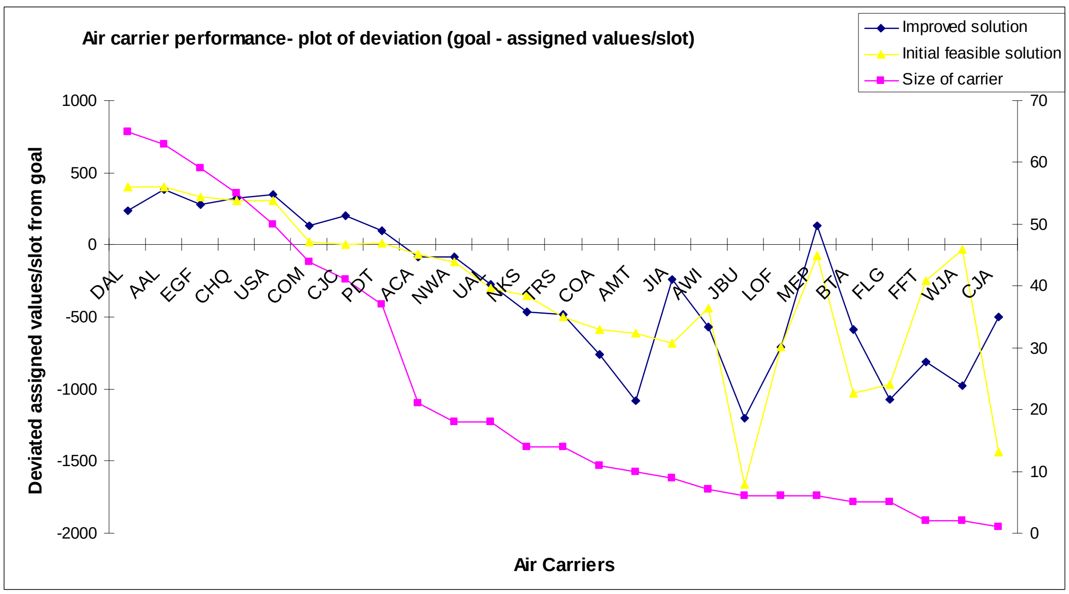


Figure 5-7 Air carrier performance change from initial feasible solution to improved solution

In terms of hourly metric y_t, which is the maximum of deviation between the goal and each air carrier's total number of assigned slot year percentage, the improved solution should be no better than the initial solution, as the latter has been optimized for each single hour *t*. However, as shown in Figure 5.6, the improved solution can reduce the range of deviations across different hours without dramatically affecting air carrier specific fairness measures in Figure 5.7.

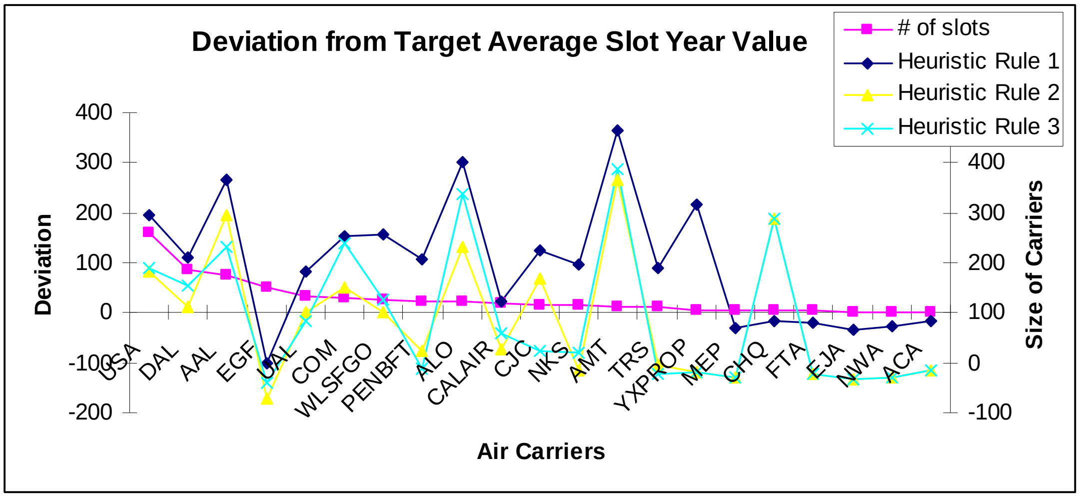


Figure 5-8 Airline deviations from target value per slot for heuristic methods

We now further compare the proposed heuristics. Figure 5.8 depicts the deviation from the target average value per slot of each air carrier among three heuristic methods using different rules. Figure 5.9 shows the deviation from the target total slot year values among these heuristics methods. Obviously, all three heuristic methods have similar performance based on the above fairness measures. Moreover, heuristics 1 and 3 tend to relieve the target value per slot deficit for airlines with more slots. Both heuristics 1 and 3 use a balanced start rule, so the following section will add additional rules and further compare the heuristic methods with the sequential optimization method.

By comparing the sequential optimization model and the proposed four heuristic rules through Figures 5.10- 5.12, we observe that the sequential procedure outperforms heuristics methods in terms of both the slot year hourly metric and slot value hourly metric. This can be explained by the fact that the sequential method has an improved solution based on the optimal solutions for each single hour. In terms of individual air carrier performance, Figure 5.12 also demonstrates that the sequential model produces a smoother air carrier deviation series, which implies more equitable resource assignment.

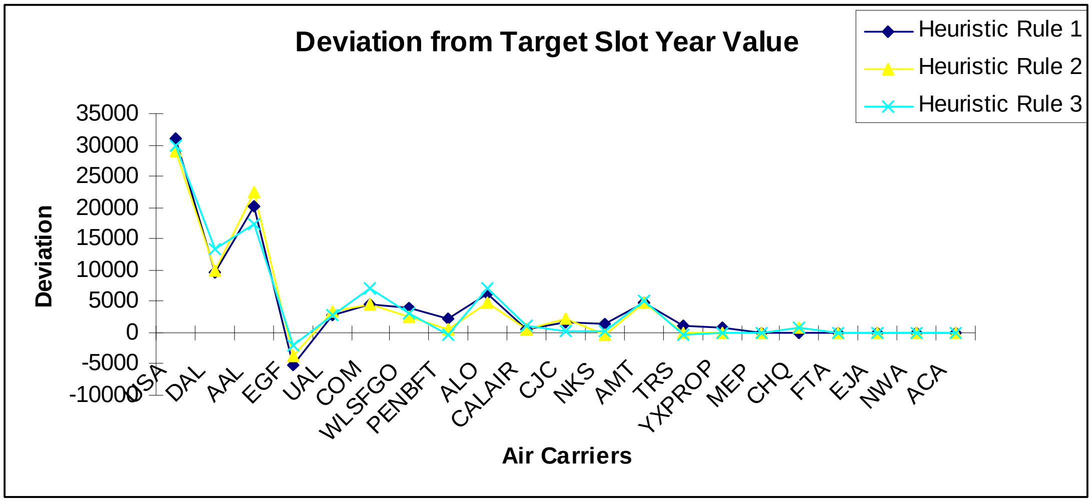


Figure 5-9 Air carrier deviations from target total slot year values for heuristic methods

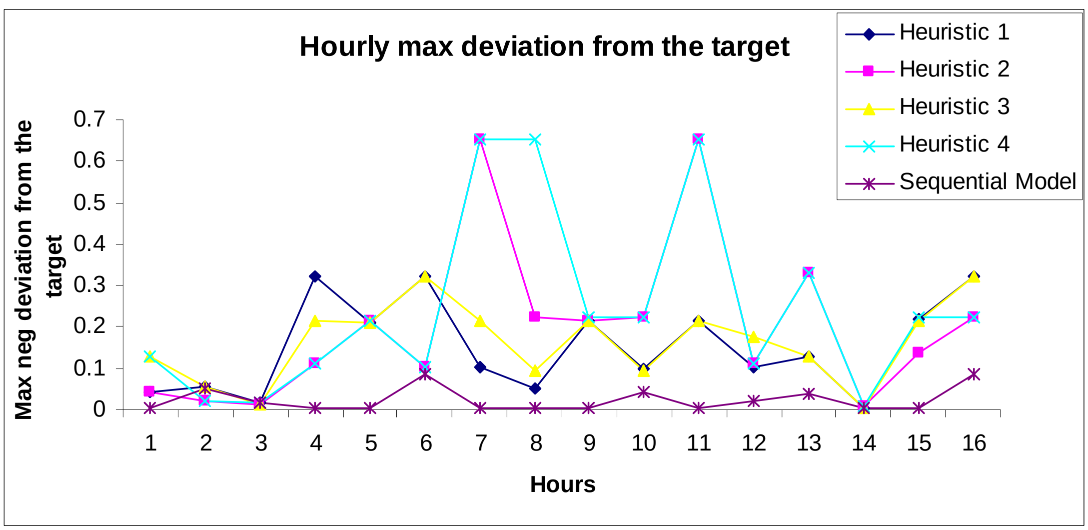


Figure 5-10 Hourly max deviations from target (using slot values percentage) from heuristic methods and sequential optimization models.

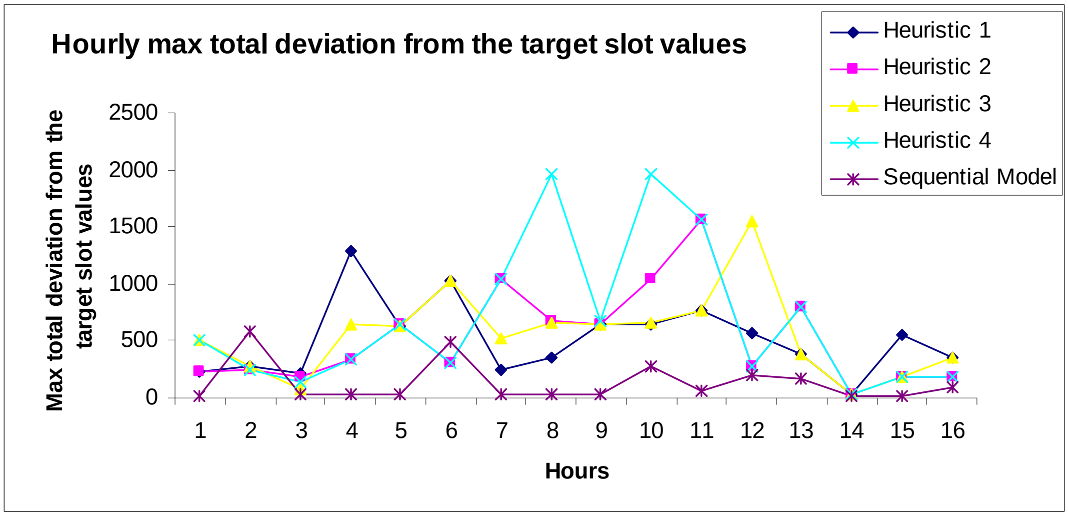


Figure 5-11 Hourly max total deviation plot (target total slot year values$_a$ – assigned total slot year values$_a$)

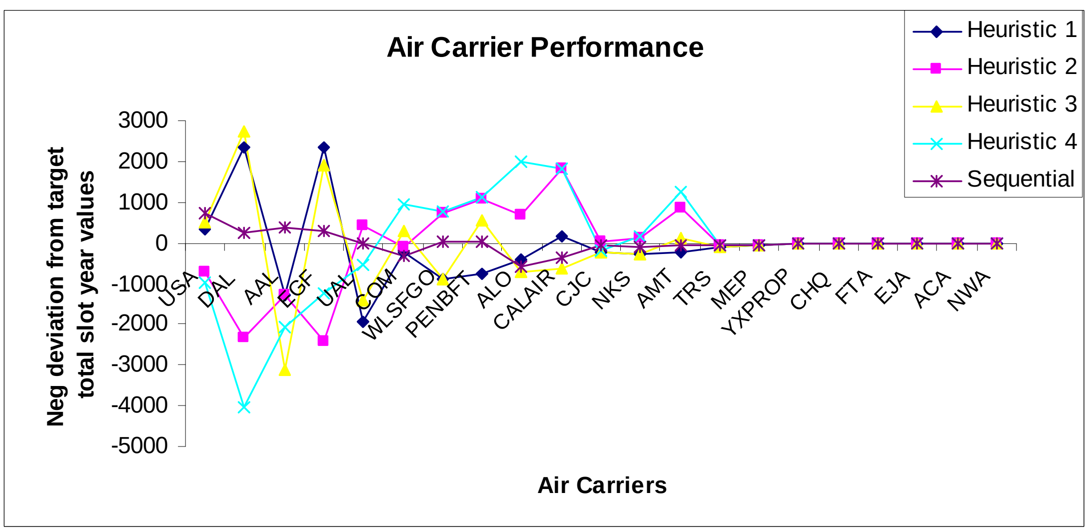


Figure 5-12 Air carrier performances in heuristic methods and sequential optimization models

5.7 Conclusion and Future Research

This chapter developed a long term slot lease assignment model. Several models and algorithms are developed to solve the slot assignment problem with equity consideration. The models and algorithms have been numerically evaluated extensively on hourly metric and air carrier specific metrics. Using the data from the LGA airport, experiment results show that the proposed sequential model solves the problem with very good solution quality with reasonable running time and resource.

The proposed models can be improved further with greater flexibility and different needs. The slot assignment can take into account air carriers' input as they are the parties that are influenced by the final decision. Although the proposed round-robin heuristic algorithms do not outperform the sequential optimization model based on experiment results, they still provide an option of adding flexibility in the assignment. For example, instead of making centralized assignment based on the values that we measure, a round-robin or a different interactive procedure could be adapted. After calculating the deviation and determining the assignment order, air carrier could select their own preferred subgroup bundle subject to certain restrictions. Furthermore, the paired assignment could be extended to 3-in-subgroup cases to enable more choices to further improve solution quality. Last, the models could be further extended to multi-day scenario, and each day's deviation could be compensated by the following day and so on. The models and methods discussed in this chapter in fact provide a starting point and possible directions for the long term slot lease assignment.

Chapter 6: Conclusions and Future Research Directions

Economic development leads to increasing air traffic demand which in turn poses increasingly stress to the National Airspace System (NAS). As a result, air traffic congestion is expected to remain as a top concern for the related public agencies and private corporations.

Table 6-1 Summary of modeling elements and contributions

Chapter	3	4	5
Topic	Airspace sector-level optimization	Airspace rerouting and ground holding decisions	Airport long-term slot assignment
Decisions to be made	Offload excess demand from different competing airlines in the congested airspace	Make flight-specific routing and ground holding decisions	Assign airport long-term slots to different air carriers
Models	Bi-criteria and ε-constraint integer programming models, network flow models with side constraints	Dynamic multi-commodity flow optimization model	Alternative models for incorporating equity metrics into assignment
Algorithms	Integer programming solution methods, efficient network flow solution algorithms with side constraints	Lagrangian relaxation to dualize stochastic capacity, non-anticipatory and equity assignment constraints	Sequential paired assignment heuristic algorithm that considers airline-specific equity metrics
Key Contributions	Construct and tested alterative network flow programming models for the resource allocation problem with equity considerations	Enable equitable assignment in flight rerouting and ground holding decision in a stochastic capacity environment	Adapt round robin scheduling principle for improving fairness measures across different airlines

As summarized in Table 6.1, this dissertation develops a number of model reformulations and efficient solution algorithms to address resource allocation problems in air traffic flow management, while explicitly accounting for equitability objectives in order to encourage further collaborations by different stakeholders, specifically air carriers with competing objectives when using capacity-limited airspace networks.

6.1 Sector-level Formulation with Equity Considerations

In Chapter 3, we discuss how to model the Flow Constraint Area (FCA) decisions and how to solve the corresponding airspace congestion problem in a real-time application (with a relatively short time period (one to several hours)). FCA/AFP is currently being used as a strategic approach to solve the airspace congestion problem due to demand/capacity imbalance or severe weather situation, but the current FCA/AFP approach does not consider the entire congested area or equity among air carriers when choosing offloaded flights.

After discussing some important modeling issues, e.g., exemption rules and offloading bias among airlines, Chapter 3 first develops a bi-criteria optimization model and a ε-constraint model to offload excess demand from different competing airlines in the congested airspace when the predicted traffic demand is higher than available capacity. Additional network flow-based reformulations, such as circulation models and the “flight on the node” model, are also developed for both single-sector and multi-sector cases. Computationally efficient network flow models with side constraints are developed and extensively tested using datasets obtained from the Traffic Flow Management Systems (TFMS) database. Representative Pareto-optimal tradeoff frontiers are consequently generated to allow decision-makers to identify best-compromising solutions based on relative weights and systematic considerations of both efficiency and equity.

The contributions of this chapter is to develop several integer programming and network flow programming models to solve the resource allocation problem with equity consideration, in particular for airspace single-sector and multiple-sector cases. Both computational running time and solution quality of those models are systematically evaluated.

6.2 Integrated Airspace Flight Rerouting and Ground Holding Problem

In Chapter 4, we further model and solve an integrated flight re-routing and ground holding problem on an airspace network. Given a network of airspace sectors with a set of waypoint entries and a set of flights belonging to different airline companies, the optimization model aims to minimize the total flight travel time subject to a set of flight routing equity, operational and safety requirements. A time-dependent network flow programming formulation is proposed with sector capacities and rerouting equity for each airline company as side constraints. Moreover, to consider stochastic airspace capacity under severe weather conditions, we use multiple scenarios to represent random realizations of predicted capacities, and further integrate non-anticipatory constraints to ensure the first-stage solutions across different scenarios have the same values. The routing equity is defined through an average travel cost threshold (per flight) for individual air carriers with a number of flights competing for the congested airspace.

A Lagrangian relaxation based method is used to dualize these side constraints and decompose the original complex problem into a sequence of simpler integer programming problems. If all three sets of side constraints are dualized, then the relaxed problems reduce into a sequence of linear programming problems with total unimodularity properties. By relaxing the coupling constraints between flights, the proposed Lagrangian relaxation-based solution method can separate the original problem into individual flight scheduling subproblems that can be efficiently solved by the shortest path algorithm in an expanded time-space network. The experiments investigate the computational time and solution quality gaps of different possible relaxations in the Lagrangian relaxation framework.

6.3 Airport Long-term Slot Assignment Problem

In Chapter 5, we develop an initial slot lease assignment model. In phasing out the High Density Rule, the FAA recognized the possibility that there could be an increase in congestion and delay at the affected airports. After exploring all possibilities, including do nothing, assigning based on a market mechanism, slot auctions, etc., the FAA proposed to assign the current landing slots finite lives with possible capacity reduction. Moreover, the expired landing slots are subject to reassignment, and flexible marketing mechanisms, such as auctions or congestion pricing.

Within a multi-objective utility maximization framework, this chapter proposes several practically useful heuristic algorithms for the long-term airport slot assignment problem. Alternative models are constructed to decompose the complex model into a series of hourly assignment sub-problems. A new paired assignment heuristic algorithm is developed to adapt the round robin scheduling principle for improving fairness measures across different airlines. Computational results are presented to show the strength of each proposed modeling approach.

6.4 Future Research

Equity in Air Traffic Flow Management

To ensure the fair allocation of en-route airspace resource, the flight operators and FAA should agree upon equity standards related to constrained airspace. In a real-time decision environment, future research needs to be conducted to systematically quantify the expectation of airlines on the fair slot and route assignment, as well as to dynamic calibration of the behavior model related to competing agents. With well-defined equity measures, it will pave the way for rapidly adapting the equity-oriented resource allocation mechanism in Air Traffic Flow Management applications.

Sector-level rerouting decisions with equity considerations

The models can be extended to incorporate the airlines' preference information. The proposed models are formulated in a centralized way, which mainly highlights the system efficiency side with equitable offloading assignment among airlines. It should be noted that, the relative importance of the flights for each airline is not modeled and a bi-level structure model can be explored to offer more control to the airlines. Moreover, the computational efficiency of the proposed models can be improved in order to meet the requirement arising in the context of real-time decision making where the decision should be rapidly modified if the congestion situation changes.

Integrated-routing and ground holding decisions with equity constraints

This study assumes stochastic sector capacity, but deterministic sector travel times, so a natural extension is to allow variable travel times and stochastic capacity for more realistic applications. However, because introducing any new problem dimensions typically increases the computational complexity quite steeply, it is undoubtedly vital to develop efficient and effective approximation and heuristic schemes. Further research will focus on how to extend a two-stage optimization model to multiple stages for emerging real-time adaptive routing applications. To search for high-quality solutions under tight equity constraints, we might need to propose alternative reformulations or solution methods to enforce the equity constraints, while allowing exceptions or compromises which should be systematically considered in a multi-objective decision-making framework with multiple agents. As the numerical experiment only tests the proposed algorithm on a small network, successful applications call for an extension and an adaptation of the current Lagrangian relaxation framework for producing optimal solutions for medium-sized or large-scale networks.

Equitable long-term airport slot assignment

The proposed models can be improved further regarding flexibility for meeting different modeling needs. The slot assignment will take into account airline's input as they are the parties that are influenced by the final decision. Although, according to the numerical experiments, the proposed heuristic algorithms do not outperform the sequential optimization model, they still provide a possibility of adding flexibility in the assignment. For example, instead of making the assignment according to the values that we measure, a more interactive procedure could be adapted. After sorting the deviation and determining the assignment order, airlines could select their own preferred subgroup assignment subject to certain restrictions. Furthermore, the paired assignment could be extended to 3-in-subgroup cases, where there will more choices and the solution quality could be further improved. Lastly, the models could be further extended to multi-day scenarios, and each day's deviation could be compensated by the following day and so on. The proposed models and methods provide a starting point and possible directions for the initial slot lease assignment.

Bibliography

Agustı´n, A., Alonso-Ayuso, A., Escudero, L.F., Pizarro, C., 2012. On air traffic flow management with rerouting. Part II: Stochastic case. *European Journal of Operational Research*, 219(1), pp.167-177.

Ahuja, R. K., Magnanti, T. L., Orlin, J. B., 1994. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, NJ.

Alonso-Ayuso, A., Escudero, L.F., Garin, A., Ortuno, M.T., Perez, G., 2003. An approach for strategic supply chain planning under uncertainty based on stochastic 0–1 programming. *Journal of Global Optimization*, 26(1), pp.97-124.

Andreatta, G., Romanin-Jacur, G., 1987. Aircraft flow management under congestion. *Transportation Science*, 21(4). pp 249-253.

Ball, M.O., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Peterson, E., Sherry, L., Trani, A., Zou, B. et al., 2010. *Total delay impact study*. NEXTOR Final Report prepared for the Federal Aviation Administration.

Ball, M.O., Barnhart, C., Nemhauser, G., Odoni, A., 2007. Air transportation: Irregular operations and control. In: Barnhart, C., Laporte, G., eds. 2007. *Handbook of Operations Research and Management Science: Transportation*, Amsterdam: Elsevier, pp 1-67.

Ball, M.O., Donohue, G., Hoffman, K., 2005. Auctions for the safe, efficient and equitable allocation of airspace system resources. In: Cramton, P., Shoham, Y., Steinberg, R., eds. 2005. *Combinatorial Auctions*, Cambridge: MIT Press, pp.507-538.

Ball, M.O., Hoffman, R.L., Knorr, D., Wetherly, J., Wambsganss, M.C., 2001. Assessing the benefits of collaborative decision making in air traffic management. *Proc. 3rd USA/Europe Air Traffic Management R&D Seminar*, Napoli, Italy, June 13-16.

Ball, M.O., Hoffman, R.L., Mukherjee, A., 2010, Ground delay program planning under uncertainty based on the ration-by-distance principle. *Transportation Science*, 33(1), pp.1-14.

Ball, M.O., Hoffman, R.L., Odoni, A.R., Rifkin, R. 2003. A stochastic integer program with dual network structure and its application to the ground-holding problem. *Operations Research*, 51(1), pp.167-171.

Ball, M.O., Lovell, D.J., Hoffman, R., Mukherjee, A., 2005. Response mechanisms for dynamic air traffic flow management, *Proceedings of 6thUSA/Europe Air Traffic Management R&D Seminar*, Baltimore, MD, June 2005.

Ball, M.O., Lulli, G., 2004. Ground delay programs: Optimizing over the included flight set based on distance. *Air Traffic Control Quarterly*,12(1), pp.1-25.

Ball, M.O., Hoffman, K., Ausubel ,L., Berardino, F., Cramton, P., Donohue, G., Hansen, M. Lovell, D. 2005. NEXTOR Congestion Management Project: Final Report. Prepared for Federal Aviation Administration and Office of the Secretary.

Barbosa-Povoa, A. P., Matos, P. L., Rocha, L., 2001. Optimization approach to support the grouping and scheduling of air traffic control sectors, *Lecture Notes in Economics and Mathematical Systems*, 505(4), pp.399-414.

Beatty, R., Hsu, R., Berry, L., Rome, J., 1998. Preliminary evaluation of fight delay propagation through an airline schedule. *Proceedings of the Second USA/Europe Air Traffic Management R & D Seminar*, Orlando, December 1-4, 1998.

Bertsimas, D. J., Lulli, G., Odoni, A. R., 2011. An Integer Optimization Approach to Large-Scale Air Traffic Flow Management. Operations Research, 59(1), pp. 211–227.

Bertsimas, D.J., Stock S., 1998. The air traffic flow management problem with enroute capacities. *Operations Research*, 46(3), pp.406-422.

Bertsimas, D.J., Stock S., 2000. The traffic flow management rerouting problem in air traffic control: A dynamic network flow approach. *Transportation Science*, 34(3), pp.239-255.

Bianco, L., Bielli, M., 1993. System aspects and optimization models in ATC planning. In: Bianco, L., Odoni, A. R., eds. 1993. *Large Scale Computation and Information Processing in Air Traffic Control*, Berlin: Springer- Verlag, pp.47-99.

Bureau of Transportation Statistics, 2011. *Transportation statistics*. [online] Available at: < http://www.transtats.bts.gov > [Accessed February 01, 2009].

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

Chang, K., Howard, K., Oiesen, R., Shisler, L., Tanino, M., Wambsganss, M. C., 2001. Enhancements to the FAA ground-delay program under collaborative decision making. *Interfaces*, 31(1), pp.57-76.

Chang, Y., 2010. *Stochastic programming approaches to air traffic flow management under the uncertainty of weather*. Ph.D. Georgia Institute of Technology.

Churchill, A.M., Lovell, D.J., 2011. Assessing the impact of stochastic capacity variation on coordinated air traffic flow management. *Transportation Research Record*, 2214, pp.111-116.

Churchill, A. M., Lovell, D.J., 2012. Coordinated aviation network resource allocation under uncertainty. *Transportation Research Part E*, 48(1), pp.19-33.

Churchill, A.M., Lovell, D.J., Ball, M.O., 2010. Flight delay propagation impact on strategic air traffic flow management. *Transportation Research Record*, 2177, pp.105-113.

Dantzig, G. B., Fulkerson, D. R., 1954. Minimizing the number of tankers to meet a fixed schedule. *Naval Research Logistics Quarterly*. 1(3), pp.217-222.

Ganji, M., Lovell, D.J., Ball, M.O., Nguyen, A., 2009. Resource allocation in flow-constrained areas with stochastic termination times. *Transportation Research Record*, 2106, pp.90-99.

Gleimer, E., 1996. Slot regulation at high density airports: how did we get here and where are we going? *Journal of Air Law and Commerce*, 61(4), pp.877-931.

Goodhart, J. R., 2000. Increasing airline operational control in a constrained air traffic system. Ph.D. University of California, Berkeley.

Helme M.P., 1992. Reducing air traffic delay in a space-time network. *IEEE International Conference on Systems, Man and Cybernetics*, 1, pp.236-242.

Hoffman, R., Ball, M., 2000. A comparison of formulations for the single-airport ground-holding problem with banking constraints. *Operations Research*, 48(4), pp.578-590.

Hoffman, R., Mukherjee, A. and Vossen, T. W. M. 2011. Air Traffic Management. Wiley Encyclopedia of Operations Research and Management Science.

Libby, M, Buckner J., Brennan, M. 2005, Operational concept for Airspace Flow Programs (AFP). Version 1, FAA Air Traffic Organization, System Operations Services, Washington, DC.

Lindsay, K. S., Boyd,E. A., Burlingame, *R*., 1993. Traffic flow management modeling with the time assignment model. *Air Traffic Control Quarterly*, 1(3), pp.255-276.

Liu, B., Hansen, M., Mukherjee, A., 2008. Scenario-based air traffic flow management: From theory to practice. *Transportation Research Part B*, 42(7-8), pp.685-702.

Lulli, G., Odoni, A.R., 2007. The European air traffic flow management problem. *Transportation Science*, 41(4), pp.431-443.

Mukherjee, A., Hansen, M., 2007. A dynamic stochastic model for the single airport ground holding problem. *Transportation Science*, 41(4), pp.444-456.

Mukherjee, A., and Hansen, M. . 2009. A dynamic rerouting model for the air traffic flow management. *Transportation Research Part B: Methodological*, 43(1), pp. 59-171.

Mulvey, J.M., Ruszczyński, A., 1992. A diagonal quadratic approximation method for large scale linear programs. *Operations Research Letters*, 12(4), pp.205-215.

Niznik, T., 2001. Optimizing the airline response to ground delay programs. *Proceedings of AGIFORS Airline Operations Meeting*, Ocho Rios, Jamaica.

Norkin, V. I., Pflug, G. C., Ruszczynski, A., 1998. A branch and bound method for stochastic global optimization. *Mathematical Programming*, 83(1-3), pp.425-450.

Odoni, A. R., 1987. The flow management problem in air traffic control. In: Odoni, A. R., Bianco, L., and Szego, G., eds. 1987. *Flow control of congested networks*. 269-288, Berlin: Springer-Verlag, pp 269–288.

Pourtaklo, N. V. 2009. Preference Based Fair Allocation Of Limited Resources,, Ph.D. Dissertation, University of Maryland, College Park.

Richetta, O., Odoni, A.R., 1993. Solving optimally the static ground-holding problem in air traffic control. *Transportation Science*, 27(3), pp. 228-238.

Rios, J., Ross, K., 2010. Massively parallel Dantzig-Wolfe decomposition applied to traffic flow scheduling. *Journal of Aerospace Computing, Information, and Communication*, 7(1), pp.32-45.

Rockafellar, R. T., Wets, R. J., 1991. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*. 16(1), pp.119-147.

Ruszczyński, A., 1989. An augmented Lagrangian decomposition method for block diagonal linear programming problems. *Operations Research Letters*, 8(5), pp.287-294.

Schumer C.E., Maloney, C.B., 2008. *Your fight has been delayed again: fight delays cost passengers, airlines, and the US economy billions*. Washington D.C.: The US Senate Joint Economic Committee.

Sherali, H.D., Hill, J.M., McCrea, M.V., 2011. Integrating slot exchange, safety, capacity, and equity mechanisms within an airspace flow program. *Transportation Science*, 45(2), pp.271-284.

Sherali, H.D., Staats, R.W., Trani, A.A., 2003. An airspace planning and collaborative decision-making model: Part I - probabilistic conflicts, workload, and equity considerations. *Transport Science*, 37(4), pp.434-456.

Sherali, H.D., Staats, R.W., Trani, A.A., 2006. An airspace-planning and collaborative decision-making model: Part II - cost model, data considerations, and computations. *Transport Science*, 40(2), pp.147-164.

Sun, D., Bayen, A., 2008. Multicommodity Eulerian–Lagrangian large-capacity cell transmission model for en route traffic. *Journal of Guidance, Control, and Dynamics*, 31(3), pp.616-628.

Sun, D., Clinet, A., Bayen, A., 2011. A dual decomposition method for sector capacity constrained traffic flow optimization, *Transportation Research Part B*, 45(6), pp.880-902.

Terrab, M., Odoni, A. R., 1993.Strategic flow management for air traffic control. *Operations Research*, 41(1) pp 138-152.

Tosic, V., Babic, O., Cangalovic, M., Hohlacov, D., 1995. Some models and algorithms for en route air traffic flow management. *Transportation Planning and Technology*, 19(2), pp.147-164.

Vossen, T., 2002. Fair allocation methods in air traffic management. Ph.D. University of Maryland, College Park.

Vossen, T., Ball., M.O., 2006. Slot trading opportunities in collaborative ground delay programs. *Transportation Science*, 40(1), pp. 29-43.

Vossen, T., Ball, M.O., Hoffman, R., Wambsganss, M., 2003, A general approach to equity in traffic flow management and its application to mitigating exemption bias in ground delay programs, *Air Traffic Control Quarterly*, 11(4), pp.277-292.

Vranas, P. B., Bertsimas, D. J., Odoni, A. R., 1994. The multi-airport ground-holding problem in air traffic control. *Operations Research*, 42(2), pp.249-261.

Warburton, A., 1987. Approximation of Pareto optima in multiple-objective shortest path problems. *Operations Research*, 35(1), pp.70-79.

Wolsey, L. 1998, *Integer Programming*, John Wiley, New York.

Young, H. P., 1994. *Equity: In Theory and Practice*. Princeton University Press

Yu, G. ed. 1997. *Operations research in the airline industry*. Boston, MA: Kluwer Academic Publishers.