ABSTRACT

Title of Dissertation: FAIR ALLOCATION METHODS IN AIR TRAFFIC MANAGEMENT

Thomas W.M. Vossen, Doctor of Philosophy, 2002

Dissertation directed by: Professor Michael O. Ball
R.H. Smith School of Business

Recently, the Federal Aviation Administration (FAA) and the major airlines in the U.S. have embraced a new initiative to improve Air Traffic Flow Management. This initiative, called *Collaborative Decision Making* (CDM), is based on the recognition that improved data exchange and communication between the FAA and the airlines will lead to better decision making. In particular, the CDM philosophy emphasizes that decisions with a potential economic impact on airlines should be decentralized and made in collaboration with the airlines whenever possible. This dissertation is motivated by the fairness issues that arise in the resource allocation procedures that have been introduced under CDM.

While the fair allocation of resources has been and continues to be a major concern in the procedures that have been developed under CDM, its interpretation is oftentimes left implicit. In this dissertation, we introduce and evaluate

several potential approaches to fair allocation, using both multi-objective optimization models and cooperative game theory models. Subsequently we study how the dynamic nature of flow management impacts fairness, and introduce methods that may be used to manage the allocation of resources in this environment. In addition, we also consider the opportunities for increased airline control in a CDM-based environment. In particular, we study the potential benefits that can be obtained by the introduction of a framework in which airlines dynamically trade resources.

FAIR ALLOCATION METHODS IN AIR TRAFFIC MANAGEMENT

by

Thomas W.M. Vossen

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
2002

Advisory Committee:

Professor Michael O. Ball, Chairman/Advisor
Professor Gnana Anandalingam
Professor Michael C. Fu
Dr. Robert L. Hoffman
Professor Paul M. Schonfeld
Professor H. Peyton Young

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Michael O. Ball. It is safe to say that this dissertation would not have been written without his constant support, encouragement, guidance, and (most of all) patience. It is difficult to express just how much I appreciate all his help.

I am also indebted to the other members of my committee: Dr. Gnana Anandalingam and Dr. Paul Schonfeld for their many valuable suggestions and comments, Dr. H. Peyton Young for taking the time to discuss our initial forays into fairness, Dr. Michael Fu for his support throughout my stay at the University of Maryland, and Dr. Robert Hoffman for introducing me to the many intricacies of air traffic flow management during those late night Seil Lab sessions.

I should also thank all the office mates who have helped and had to put up with me over the years, from Andrew Vakhutinski and

Heshmat (sorry Rosa) Beirjandi to the Nextor office gang: Bala Chandran, Jason Burke, and Narender Bhogadi. I especially want to thank Guglielmo "Giullare" Lulli and Chien-Yu Chen, for being great guys and great friends. Thanks also to my fellow Dutch expatriates Manny Blits and Frank Groen, for keeping me (at least somewhat) sane along the way. And I shouldn't forget "good old" Toon Bruining, for keeping in touch through the years.

Finally, I would like to thank my family: Diana, Wilbert, and "Oos Mam en Oos Pap". You have always been there when I needed you. I love you and I miss you.

This work was supported in part by the U.S. Federal Aviation Administration through NEXTOR, the National Center of Excellence for Aviation Operations Research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FAIR ALLOCATION METHODS IN
AIR TRAFFIC MANAGEMENT

Thomas W.M. Vossen

February 4, 2003

This comment page is not part of the dissertation.

Chapter 1

Introduction

In the last several decades, the growth in air traffic has been dramatic. From a relatively minor industrial sector, air transportation has evolved into a branch of the economy which currently accounts for 6% of the Gross Domestic Product in the United States, and employs approximately 1.5 million people ([25]). Moreover, there are no signs that this growth is slowing down. Indeed, current projections expect air traffic to grow at an annual rate of 3% to 5% over the next 15 years.

Unfortunately, however, the increase in air traffic at the major airports in the United States has vastly outgrown the increase in airport resources. As a consequence, the level of congestion has risen consistently, leading to increased delay during peak periods of travel. These delays result in substantial costs: in 1995, the FAA estimated that the cost of delays to the airlines was approximately $2.5 billon in operating expenses ([25]). As such, it is clear that the imbalance between stagnating capacity and increasing demand has (and will have) an enormous impact on the performance of the air transportation system.

Not surprisingly, the current level of delays and projected increase in demand have led to a number of initiatives that aim to alleviate congestion. These initiatives are both varied and numerous. Some airports are considering increases in capacity by adding runways. Other initiatives consider the potential of demand management measures, such as the use of auctions as LaGuardia Airport and Congressional regulation that would allow airlines to coordinate schedule reductions at certain airports ([57]). In addition, the FAA has implemented (and is considering) procedural changes during the management of daily operations which aim to increase flexibility.

So far, these efforts to reduce congestion have perhaps had their biggest impact on the management of daily operations. Until recently the management of daily operations was largely centralized, in that the FAA would unilaterally make all relevant decisions and force airlines to operate within narrow guidelines. Spurred by a joint government-industry effort known as Collaborative Decision Making (CDM), however, the last five years have seen a major shift in this paradigm. The major philosophical components of CDM are: (1) improved data exchange and communication between the FAA and the airlines will lead to better decision making in air traffic flow management and (2) that, whenever possible, those decisions which have a potential economic impact on airline operations should be decentralized and made in collaboration with the airlines.

While the CDM paradigm encompasses a wide range of applications in air traffic flow management, its primary focus so far has been the implementation and enhancement of Ground Delay Programs, which are used to manage periods of congestion at an airport. The number of enhancements that have recently been implemented are numerous: examples include improved data-exchange, better situational awareness tools, and increased flexibility for the airlines. Without a doubt, however, the biggest changes have come through the introduction of new methods for the allocation of available resources. These procedures have had a profound impact on the interaction between the FAA and the airlines, in that they have solidified the FAA's role as a discoverer of constraints and as an arbiter of rationed capacity. The resulting allotments of scarce capacity allow airlines to trade off operating options based on internal business objectives.

1.1 Motivation

This dissertation is motivated by the fairness issues that arise in the allocation procedures that have been introduced under CDM. Fairness concerns have played an important role throughout the development of the allocation procedures, and continue to be an essential factor whenever extensions or modifications to these procedures are proposed. It is therefore surprising that, oftentimes, it is not clear what is meant by fairness within the context of the procedures developed under CDM. Because the notion of fairness is largely left implicit in the procedures, there is no well-defined set of principles that defines what constitutes a fair distribution of the resources. Moreover, it is not obvious how the concepts embedded in the different procedures relate to each other and to the metrics that are used to measure equity ex-post (for analysis purposes). As such, the absence of an overall set of guiding principles complicates the extension of CDM to a more general environment (e.g., the management of en-route resource constraints).

It is important to note, however, that the use of fairness as a basis for allocating scarce resources presents a immediate restriction in the focus of our research. While the attention to fairness has evolved into one of the pillars of

CDM, a number of other approaches could also be considered. In fact, a number of different approaches to the allocation of resources have been proposed, ranging from auctions ([61]) and congestion pricing ([57]) to bargaining schemes ([2]). However, these proposals address the allocation of airport arrival slots in the long run, that is, the resources assigned may be viewed as long-term capacity reservations. The allocation of slots under CDM, on the other hand, is markedly different, in that slots have to be assigned on a day to day basis due to fluctuations in an airport's capacity (caused by weather conditions). This introduces a number of complications, such as the significant levels of uncertainty, the complexities of airline trade-offs, and the dynamic nature of the allocation processes, which significantly complicate each of the above mentioned possibilities. As such, the attention to fairness may be more amenable in this environment.

Therefore, the main purpose of the research in this dissertation is the development of fair resource allocation mechanisms in a collaborative air traffic management environment. Our first objective is to analyze potential concepts of fairness that might be applicable in this environment. A subsequent objective is to show how these principles can be applied to devise fair allocation mechanisms that can be used within a context that is characterized by significant dynamics and uncertainty. Finally, we also consider the opportunities for increased airline control in a CDM-based environment. In particular, we study the potential benefits that can be obtained by the introduction of a framework in which airlines dynamically trade resources.

1.2 Outline and Research Contributions

The remainder of this dissertation is organized as follows.

Chapter 2 presents a brief overview of air traffic management, in particular the management of daily operations. We summarize the flow management initiatives employed by the FAA, as well as the airlines' response to these initiatives.

Chapter 3 discusses the current move towards decentralization of air traffic management, with a focus on the Collaborative Decision Making paradigm and the related notion of "Free Flight". We present an overview of the allocation procedures introduced under CDM, and discuss their relationship to other potential approaches. In particular, we motivate the use of fairness as a basis for resources allocation decisions in the management of daily operations.

Chapter 4 investigates concepts of fairness for the allocation of arrival slots under CDM. The fair allocation of arrival slots poses a number of fundamental questions. Who are the slots to be assigned to, i.e. who are claimants ? On what basis do we compare the claimants' demands ? Given such a basis for comparison, what are the resulting allocation mechanisms and how applicable are they within the context of Ground Delay Programs ? To address these questions, we first interpret the problem as a cooperative game in which claimants share the delay imposed by their respective demands. This approach, however, appears to be less applicable within the context of GDPs. We therefore pursue a more direct approach, in which we postulate a number of intuitive axioms and characterize the resulting class of allocation mechanisms. Besides the mechanism currently used under CDM, this yields a number of potential alternatives. We analyze the differences between these methods, and compare their allocations using historical GDP data. The research contribution in Chapter 4 are twofold.

First, we introduce formal fairness concepts within the framework of air traffic flow management. Secondly, we derive a new class of allocation schemes based on the approach discussed by Young ([86]), which extend the models proposed by Moulin ([48]).

The mechanisms discussed in Chapter 4 define fair shares of the resources for each airline. In Chapter 5, we propose methods to approximate these shares in situations where the "ideal" may not be attainable. A practical motivation for these procedures stems from the dynamic nature of GDPs. We show how these methods yield a unified approach to the different allocation procedures currently used under CDM. Moreover, we discuss how these methods may be applied to reduce certain systematic biases caused by the timing of GDPs. The main contribution of Chapter 5 is the introduction of a general framework for the allocation of slots during GDPs, based on a novel application of models developed for balanced just-in-time scheduling problems. In addition, Chapter 5 shows the extent to which practical issues can affect fairness, and proposes methods to mitigate the resulting biases.

Chapter 6 explores opportunities for increased coordination during Ground Delay Programs. In particular, we propose a general framework by which the airlines can trade arrival slots, in which the FAA acts as a mediator, and introduce an optimization model for the mediation problem. Using two different models of airline decision-making, we evaluate the potential benefits of increased coordination. The research contribution of Chapter 6 is first that it shows the potentially significant benefits of increased coordination. In addition, we introduce novel IP formulation for the mediation problem and demonstrate their efficiency in near real-time settings.

Chapter 7 provides conclusions and discusses areas for further research.

Chapter 2

Air Traffic Management

The air transportation system in the U.S. is one of the most complex logistical systems imaginable. On a daily basis, the system supports approximately 60,000 flights of commercial, military, and general aviation aircraft, and as many as 6,000 aircraft may simultaneously occupy the airspace. Besides the sheer volume, the air transportation system is further complicated by significant variations in airspace capacity (due to factors such as fluctuating weather conditions and equipment outages). It is therefore safe to say that the coordination of air traffic presents a formidable task, which requires a multitude of processes and involves a large number of stake holders. The broad term "Air Traffic Management" is commonly used to represent the overall collection of these processes.

This chapter presents a general overview of Air Traffic Management, with a particular focus on operational decision and coordination processes. We start with a high-level classification of Air Traffic Management initiatives, which primarily serves to clarify the context in which operational decisions are made. Next, we describe the major operational decision processes employed by the FAA, and review the manner in which airlines respond to these initiatives. To

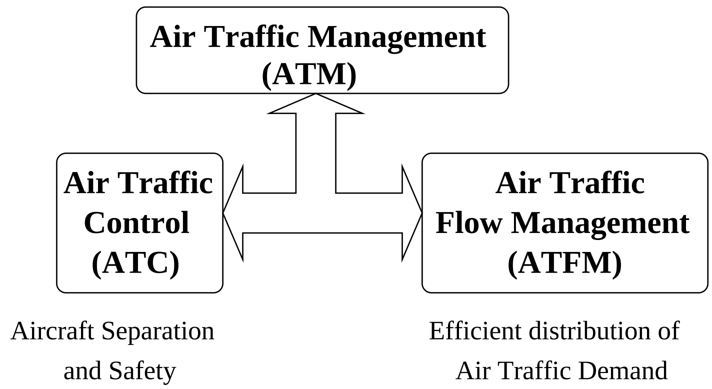


Figure 2.1: ATM Components

conclude, we discuss the (often implicit) decision-making hierarchy and summarize important characteristics of the ATM environment.

2.1 Air Traffic Control vs. Air Traffic Flow Management

Air Traffic Management (ATM) can be defined as the composite of processes that support the ultimate goal of safe, efficient, and expeditious aircraft movement. It is common to distinguish two basic ATM components: Air Traffic Control and Air Traffic Flow Management (see Figure 2.1). Air Traffic Control (ATC) refers to processes that provide tactical separation services, that is, real-time separation procedures for collision detection and avoidance. As such, ATC actions are of a more "microscopic" nature and primarily address immediate safety concerns of airborne flights. Air Traffic Flow Management (ATFM), on the other hand, refers to processes of a more "macroscopic" nature. Typically, ATFM considers strategic procedures, which aim to detect and resolve demand-capacity imbalances by adjusting aggregate traffic flows to match scarce capacity resources. Accordingly, ATFM actions have a greater potential to address system efficiency.

For the majority of the previous century, the coordination of air traffic proceeded largely through tactical air traffic control procedures. This was deemed sufficient, as the demand for air traffic was generally well within the capacity limits. Periodic congestion was usually resolved by procedural changes or technologic advances (see [55] for a comprehensive review of the evolution of ATM). It was not until the aftermath of the air traffic controllers' strike of 1981 that the FAA first implemented a systematic form of flow management known as ground holding. Under ground holding, aircraft departures are restricted until it is determined that sufficient airspace is available for each aircraft[1]. Initially, the use of ground holding was primarily instituted to reduce workload for the inexperienced controllers that were hired in the wake of the mass firings that accompanied the strike. However, the continued growth in air traffic that followed the airline deregulation act of 1978, as well as changes in traffic patterns[2], gradually increased the scope of ATFM practices. Over the past two decades, the levels of congestion in the system have risen consistently (see [25]), which has resulted in increasing delays during peak periods of travel. The use of ATFM initiatives has therefore become increasingly important, and will undoubtedly play an even more important role in the future.

A systematic description of the application of flow management to resolve air traffic congestion is given by Odoni [56], who classifies ATFM initiatives as

[1] A more detailed description follows in Section 2.2.

[2] Caused in particular by the so-called "hub and spoke" scheduling practices used by airlines.

long-, medium-, or short-term:

- Long-term approaches typically focus on increasing capacity. Examples include the construction of additional airports (which may take 10 to 15 years), the introduction of new technologies (e.g., satellite-based navigation tools), and the addition of runways to existing airports. Though effective, such initiatives are usually very costly and may be difficult to implement[3].

- Medium-term approaches are mostly administrative or economic in nature, and try to alleviate congestion by modifying spatial or temporal traffic patterns. For example, at some airports flight schedules are coordinated bi-annually according to IATA guidelines ([32]). Recently, Congress proposed a bill that would allow airlines to coordinate flight schedule reductions at congested airports (The HD1407 bill, [78]). Similar medium-term approaches include the recent use of slot lotteries at LaGuardia Airport ([23]), as well as current proposals for slot auctions and congestion pricing.

- Short-term approaches consider the strategic adjustment of air traffic flows to match available capacity, and typically span a planning horizon that is less than 24 hours. These operational ATFM initiatives attempt to mitigate the unavoidable congestion that may arise from unforeseen and unpredictable disruptions as efficiently as possible. Such periods of congestion arise frequently when bad weather causes sudden capacity reductions.

[3]Airport expansions frequently encounter the resistance of local communities and other special interest groups, who may be concerned with noise, real estate depreciation and other factors; Moreover, they are usually subject to strict environmental regulations.

Throughout this dissertation we focus on strategic, short-term ATFM initiatives. It is important to note that these operational processes are a critical and indispensable part of ATFM: while long- and medium-term initiatives may help to alleviate congestion, the significant impact of weather conditions on airspace capacity[4] make it unlikely that periodic congestion can ever be eliminated. In the remainder of this dissertation, we will use the term ATFM to represent only these short-term initiatives.

2.2 Air Traffic Flow Management Initiatives

In the U.S., the Federal Aviation Administration (FAA) is responsible for the coordination of air traffic. Its primary task is the enforcement of proper separation requirements in the controlled airspace. To carry out this function, the FAA has divided the airspace in the continental United States into 22 areas. Aircraft separation responsibility within each area belongs to associated Air Route Traffic Control Centers (ARTCCs). Because a single controller cannot handle all aircraft within an ARTCC's boundaries, each ARTCC is further divided (both vertically and horizontally) into 20 to 80 smaller areas called sectors. Air Traffic Controllers guide aircraft from sector to sector until they arrive within roughly 200 miles of their destination airports, at which point control of the aircraft is assumed by terminal radar approach control facilities (TRACONs). Finally, airport towers control aircraft while they taxi to and from runways and during takeoffs and landings. Accordingly, the ATC functions performed by the FAA

[4] It is not unusual that occurrences of bad weather reduce airport capacities by a factor of 2 or 3.

form a highly distributed process. Air traffic controllers (cf. TRACON/control tower representatives) are only responsible for the movement of aircraft within their region of airspace, and their decisions are mainly based on local and near real-time information about the flights entering their sectors. Typically, there is little coordination in ATC procedures; coordination occurs largely between controllers at adjacent sectors, by handoff procedures that transfer the responsibility for an aircraft when it passes sector/facility boundaries[5].

The (strategic) ATFM functions performed by the FAA, on the other hand, are primarily coordinated by the FAA's command center, the Air Traffic Control Systems Command Center (ATCSCC). The ATCSCC continuously monitors current and projected demand within the NAS, and identifies system constraints or other conditions (e.g. weather) that may affect the capacity limits. Whenever it is predicted that demand will exceed capacity limits within a 15-minute interval, FAA regulation mandates a response. In that case, the ATCSCC generates and implements strategies to resolve the projected congestion. The short-term flow management procedures that are used most often are ground delay programs, metering, and rerouting. These initiatives may be outlined as follows.

- Ground Delay Programs (GDPs) are used in response to periods of airport congestion. Typically, this is caused by a reduction in the airport's arrival capacity due to bad weather (although airport construction, special runway operations and limited surface capacity may also be possible reasons). In a GDP, flights bound for congested airports are delayed on the ground, so as to balance the total arrivals with the reduced capacity at the airport under

[5]Occasionally though, controllers may also be concerned with downstream effects, so as to prevent the simultaneous operation of too many aircraft in an area ([25])

consideration. Ground holding therefore consists of delaying a flight's take-off beyond its scheduled arrival time. The underlying motivation is that, as long as a delay is unavoidable, it is both safer and less costly for the flight to absorb this delay on the ground before take-off.

GDPs are the most important traffic management procedure used by the ATCSCC; in spite of the fact that GDPs can only control aircraft destined for a single airport, they are sometimes even used to help resolve congestion in other areas of the airspace. Closely related to GDPs are are so-called Ground Stops, which are implemented when an airport has an unexpected problem (e.g. a runway closure or a severe snowstorm). Ground stops allow the ATCSCC to stop all inbound traffic (i.e. delay their departure) to reduce traffic flows. When ground stops become excessive or delays can be foreseen, a regular GDP often follows the ground stop(see also [24]).

- Metering restrictions control traffic flows in the enroute environment. Metering procedures may be subdivided into (1) time-based metering, which controls the time at which an aircraft is to pass over a certain geographical point, and (2) distance-based metering, which places a limit on how closely aircraft can follow each other. Distance-based metering is better known as "Miles-In-Trail", which specifies a minimum separation (in miles) between aircraft moving in the same direction.

  Time-based metering is used primarily when excessively large airborne holding queues have built up around an airport (e.g. due to severe capacity reductions or airport closures). In such cases, time-based metering can be used to control holding patterns precisely, and to efficiently space aircraft for final approach. Miles-In-Trail restrictions are commonly used in conjunction with so-called Enroute Spacing Programs, to manage the (merging of) traffic streams entering an airport's terminal area (cf. [24]).

- Rerouting of aircraft occurs primarily when bad weather threatens the accessibility of certain regions of the airspace. Oftentimes, rerouting is instituted as part of Severe Weather Avoidance Programs (SWAPs), which are typically enacted when traffic flows are affected by widespread severe weather in the airspace. SWAP plans usually have a major impact on air traffic, and oftentimes include metering restrictions and/or GDPs along with rerouting.

In addition to these major initiatives, there are also a number of procedures with a smaller scope. For instance, Low Altitude Arrival and Departure Routes (LAADR) embodies a set of procedures for the use of low altitude routes to avoid congested airspace, and Coded Departure Routes (CDR) involves procedures and a database for the creation, storage, and dissemination of alternate routes used to avoid airspace blocked by severe weather. Other examples include the Pacific Track Advisory Program, which is used to allocate a series of tracks for aircraft to transit the North Pacific from U.S. airports to airports in Asia, and the National Route Program (NRP), which allows airlines to file flight plans other than those normally preferred by the FAA ([1]). Typically, such processes are of a more "local" nature, in that they are not (or only partially) coordinated by the ATCSCC. The reason for this is that they usually apply only to certain specific region of airspace and heavily rely on local conditions. Generally speaking however, we may classify the ATFM actions employed by the FAA as (1) imposing ground delays, (2) imposing airborne delays, and (3) imposing alternate routes.

2.2.1 Airline Response

An airline's operational objectives are usually markedly different from those that underly the FAA's ATFM initiatives: whereas the FAA is concerned with aggregate flows and capacity limits, the ultimate goal of airline operational control is to preserve its published flight schedule. An airline's flight schedule represents its primary product, and often reflects its competitive strategy.

Airlines typically coordinate their daily operations at centralized Airline Operational Control Centers (AOCs), which interact with airport and maintenance stations and with individual pilots. Schedule preservation needs to consider both individual flights and schedule interdependencies. Therefore, airline operations require a level of coordination that is usually much higher than it is for the FAA, because of the potential cascading effects of flight delays[6]. This presents a challenge in particular when airlines face so-called *irregular operations*, that is, when they need to respond to ATFM restrictions imposed by the FAA or to other schedule disruptions .

Important functions that need to be performed by airline operational control include the following (also see [1] and [24]):

- *Schedule Adjustment.* On a daily basis, unforeseen events, such as delays or mechanical problems, may disrupt an airline's flight schedule. To prevent the cascading effects these disruptions may have, the AOC will make schedule adjustments that allow a return to a more balanced condition.

[6]The propagation of delays is of course caused by connections that passengers, flight crews, and aircraft oftentimes have to make.

Schedules may be adjusted in several ways. One option is to delay selected flights. Other possibilities are to reallocate the resources needed to operate flights (e.g. aircraft, crews, but also airport arrival slots), or even to cancel flights to reduce the demand on those resources. In addition, airlines may sometimes create flights to balance the schedule.

It should be noted that balancing the schedule may be interpreted differently by individual airlines: For one airline the objective might be the ability to return to the normal schedule by the next day, while for another it might mean flying as many of its scheduled flights as possible (cf. [24]).

- *Flight Planning and Dispatch.* An important aspect of airline operations is to determine flight routes and payload that minimize costs and meet the overall airline flight objectives. Winds, aircraft type and restrictions all affect the choice of route, which involves a complex trade-off between speed, altitude, payload and fuel load. In addition, flight planning may have to take into account that regions of airspace may be congested or temporarily inaccessible.

- *Flight Monitoring.* This includes monitoring all aspects of flights in progress, such as ensuring that the flight stays within safe and legal limits, assessing weather conditions en route and at destination and alternate airports, and assisting crews in solving problems that may arise. Thus, AOCs are in constant communication with crews during flights.

Schedule planning is usually performed by dedicated coordinators. Flight planning, dispatch, and monitoring are performed by flight dispatchers, which are licensed personnel responsible for individual flights. By law, the responsibility for the safety and control of flights is shared between the dispatcher and pilot; thus, dispatchers at the AOC maintain frequent contact with pilots prior to and during the flight. Other tasks of airline operational control include crew scheduling and tracking, aircraft maintenance operations, and gate management. Typically, these tasks are performed by separate departments that interact with the AOC. For instance, airport stations manage gate allocations and other ground-based resources (e.g. passenger and baggage handling); maintenance stations handle the coordination of required aircraft maintenance checks (e.g. ensure that aircraft are routed through the maintenance stations).

2.2.2 Interaction

Both on the side of the FAA and on the side of the airlines, decision-making responsibilities are shared between a number of stake holders. The actions these stake holders may perform are of course highly interdependent, and therefore necessitate a significant degree of coordination. On the FAA's side, operational processes are essentially distributed among three organizational levels. At the first level, we find the ATCSCC. The ATCSCC oversees aggregate traffic flows and monitors current and projected capacity limits and demands. Major flow management actions, such as GDPs and SWAPs, are usually initiated by the ATCSCC. The ATCSCC coordinates these ATFM initiatives with traffic management units at the various ARTCCs, TRACONs, and Towers, which form the second organizational level. The entities at this level are responsible for coordinating air traffic in their assigned regions of the airspace. Besides their interaction with the ATCSCC, adjacent centers at this level also interact to coordinate the air traffic between their regions. ARTCCs, TRACONs, and Towers further

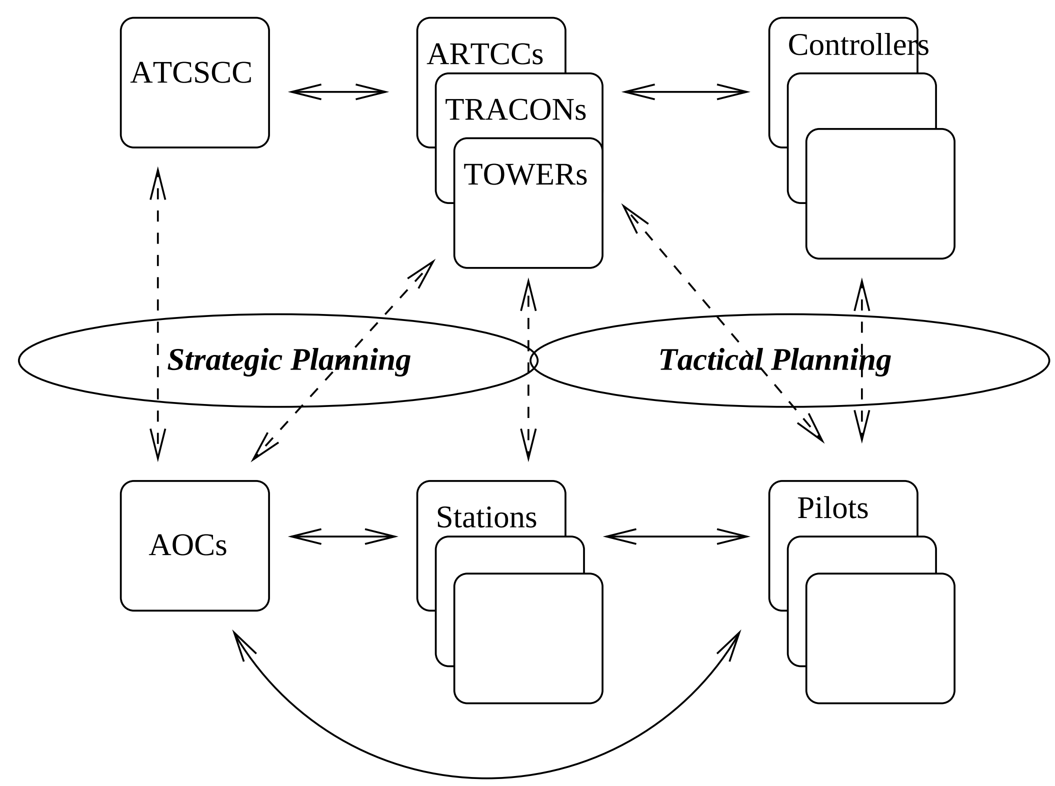


Figure 2.2: Operations Level Interactions between FAA and airlines

delegate responsibilities to the individual air traffic controllers, which form the third organizational level. The primary interaction at this level is between controllers at adjacent sectors to transfer control of aircraft. On the airlines' side, on the other hand, daily operations are primarily coordinated at centralized AOCs. Specific tasks, such as gate assignments and maintenance are coordinated with various stations, and flight dispatch is of course in constant communication with pilots to monitor and control the progress of individual flights.

The interaction between the FAA and airlines during daily ATFM operations may be separated according to interactions at the strategic and the tactical level, as is shown in Figure 2.2. At the strategic level, interactions occur primarily between between the ATCSCC and the airlines' AOCs. When the ATCSCC predicts a sustained period of congestion, it may respond with an ATFM initiative (e.g. a GDP or a SWAP plan), which is communicated to the airlines' AOCs. Typically, these plans are formulated two to four hours in advance. In turn, airlines communicate the schedule adjustments they intend to make in response to these disruptions. It should be noted that this flow of information is highly important, as the FAA's decisions are partly based on the information they receive. At the tactical level, the interactions occur primarily between controller, pilots (to ensure separation), centers and stations. Typically, these interactions concern ATC initiatives (e.g. ensuring the separation standards), and other near real-time initiatives.

2.3 Discussion

The current structure of ATM in the U.S., with its amalgam of flow management initiatives and variety of stake holders, is the result of an evolutionary process stretched across a number of decades. When faced with a frequently recurring problem, the typical response would be a "local" solution, with limited concern or consideration to the overall system effects. While air space capacity was readily available, the congestion resulting from disruptions to the system (e.g., bad weather) was relatively minor. Traditionally, the FAA would unilaterally decide how to resolve this congestion, with little or no input from the airlines. On the whole, airlines could reasonably absorb the resulting flight delays with limited effects on the integrity of their flight schedules (e.g., by accounting for these effects in the flights' block times [25]).

However, the steady growth in air traffic during the past decade has pushed this approach to its limits. As demand levels approximate available capacity, even minor disruptions may have significant ripple effects and lead to sustained periods of congestion. This became painfully clear in September 2000, when the relaxation of slot controls mandated by Congress led to a daily reoccurrence of gridlock at LaGuardia. The resolution of these disruptions necessitates an increased role for the ATFM initiatives imposed by the ATSCC, in particular with regard to the possible network effects in the system. At the same time, the impact of these effects on the integrity of flight schedules has significantly increased the management responsibilities at the airline side.

As a result, the coordination and cooperation between the FAA and the airlines has become increasingly important. To implement appropriate ATFM actions, the FAA needs an accurate picture of flight status and intent. Airlines, on the other hand, need the flexibility to adjust their schedules, and can only provide accurate information if they know the actions planned by the FAA. Given the relatively short response times, the real-time exchange of information between the FAA and the airlines is therefore a critical component of ATFM functionality. In addition, it has become increasingly clear that the ATCSCC should not be solely responsible for determining the delays, reroutes, etc. required to resolve congestion. While both the FAA and airlines can possibly delay or reroute flights, certain actions that may alleviate congestion are only available to airlines. For example, only an airline can decide to cancel flights or to reassign passengers, crew, and aircraft. Consequently, any successful attempt at flow management will require a significant input from and role for airline decision-making. Such decisions involve economic trade-offs, which the

FAA is not in a position to make. As such, it is not surprising that current efforts to improve ATM, which are discussed in the next Chapter, envision a more decentralized system for managing air traffic.

Chapter 3

Towards Decentralized Air Traffic Management

Recent studies estimate that air traffic will increase at an annual rate of 3% to 5% over the next 15 years. Accommodating this increase in air traffic will likely require significant changes in the structure of ATM functions, especially in light of the already reoccurring periods of gridlock in the system. The FAA has responded to this challenge by formulating a comprehensive vision for the future of ATM, better known as *Free Flight*. In addition to extensive technology upgrades, the notion of free flight is characterized by a significant move toward decentralized decision-making.

This chapter presents an overview of Free Flight and the related concept of Collaborative Decision-Making (CDM). In addition, we discuss the effect these ideas have had on the implementation of GDPs. It should be noted that a move toward decentralization in such a complex environment may bring forth a variety of issues, such as human factors problems, software development, etc.. However, this chapter focuses on the issues related to resource allocation problems that arise in the implementation of these ideas.

3.1 The Future of Air Traffic Management

The current ATM structure presents a myriad of rules and procedures for airspace users. Notwithstanding recent initiatives, users are often forced to operate within narrow and highly restricted guidelines. While this approach provides a high level of predictability (and therefore safety), it is safe to say that the structure of the airspace system was simply not designed to deal with the current and projected volume of traffic. As a result, the FAA has been subject to widespread criticism. In particular, there is a general consensus among airlines that the restrictions implemented by the FAA are often overly severe, which results in unnecessary delays, congestion, and costs for the airlines. In response to these criticisms, the FAA has formulated a wide-ranging set of plans known as "free flight". The first phase of the implementation is currently underway, and started in 1997 (see [55]).

3.1.1 Free Flight

According to the FAA, Free Flight represents

> "a concept for safe and efficient flight operating capability under instrument flight rules (IFR) in which the operators have the freedom to select their path and speed in real-time. Air traffic restrictions are imposed only to ensure separation, to preclude exceeding airport capacity, to prevent unauthorized flight through special use airspace, and to ensure the safety of flight. Restrictions are limited in extent and duration to correct the identified problem. Any activity that removes restrictions represents a move toward free flight."

The concept of free flight embodies a different philosophy toward ATM functions. The traditional approach largely followed a central planning paradigm, in which users had to adhere to ATC decisions (e.g. using ATC-preferred routes). In contrast, free flight envisions increased collaboration between users and air traffic managers, greater flexibility for airlines to make decisions to meet their unique operational goals, and the replacement of broad restrictions with more tailored responses. In theory, free flight would let pilots assume a significant portion of the separation responsibilities, and choose routes as they see fit using advanced technologies. ATC interventions would only occur if flight separation standards were threatened to be violated.

There are, however, a number of steps that need to be taken before these ideas can be put into practice. As a first step toward free flight, the FAA has instituted the National Route Program (NRP), which gives airlines and pilots greater liberties in choosing their routes. Under this program, certain flights may proceed unrestricted from origin to destination[1]. The NRP program has had considerable success (see [1] and [55]), showing the potential benefits of free flight. Other efforts currently underway focus on the necessary technology improvements, such as digital communication systems and satellite-based navigation technology.

In the previous chapter, we separated ATM functions according to two basic components, tactical ATC and more strategic ATFM. It should be noted that with its focus on separation insurance and dynamic conflict probing and resolution, free flight is perhaps best viewed as the future vision for the ATC functions

[1]subject to terminal area restrictions within a 200-mile radius of take-off and landing, as well as certain altitude restrictions

in the air transportation system.

3.1.2 Collaborative Decision-Making

Collaborative Decision-Making (CDM) is a concept that goes hand in hand with free flight, in that it may be viewed as the future direction of ATFM functions (for an overview of CDM, see [8], [10], and [83]). Under CDM, the management of traffic flows and the associated resource allocation decisions are conducted in a way that gives significant decision-making responsibilities to AOCs. The overall objectives of CDM can be summarized as:

- generating better information, by merging flight data from the airspace system with information generated by airspace users;
- creating common situational awareness by distributing the same information both to traffic managers and to airspace users;
- creating tools and procedures that allow airspace users to respond directly to congestion and to collaborate with traffic flow managers in the formulation of flow management actions.

CDM was initially conceived in the mid-1990s within the FAA Airline Data Exchange (FADE) project, which originally was created as a short-term experiment to see if up-to-date airline schedule information would result in improved flow management decisions. The issues revealed during extensive human-in-the-loop experiments eventually led to the initial implementation of CDM, which primarily focused on the development of new operational procedures and decision support tools for implementing and managing GDPs.

The initial implementation of CDM, known as GDP enhancements (GDP-E) began its prototype operations at San Francisco and Newark airports in January of 1998. In GDPs under CDM, airlines send operational schedules and changes to schedules to the ATCSCC on a continual basis. The schedule information includes flight delay information, cancellations, and newly created flights. The ATCSCC uses this information to monitor and possibly implement GDPs, using a newly developed decision support tool called Flight Schedule Monitor (FSM). It is important to note that this information is shared with all users (e.g. airlines also have access to FSM), creating a common picture of current and projected airport conditions. Essential to these procedures is the use of newly defined resource allocation procedures, which have removed previously existing disincentives for airlines to provide accurate information. The effects of these procedures has been significant: it has been stated that since their initial implementation in January of 1998, over six million minutes of assigned ground delay have been avoided (cf. [9]). While one can point to a variety of concepts and technologies that are fundamental to CDM's success, probably the most vital underlying element has been a strong and continuous interaction among all stake holders. Airline input was sought from the very beginning, and regular meetings between the various groups involved in CDM have been held through the life of the CDM project.

The success of these initial CDM efforts has highlighted the potential benefits of increased collaboration in ATFM, and led to a number of projects that aim to enhance the basic application of CDM to GDPs. Examples include the incorporation of uncertainty trade-offs during a GDP (e.g., due to weather predictions, see [28]) and the possible inclusion of airport departures into the GDP planning process ([25]). Another example is the current Collaborative Routing (CR) effort, which intends to improve handling of potential en-route congestion; whereas GDPs under CDM give airlines more flexibility in distributing FAA-assigned delays among its flights, CR would also give airlines greater input in rerouting flights.

3.2 Decentralized Ground Delay Programs

So far, the efforts of the CDM working group have primarily concentrated on GDP enhancements. These efforts have led to substantial changes in the procedures for allocating ground delays, which provide airlines a much greater input. As such, these procedures present a significant move towards decentralized ATFM. This section introduces the main GDP procedures introduced under CDM, and contrasts these procedures with traditional decision models for the allocation for slots.

3.2.1 Models for the Ground Holding Problem

The use of ground holding to resolve air traffic congestion was first described systematically by Odoni [56]. However, the generic flow management problem defined by Odoni is extremely general, in that it addresses congestion anywhere in the network. Therefore, a common assumption (both in theory and, more implicitly, in practice) is that the only capacitated element in the air traffic network is the arrival airport. Under this assumption, the problem is commonly known as the Ground Holding Problem (GHP). The basic version of the GHP (see [72]) requires the following additional assumptions:

I. *Discrete Time Horizon*: The planning horizon consists of a fixed and finite time period, which has been discretized into contiguous time periods (slots).

II. *Deterministic Demand*: At the beginning of the planning horizon, a complete list of flights bound to arrive at the congested airport is known. Moreover, the travel times of these flights are deterministic and known in advance.

III. *Deterministic Capacity*: At each time period, the airport arrival capacity in each time period is deterministic and known in advance (Without loss of generality, we assume each slot can service 1 flight).

Given these assumptions, the GHP can be formulated as an Integer Programming problem. We represent the flights as a set $\mathcal{F}$ and the slots as a set $\mathcal{S}$. We let oag_f denote the scheduled arrival time of a flight $f \in \mathcal{F}$, and t_s the time of a slot $s \in \mathcal{S}$. The resulting LP formulation is shown in Figure 3.1. It should be noted that the constant capacity assumption implies that no flight will be allocated airborne delay (since airborne delay is more expensive than ground delay). Thus, this version of the GHP allocates ground delays based on the costs $C_f(d)$, which are a function of delay.

While the distribution of delays among flights is an important topic, it has received relatively little attention in literature. Most models that address the GHP assume constant marginal costs of both airborne and ground delay [2], and instead concentrate on the trade-off between them in the case of stochastic ca-

[2]Note that in this case, the previous problem can be simplified further, since a first-come, first-served ordering will be optimal.

Decision variables:

- $x_{fs} \in \{0, 1\}$, for all $f \in \mathcal{F}$, $s \in \mathcal{S}$, $t_s \geq a_f$.

LP formulation:

$$
\begin{array}{lll}
\text{Min} & \sum_{f \in \mathcal{F}, s \in \mathcal{S}, t_s \geq a_f} C_f(t_s - oag_f) x_{fs} & \\
\text{subject to:} & & \\
& \sum_{s \in \mathcal{S}, t_s \geq oag_f} x_{fs} = 1 & \text{for all } f \in \mathcal{F} \\
& \sum_{i \in \mathcal{F}, t_s \geq oag_f} x_{fs} \leq 1 & \text{for all } s \in \mathcal{S} \\
& x_{fs} \geq 0 &
\end{array}
$$

Figure 3.1: Assignment problem formulation of the static, deterministic GHP

pacity (i.e., by relaxing assumption 3). This version of the GHP was first studied in [5] and [56]. More efficient models, as well as several extensions, were proposed in [63], [64], [73], and more recently [11] and [27]. A systematic review of some of these results may be found in [4]. Other related work has focused on different aspects of the ground holding problem, in particular on the effects of delay propagation through the air traffic network (e.g., [3] [82], and citeVranas94a) and on more general air traffic flow management problems (see [14] and [15]).

Generally speaking, one might argue that the focus on aggregate trade-offs between airborne and ground delays limits the attention that can be given to airline-specific preferences. Even though airline specific delay costs could, in principle, be incorporated into the decision problems, the "global optimization" perspective would likely introduce systematic biases against or in favor of individual airlines[3]. Consequently, the models described here are perhaps primarily suited for making aggregate decisions (e.g., determining overall flow rates per period, as discussed in [11]).

3.2.2 Airline Decision-Making during GDPs

Whereas the FAA is primarily concerned with aggregate traffic flows and overall throughput during periods of airport congestion, the decisions and trade-offs faced by individual airlines in a GDP are of a different nature. When faced with a GDP, an airline typically responds to the resulting schedule disruptions by trading-off flight cancellations and delays. Such decisions are based on a multitude of factors, such as the disruption of and the cost of crew schedules, the passenger costs of delay, possible flight connections, etc.

The ground delays imposed by a GDP create severe disruptions of an airline's flight schedule, which not only affect the delayed flights but may also propagate delays to other flights. To mitigate these disruptions, airlines may cancel flights and substitute flight-slot assignments. A decision model to support this slot swapping process was first presented in [79], which also describes its application at American Airlines. Other models for resolving schedule disruption through slot swapping are proposed in [29], [40], [41], and in [54]. Another approach, which explicitly considers the connection dependencies of hub operations but leads to less efficient algorithms, is proposed in [43] and further extended in [22]. It should be emphasized that none of these models fully reflect the complexity

[3]See [56]. Typically, most ground-holds would be assigned to aircraft with smaller per-unit delay costs (e.g. regional aircraft), while aircraft with higher delay costs would be given priority (e.g. wide-body aircraft).

of airline decision-making during GDPs. For instance, none of these models incorporates the decision to cancel a flights, which is one of the most important decisions during a GDP.

Another family of decision models for resolving schedule disruptions may be found in [12], [20], [21], [70], [71], [74], and [75]. Generally speaking, this class of models attempts to find an operable, system-balanced flight schedule when aircraft shortages disrupt an airline's flight schedule (that is, they consider an airline's entire network of flights). The application of these models at United Airlines is described in [35] and [60]. These models, however, typically do not incorporate arrival slot constraints. Their use is primarily in schedule recovery after the disruptions from a GDP have occurred.

3.2.3 Ground Delay Programs under CDM

In contrast to the models proposed in the literature, the allocation procedures instituted under CDM primarily address the distribution of delays among individual flights. CDM has its origin in early efforts by the FAA (through the FADE program[4]) to acquire up-to-date airline schedule information (cf. [83]). Though human-in-the-loop experiments with ATFM specialists clearly showed the benefits of this information, airlines remained highly reluctant to submit this data. The reason for this was that the GDP procedures used at the time could actually penalize an airline for providing that information. The main reason for this problem was that flights were essentially allocated slots on a first-come, first-served basis according to their most recent estimated arrival times, using a algorithm called "Grover-Jack". With this mechanism, a so-called "double

[4]FADE: FAA Airline Data Exchange

penalty" could occur. For example, if a flight were delayed for 30 minutes due to mechanical problems and the airline reported this delay during a GDP, the FAA might assign another 30 minutes of ground delay (for a total delay of 60 minutes). However, the GDP-assigned delay would have been absorbed if the airline had not reported its mechanical delay! In addition, the Grover-Jack mechanism reallocated slots that were assigned to cancelled flights, thus preventing an airline from substituting other flights into those slots.

The resource allocation schemes implemented under CDM have addressed these issues through a fundamental change in the allocation of capacity. Rather than an assignment of individual flights to arrival slots, the central paradigm under CDM is that slots are allocated to airlines. This has led to the introduction of two new allocation mechanisms, RBS and Compression. The RBS algorithm creates an initial allocation of slots to airlines, based on the consensus recognition that airlines have claims on the available arrival capacity through the original flight schedules. Given their slots, airlines are free to reschedule flights according to their private objectives, through flight substitutions and cancellations. The Compression Algorithm is a reallocation procedure that prevents the underutilization that might be caused by flight cancellations and delays.

Ration-By-Schedule As the first step in a GDP, the RBS procedure rations the arrival slots among airlines. As in the Grover-Jack procedure, RBS assigns flights to slots on a first-come, first-served basis. In RBS, however, flights are ordered according to their *original scheduled time of arrival* as opposed to the *most recent estimated time of arrival* that was used before. Consequently airlines will not forfeit a slot by reporting a delay or a cancellation, which is what

RBS Algorithm:

Step 1. Order the flights in $\mathcal{F}$ by increasing scheduled time of arrival.

Go to step 2.

Step 2. Select the first flight in $\mathcal{F}$ that has not been assigned a slot.

If no such flight exists, the algorithm is terminated.

Otherwise, the flight is assigned the earliest unassigned slot it can meet.

Figure 3.2: The Ration-By-Schedule Procedure

happened prior to CDM. Figure 3.2 provides a conceptual overview of the RBS algorithm.The actual RBS algorithm has to take into account several complicating factors, such as flights being airborne, flights exempted from the GDP, and the possibility that a GDP was already executed before. (see [28] for a discussion of these details).

It should be noted that the resulting flight schedule may be inefficient in its utilization of arrival capacity. Arrival slots may have been assigned to flights that have been cancelled or delayed and therefore cannot use their assigned slot. However, the end result of RBS should not be viewed as an assignment of slots to flights but rather as an assignment of slots to airlines. After RBS, an airline can reassign the slots it owns to any of its flights by using the cancellation and substitution process.

Compression After a round of substitutions and cancellations the utilization of slots can usually be improved. The reason for this is that an airline's flight cancellations and delays may create "holes" in the current schedule, that is, there will be arrival slots which have no flights assigned to them. The purpose of the

Compression Algorithm is to move flights up in the schedule to fill these slots. The basic idea behind the Compression Algorithm is that airlines are "paid back" for the slots they release, so as to encourage airlines to report cancellations. The extent to which a flight can be moved up will be limited, e.g. a flight cannot depart before its scheduled departure time. To capture this, each flight has an *earliest arrival time* specified by the mapping $e : \mathcal{F} \to \mathcal{S}$. Moreover, it is assumed that a flight cannot be moved down from its position in the current schedule I. Thus, the set of slots $\{e(f), \ldots, I(f)\}$ defines the window in which flight f can land. A conceptual overview of the Compression Algorithm is shown in Figure 3.3. It should be noted that there are two ways for an arrival slot to become open; either the flight assigned to that slot has been cancelled, or it has been delayed by a cause unrelated to the GDP, e.g. mechanical delay. In either event, the controlling airline will release the slot to the Compression Algorithm.

The important features of the Compression Algorithm are that (i) arrival slots are filled whenever possible, (ii) flights from the airline that owns the current open slot are considered before all others, (iii) if the controlling airline cannot use a slot, then it is compensated by receiving control over the slot vacated by the flight which moves into its slot, and (iv) airlines do not involuntarily lose slots they own and can use.

To illustrate the Compression Algorithm, let us consider the example shown in Figure 3.4. The leftmost figure represents the flight-slot assignment prior to the execution of the Compression Algorithm. Associated with each flight is an earliest time of arrival, and each slot has an associated slot time. Note that there is one canceled flight. The rightmost figure shows the flight schedule after execution of the Compression Algorithm: as a first step, the algorithm attempts

Compression Algorithm:

Step 1. Order the flights according to the current schedule. Determine the set of open slots $\mathcal{C}_{\mathcal{S}}$. For each slot $c \in \mathcal{C}_{\mathcal{S}}$, execute step 2.

Step 2. Determine the owner of slot c, that is, the airline a that owns the cancelled or delayed flight f that has been assigned to slot c. Try to fill slot c, according to the following rules:

2.1. Determine the first flight g from airline a (in the current schedule) that can be assigned to slot c, that is, for which $c \in \{e(g), \ldots, I(g)\}$. If there is no such flight, go to Step 2.2. Otherwise, go to Step 3.

2.2. Determine the first flight g from any other airline that can be assigned to slot c. If there is no such flight, go to Step 2.3. Otherwise, go to Step 3.

2.3. There is no flight that can be assigned to slot c. Return to Step 1 and select the next open slot.

Step 3. Swap the slot assignments of flights f and g, i.e, assign flight g to slot i, and flight f to slot $I(g)$. Note that airline a is now the owner of open slot $I(g)$. Next, slot $I(g)$ is made the current slot, and Step 2 is repeated.

Figure 3.3: The Compression Procedure

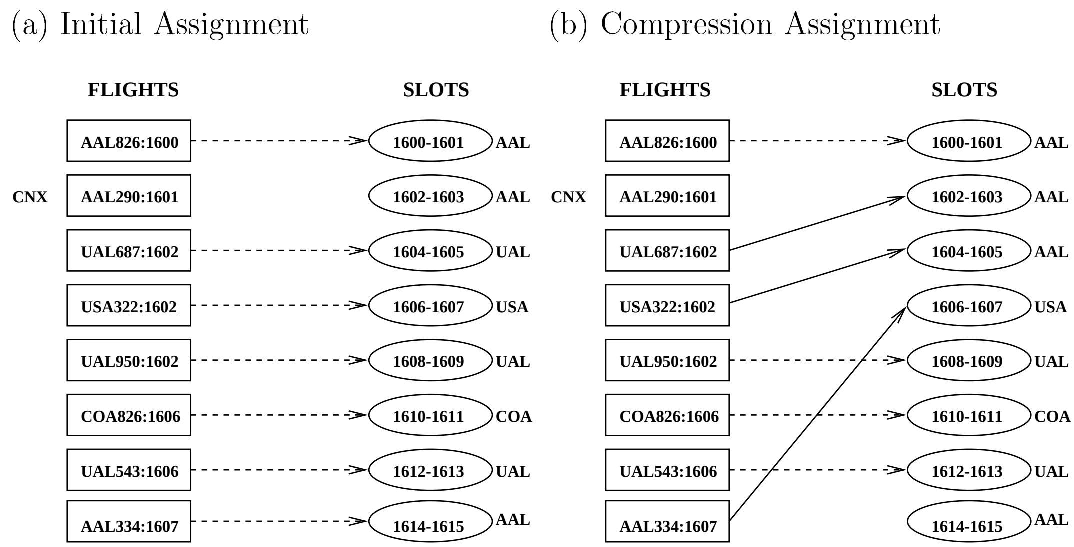


Figure 3.4: Compression Example.

to fill AAL's open slot. Since, there is no flight from AAL that can use the slot, the slot is allocated to UAL, and the process is repeated with the next open slot.

3.3 Discussion

The move toward free flight presents a number of dramatic changes in ATM functions. It should be emphasized, however, that the path toward implementing these changes is necessarily incremental, consisting of a large number of small steps. Radical, large-scale modernization efforts by the FAA have had a history of failure, and the FAA has therefore adopted the more cautious approach of "build a little, test a little, field a little". This approach is further motivated by the vast complexity of the airspace system, in which many of the constraints and rules are difficult to represent formally. The current free flight efforts and the initial implementation of CDM are examples of this approach, and for which the restricted focus has led to considerable success.

The GDP enhancements introduced under CDM, for instance, have had a profound impact on the interaction between the FAA and airlines. From these procedures, a general protocol for the interaction between the FAA and airlines has begun to emerge, which defines each side's roles and responsibilities. In particular, CDM has solidified the FAA's responsibility to monitor the system and its authority to ensure that demand does not exceed capacity. Furthermore, the procedures implemented under CDM define the role of the FAA as a discoverer of constraints and as an arbiter of rationed capacity. The resulting allotments of scarce capacity allow airlines to trade off operating options based on internal business objectives. At the same time, airlines are responsible for providing the FAA with accurate data, especially in light of their possible schedule adjustments (see also [84]).

The evolution of this framework has been, at least to a large degree, the result of the environment in which it is used. In fact, one might argue that these procedures have been successful exactly because they took into account the context in which GDPs are executed (and, through extensive meetings, the concerns of all the parties involved). In particular, a key characteristic of GDPs under CDM is that slots are allocated initially, using RBS, followed by periodic inter-airline slot exchanges, using Compression. The use of slot exchange (as opposed to a single allocation step) can be interpreted as a compromise between two opposing factors. On the one hand, GDPs are executed in a dynamic environment that is characterized by significant uncertainty. In and of itself, this presence of uncertainty would suggest a postponement in the allocation of slots (i.e. the allocation of flights to slots would be postponed as much as possible). On the other hand, however, airline operations would be severely hampered by such a postponement in the allocation of arrival slots. The impact of GDPs on flight schedules forces airlines to respond with an oftentimes elaborate set of strategies [5]. The time required to formulate a response (as well as the internal communication of schedule changes) may be significant, and suggests an early allocation of arrival slots. Specifically, delays in the allocation would severely limit an airline's possible responses and would expose them to the risk of not getting timely slots. The initial allocation followed by an exchange addresses both these issues: the initial allocation allows airlines ample time to formulate a response, while subsequent reallocations address dynamic changes in schedule and capacity.

Another factor that has contributed to the success of CDM is the recognition that GDPs are part of a much larger and more complex air traffic system. By limiting the scope of its decisions, the GDP processes take into account that determining airport arrival capacities is just one of the decisions made by the the ATCSCC, and that arrival slot allocation is part of a much more complex set of decision processes faced by the airlines. This is reflected in the following two features of the GDP process. First, the direct interaction in GDPs under CDM primarily involves the ATCSCC and airline AOCs, and does not incorporate direct communications with, or input from, other stake holders (e.g. Pilots, Tracons, etc.). This does not imply that the concerns from other stake holders are not taken into account; rather, these intra-FAA and intra-airline decisions are explicitly incorporated into the scope of GDP procedures. Second, GDPs can be implemented with little knowledge or information about airline preferences.

[5] To maintain the schedule's integrity, an airline may cancel or delay flights, and reallocate resources (e.g. crews, aircraft, etc.).

The reason for this is that after the initial allocation, airlines can adjust their schedules using the substitution/cancellation process. For instance, suppose that flight AL100 is a lightly loaded flight with few connecting passengers and a CTA of 12:00 and that AL500 is a fully loaded flight with many connecting passengers and a CTA of 12:45. Since the timely arrival of the first flight is not that crucial, airline AL might want to cancel AL100 and substitute flight AL500 into the 12:00 time slot, thus saving AL500 45 minutes of delay. It should be emphasized that an approach that might require substantial a-priori revelation of airline preferences could be difficult to implement; airlines may not yet have completed formulating their response strategies, and it may be difficult for an airline to evaluate all possibilities. Not only do an airline's decisions involve an amalgam of factors, they also may difficult to quantify (e.g., when taking into account factors such as workload distribution). Moreover, different airlines may have different planning capabilities. To summarize, the GDP processes implemented under CDM provide a set of clearly defined roles and responsibilities, with compact interactions yet substantial flexibility, which can be implemented on a relative stand-alone basis.

It should be noted that the allocation procedures developed under CDM are markedly different from the approaches that have so far been proposed in the literature (see [2], [25], [43]). In [2], an evolutionary framework is proposed in which airlines coordinate their decisions through bargaining and auctions. These ideas are further pursued by Hall ([25]) and by Milner ([43]), who proposes a procedure for allocating slots during GDPs that is closely related to the well-known Vickrey auction ([80]). To illustrate their differences from the procedures implemented under CDM, it is instructive to classify them according to the general framework proposed by Moulin ([52], Ch. 1), who recognizes three fundamental "modes" of cooperation - direct agreements, decentralization (by prices), and justice. Under the mode of *direct agreements*, the coordinator has no active role. Instead, agents are allowed to engage freely in direct transactions as they see fit. Under the mode of *decentralization*, the coordinator's role is to enforce certain rules of interaction (either explicitly or implicitly, through the "invisible hand"), The prime example is the model of competitive markets, in which agents coordinate by responding to price signals. In the *justice* mode, the coordinator takes a more active role: resources are allocated according to a mechanical formula that distributes the resources equitably among the agents. Based on this classification, the models proposed in [2], [25], and [43] correspond to the mode of direct agreements and the mode of decentralization. Interestingly enough, however, the procedures implemented under CDM fall under the mode of distributive justice. As illustrated in our previous discussion, there are a number of reasons why the use of procedures based on concepts of distributive justice (i.e. notions of equity and fairness) may be more applicable within the context of GDPs.

While the resource allocation schemes developed under CDM are still evolving, this general interaction protocol - and in particular the notion of resource rationing - is often viewed as a blueprint for decentralized decision-making within ATFM, and seen as the basis for all further efforts. Given the success of CDM and its acceptance within the airline community, this appears to be a natural development. As the reach of these efforts expands, however, the need for a set of guiding principles with regard to fairness and equity is becoming more and more apparent. The allocation schemes implemented under CDM have been developed to address certain specific problems the airlines and the FAA were dealing with. As a result, many of the overall fairness concepts that were introduced have largely been left implicit. That is, there has not been a clear distinction between the algorithms and the principles. Clearly, this complicates further enhancements, and introduces the danger of creating another over-constrained system with a myriad of rules and restrictions.

Chapter 4

Fair Slot Allocation

A primary objective of the FAA's ATM functions is to provide fair and equitable access to the National Air Space[1]. Traditionally, the FAA has interpreted fairness as prioritizing flights on a "first-come, first-served" basis. The allocation procedures introduced under CDM, however, represent a departure from this paradigm: airlines receive allotments of slots based on their original flight schedules. Yet in spite of these significant changes, it is often not clear what is meant by a fair or equitable allocation within the context of GDPs under CDM. The embodiment of fairness under CDM is largely left implicit in the allocation procedures (Ration-By-Schedule and Compression), and in fact, different and even conflicting concepts are sometimes used to describe these procedures. This not only generates frequent complaints, but also complicates the introduction of CDM in more complex settings. This chapter therefore aims to formalize concepts of fairness for the allocation of slots during GDPs. Using an axiomatic approach we derive a class of potential allocation procedures, which introduce a number of potential alternatives to RBS. We discuss the interpretation of these

[1] see http://www.faa.gov/atpubs (order 7110.65)

alternatives, and empirically analyze their difference with RBS.

4.1 Introduction

The allocation procedures instituted under CDM have created a connection between planned schedules and operational schedules that did not previously exist: namely, airlines are entitled to a share of the operational resources based on their planned schedules. In fact, the IATA scheduling guidelines (which are used to create planned schedules at biannual conferences, see [32]) explicitly state that

> "The Conferences deal with adjustments to planned schedules to fit in with the slots available at airports. This activity has nothing to do with adjustments to schedules on the day of operation for air traffic flow management. The two types of slot allocation are quite different and unrelated."

Moreover, at the four airports (Kennedy, LaGuardia, O'Hare, Reagan National) that fall under the High-Density rule, the slots owned by airlines are often interpreted as "the right to schedule or advertise a flight at a specific time" (see [19], [65]), which entails no explicit connection to a right on the day of operation. As such, the Ration-By-Schedule (RBS) procedure introduced under CDM has implicitly introduced a significant change to ATM practices.

The RBS algorithm is based on the notion of "first-scheduled, first-served", and iteratively assigns the next arrival slot to one of the remaining flights with the earliest scheduled time of arrival (or equivalently, its delay up to that point). Thus, slots are assigned to flights according to a priority ordering based on their respective scheduled times of arrival. While intuitively appealing, the use of this paradigm poses a number of questions. First, it is by no means clear that RBS is the only or even the most desirable possible allocation schemes. An important issue is therefore whether other standards of comparison may be applicable within the context of GDPs, and how they compare with RBS. The appropriateness of using of flights, as opposed to airlines, to compare possible allocations poses another question. While RBS allocates slots on a flight-basis, equity is measured ex-post (for analysis purposes) on an airline basis. Moreover, the use of flights as a basis for allocating delays can further be questioned by the existence of the subsequent substitution process, which allows an airline to redistribute its assigned delays in any way it sees fit. Consequently, all the allocation procedure can possibly achieve is an allocation of the slots or delays among airlines. Therefore, another issue is whether airline-based allocation procedures could be more applicable.

4.1.1 Model Description

Whenever the FAA implements a GDP, air traffic managers first have to determine the affected flights and the available arrival capacity. In our model we assume these are given, that is, we let $\mathcal{F} = \{f_0, \ldots, f_{m-1}\}$ represent the flights in the GDP, and $\mathcal{S} = \{s_0, \ldots, s_n\}$ the slots available during the GDP. Each slot $s_j (0 \leq j < n)$ has a capacity $c_j \in \{0, 1\}$, and we assume that the capacity c_n of slot s_n is unbounded. In addition, each slot s_j has an associated slot time t_j, and we assume that the slots are equally spaced. More specifically, $t_{j+1} - t_j = 1$ for all $0 \leq j < n$, and $t_0 = 0$. Each flight f_i has an associated originally scheduled time of arrival $oag_i \in 0, \ldots, n-1$ corresponding to one of the slots. A flight's originally scheduled time of arrival represents the earliest time it could possibly land. In addition, we represent the airlines involved in the GDP by a set $\mathcal{A}$. For each airline $a \in \mathcal{A}$, $\mathcal{F}_a \subseteq \mathcal{F}$ represents the flights operated by airline a.

4.2 Delay-Based Slot Allocation

The principal output of the RBS procedure is a controlled time of departure for each flight in the GDP; based on its assigned slot each flight is assigned a certain amount of ground delay. In this section, we consider approaches to fair slot allocation that are based on comparisons of the delay incurred by flights and/or airlines. We discuss the fundamental principles underlying these methods, and analyze key properties of the resulting allocation schemes.

4.2.1 Multi-Objective Optimization Methods

In the Operations Research literature, the equitable allocation of limited resources is commonly approached by using multi-objective or goal programming methods ([33]). While a number of allocation criteria are possible, a popular concept of equity for these problems is the so-called lexicographic minimax criterion (see [42] for an overview of its use in a large number of resource allocation problems). Given a number of resources, a set of activities, and a set of performance functions associated with each activity, the lexicographic minimax solution states that the resources are allocated equitably among the activities if no performance function value can be improved without degrading an already equal or worse-off performance function value (cf. [42]). This minimax or difference principle has its origins in the general theory of social justice proposed by Rawls ([62]), in which the central distributive principle states that the least well-off group in society should be made as well off as possible (see also [66]).

The lexicographic minimax criterion provides a natural interpretation of the notion of fairness embedded in the RBS procedure. This is shown in the following Theorem.

Theorem 4.2.1. *The flight-slot assignment obtained by the RBS algorithm lexicographically minimizes the maximum delay with respect to the original flight schedule; that is, let x represent a flight-slot assignment, T represent the maximum delay under RBS, and define for each k, $0 \leq k \leq D_{max}$, the performance function $d_k = \mid \{(i,j) \in x : t_j - oag_i = k\} \mid$. Then, the allocation obtained by RBS lexicographically minimizes the vector $d = (d_T, \ldots, d_0)$ over all possible flight-slot allocations.*

Proof. We assume w.l.o.g. that all oag times are different. The proof follows by a sequential exchange argument. Let A_1 be a lexicographical min-max assignment and A_2 an assignment generated by RBS. We now will argue that A_1 and A_2 necessarily assign the same flight to the first slot. Suppose this is not the case so that flight f occupies the first slot, s_1, in A_2, but slot $s_k > s_1$ in A_1, and let g be the flight assigned to s_1 in A_1. It follows from the basic properties of RBS that $oag_f < oag_g$, which implies $\mathrm{Max}\{s_1 - oag_f, s_k - oag_g\} < \mathrm{Max}\{s_1 - oag_g, s_k - oag_f\}$. It then follows that the lexicographical min-max objective function can be improved for $A1$ by interchanging f and g. This is a contradiction to the optimality of $A1$. Repeating this argument for slots $2, \ldots, n$ yields the desired result. □

In other words, the allocation obtained by RBS is such that any flight's allocated delay d cannot be reduced without increasing the delay of another flight to a value of least d. Thus, each flight is allocated a delay that is "as close as possible" to the average delay. Under this interpretation, each flight in the GDP is therefore entitled (or rather, responsible for) an equal share of the resulting overall delay.

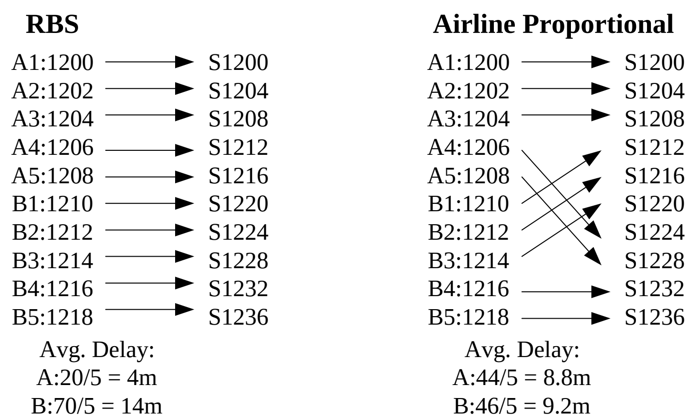


Figure 4.1: Example: Airline-based Delay Allocation

Based on this principle, one could easily envision airline-based allocation methods which would lexicographically minimize the maximum airline delay or the average airline delay. If we were to minimize the maximum airline delay, the implicit assumption would be that each airline is responsible for an equal share of the delay. If, on the other hand, we were to minimize the maximum average airline delay the principle would entail that each airline is responsible for a proportional share of the overall delay. The potential differences between such airline-based approaches and the RBS procedure are illustrated in Figure 4.1. Figure 4.1 shows a simple GDP instance in which one of the airlines has its flights at the end of the program (one could think of this as banks of flights in a hub-and-spoke network). Note that the oag times range from 12:00 to 12:18 and the slot time to be allocated from 12:00 to 12:36. The example shows that under

RBS the second airline would absorb most of the delay, while an airline-based approach could lead to a more even distribution of the delays.

The potential to reduce the disadvantage for airlines whose schedule tends towards the end of a GDP indicates the potential attractiveness of an airline-based approach. At the same time, however, the use of this multi-objective approach also has disadvantages. In the example, for instance, one could argue that the first airline is unduly penalized: only airline A can use the first slot, yet assigning this slot to airline A implies a larger delay for its remaining flights. Another potential disadvantage is that the multi-objective approach does not necessarily uniquely define the allocation of slots to airlines: there may be a large number of "optimal" assignments with significant differences in the distributions of the flight delays within an airline.

4.2.2 Cost-Sharing Methods

An alternate approach to the allocation of slots during a GDP follows by interpreting the distribution of delays as a cost-sharing problem. Intuitively, a cost-sharing problem is perhaps best explained by considering a production technology that is jointly owned by a given set of users (cf. [52]). Each of the users may have certain demands, the sum of which can only be produced at a certain cost. The resulting problem is how to distribute this cost among the users. Examples include the allocation of joint overhead costs of a firm among its divisions ([68]), and setting fees for the use of a communication network ([17]). Cost-sharing problems may be categorized according to the structure of the cost function representing the production technology (see [49]). In *homogeneous* cost-sharing problems, the production technology corresponds to a "one input-one output" model, that is, the technology produces a single type of good. Each user i demands a quantity q_i of the good and the total cost equals $C(\sum_i q_i)$, with $C : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. *Heterogeneous* cost-sharing models, on the other hand, correspond to technologies that may produce multiple types of goods.

An important strand of the literature on cost-sharing problems follows the "axiomatic" approach ([49]). The axiomatic approach imposes certain normative criteria, that is, a set of axioms that represent properties desired in a rationing method. These axioms may represent not only equity concepts, but also structural invariance and incentive criteria (see [49]). This has led to a number of different cost-sharing mechanisms, each of which is characterized by a different set of axioms. One example is the *proportional mechanism*, in which cost shares are simply proportional to demands. In the case of heterogeneous demands , however, its application is limited, since the different goods may not be comparable. Another example is the *serial mechanism* ([50]), which is similar to the uniform gains rules used in rationing problems. However, this approach may also be difficult to extend to heterogeneous cost-sharing problems ([36]). Finally, *value mechanisms* for cost-sharing are inspired by the Shapley value used in cooperative games ([59]). These methods rely, in one way or another, on the incremental or marginal cost imposed by a user's demand. Two important cost-sharing methods are the Shapley-Shubik and the Aumann-Shapley rules (see [17], [45], [68]).

It should be noted that in certain situations a cost-sharing method may yield a decentralization device, in which users may strategically submit their demands (i.e. the rule leads to a non-cooperative game). In this case the incentive properties, in particular the strategy-proofness of the method, become important (see [46], [50]). However, these issues are less of a concern within the present context: an airline's claims/demand in a GDP are defined by the planned flight schedule, which are fixed well before the GDP is executed and cannot be modified on a daily basis[2].

Model Formulation

Here, we interpret the allocation of slots during a GDP as a cost-sharing problem in which the cost corresponds to the resulting delays. The basic idea is to interpret the airport as a production technology that is jointly owned by a set K of agents. The outputs "produced" by the airport are flight arrivals (arrival slots), which are differentiated by their arrival time. The set of agents can be either the individual flights or the airlines, that is, $K = \mathcal{F}$ or $K = \mathcal{A}$. In both cases, each agent k will demand a certain amount $q(k) \in \mathbb{N}_+^n$ of the output (e.g. if $q(k)_j = 2$ agent k demands two arrivals at time j). If the agents are individual flights ($K = \mathcal{F}$) the demands $q(f)$ will be unit vectors, with $q(f)_j = 1$ if $j = oag_f$ and $q(f)_j = 0$ otherwise. If, on the other hand, the agents are airlines ($K = \mathcal{A}$) the demands $q(a)$ are defined as $q(a)_j = |\{f \in \mathcal{F}_a : oag_f = j\}|$. The aggregate demand $q \in \mathbb{N}_+^n$ is simply the sum of the individual demands, i.e. $q = \sum_{k \in K} q(k)$.

Given the capacities c and an aggregate demand vector q, we have to determine the cost, that is, the delay required to produce the arrivals demanded. This can be done by introducing a delay vector $d(c, q) \in \mathbb{N}_+^n$, in which an element $d_j(c, q)$ represents the delay incurred at slot j. The delay vector can easily be

[2] In theory, an airline could artificially inflate its planned schedule to secure additional slots during a GDP. This, however, appears to be highly unlikely since there are several detrimental effects in doing this.

defined recursively:

$$d_0(c,q) = \max(q_0 - c_0, 0),$$
$$d_j(c,q) = \max(d_{j-1}(c,q) + q_j - c_j, 0) \quad (1 \leq j \leq n-1).$$

In other words, the delay at a slot equals the number of flights at that slot that cannot be assigned to the slot. The total delay $D(c,q)$ is then expressed straightforwardly as

$$D(c,q) = \sum_{j=0}^{n-1} d_j(c,q).$$

Under the assumption that the number of slots n and the number of agents K remain fixed, a cost-sharing *problem* is then defined as the tuple $(c, q(k)_{k\in K})$. A *solution* to the cost-sharing problem is a vector $x \in \mathbb{R}_+^K$ specifying a cost(delay) share for each agent such that

$$\sum_{k\in K} x_k = D(c,q).$$

More generally, a cost-sharing *method* can be defined as follows.

Definition 4.2.2. *A cost-sharing method is a mapping $x : \mathbb{N}_+^n \times \mathbb{R}_+^{K\times n} \to \mathbb{R}_+^K$ that associates with each cost-sharing problem $(c, q(k)_{k\in K})$ a solution, such that*

$$\sum_{k\in K} x(c, q(k)_{k\in K})_k = D(c, \sum_{k\in K} q(k)).$$

□

In other words, a cost-sharing method associates with each instance of a GDP an allocation of delays. A simple method would be to divide the costs equally. This, however, violates a basic principle that cost shares should reflect the agents' contribution to the delay. This idea is taken into account in the well-known Shapley Value, which has its origins in cooperative game theory.

The Shapley Value

The Shapley value is based on imposing certain minimum requirements on the possible cost-sharing methods, the so-called dummy, impartiality, and additivity axioms. For a general cost function $C : 2^K \rightarrow \mathbb{R}_+$ which associates a cost with each group of agents in K, these properties may be defined as follows. For any $l \in K$ and $L \subseteq K$, we define $\delta_i(C, L) = C(L + \{i\}) - C(L)$.

Definition 4.2.3. *A cost sharing method $x(C)$ satisfies the* dummy *property if*

$$\delta_i(C, L) = 0 \quad \textit{for all } L \subseteq K \quad \Rightarrow \quad x(C)_i = 0 \quad \textit{for all } i \in K.$$

□

Definition 4.2.4. *A cost sharing method $x(C))$ is* impartial *if, for any $i_1, i_2 \in K$,*

$$\begin{aligned} & \delta_{i_1}(C, L) = \delta_{i_2}(C, L) \quad \textit{for all } L \textit{ s.t. } i_1, i_2 \notin L \\ \Rightarrow \quad & x(C)_{i_1} = x(C)_{i_2}. \end{aligned}$$

□

Definition 4.2.5. *Let $C_1, C_2 : 2^K \rightarrow \mathbb{R}_+$ be two cost functions such that $C = C_1 + C_2$. Then, a cost sharing method $x(C)$ is* additive *if*

$$x_i(C) = x_i(C_1) + x_i(C_2) \quad \textit{for all } i \in K.$$

□

Informally, the dummy property states that players who do not contribute to the cost are not charged any cost. The impartiality property implies that players who enter the cost function symmetrically are charged the same amount. The additivity property states that if cost function can be decomposed, the resulting cost allocation can also be decomposed. Observe that the additivity property directly applies to the allocation of slots in a GDP, as the cost function $D(c, q)$ is expressed as the sum of the delays incurred at each slot.

The Shapley value is the unique method that satisfies these three axioms, and can be characterized as follows

Definition 4.2.6. *The Shapley value $x(C)$ is defined as*

$$x(C)_i = \sum_{0 \leq s < n} \frac{s!(n-s-1)!}{n!} \sum_{S \subseteq N - \{i\} : |S| = s} \delta_i(C, S),$$

for $0 \leq i < N$. □

Intuitively, the Shapley value can be interpreted as a (random) priority method. For a given priority ordering of the players N, a *priority method* allocates to each player its incremental cost, i.e., the additional cost incurred by its addition to the coalition (after all players with a higher priority have been added). The Shapley value assigns each agent his *average* incremental cost over all priority orderings, that is, a priority ordering is chosen randomly. As such, the Shapley value is related to the RBS algorithm. Consider, for example, the case in which slots are allocated to individual flights. If the ordering corresponded to the ordering of the flights by OAG times, the priority method would equal the RBS algorithm. Thus, the Shapley value differs from RBS in that a priority ordering is chosen randomly. If, on the other hand, slots are allocated to airlines, the Shapley value randomly prioritizes airlines (i.e. under any particular ordering the airline with the highest priority would receive the "best" slots for its flights, etc.).

4.2.3 Issues

The Shapley value is a well-known solution concept, and is commonly applied in cost-sharing problems (cf. [49]). When applied within the current context of allocating slots during GDPs, however, the use of the Shapley value introduces a number of questions. These issues can be explained using the example shown in Figure 4.2.

GDP Instance:

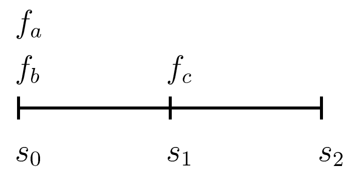


Figure 4.2: Example: Shapley value

Figure 4.2 shows a simple GDP instance, in which flights f_a, f_b and f_c have to be assigned to slots s_0, s_1 and s_2. Each slot has capacity 1, and $oag_a = oag_b = 0$, $oag_c = 1$. Alternatively, we can also say that one unit of delay is to be assigned at slot s_0 and one unit at slot s_1. The Shapley value can be calculated by determining, for each group of flights, the resulting delay if only its members were present (i.e. the characteristic function):

- $D(\{a\}), D(\{b\}), D(\{c\}), D(\{a, c\}), D(\{b, c\}) = 0;$

  $D(\{a, b\}) = 1;$

  $D(\{a, b, c\}) = 2;$

Alternatively, we can use the additivity property to decompose D into two cost functions D_0 and D_1. D_0 can be associated with the first unit of delay, and D_1 with the second unit of delay. The resulting characteristic functions are defined as

- $D_0(\{a\}), D_0(\{b\}), D_0(\{c\}), D_0(\{a,c\}), D_0(\{b,c\}) = 0$;
  $D_0(\{a,b\}), D_0(\{a,b,c\}) = 1$;

- $D_1(\{a\}), D_1(\{b\}), D_1(\{c\}), D_1(\{a,c\}), D_1(\{b,c\}), D_1(\{a,b\}) = 0$;
  $D_1(\{a,b,c\}) = 1$.

The delay distributions d_0 and d_1 that correspond to the Shapley value of these games are easily determined:

- $d_0(a) = d_0(b) = \frac{1}{2}; d_0(c) = 0$;

- $d_1(a) = d_1(b) = d_1(c) = \frac{1}{3}$;

Corresponding to this allocation of delays there is an overall slot allocation, i.e.

- $s_0(a) = s_0(b) = \frac{1}{2}$, $s_0(c) = 0$;

- $s_1(a) = s_1(b) = \frac{1}{6}$, $s_1(c) = \frac{2}{3}$;

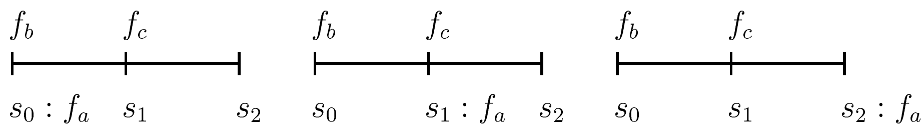
- $s_2(a) = s_2(b) = \frac{1}{3}$, $s_2(c) = \frac{1}{3}$;

Note that these values may be interpreted as the probability of being assigned a slot (i.e., $s_0(a)$ represents the probability that f_a is assigned to slot s_0).

As discussed before, the Shapley value is the unique allocation that satisfied the dummy, impartiality, and additivity properties. The questions that arise from using the Shapley value in this case, however, lie in some intuitive properties the allocation does not have. For instance, a basic fairness principle states that allocation is fair only when every subgroup believes it be so; that is, every subgroup should be satisfied that they share the slots assigned to them in a fair way (cf. [86], p.170). This concept is also known as the consistency principle, and plays an important role in a variety of allocation problems. While different

Subproblems:

a assigned to 0 w.p. $\frac{1}{2}$	a assigned to 1 w.p. $\frac{1}{6}$	a assigned to 2 w.p. $\frac{1}{3}$



Shapley value associate with each subproblem:

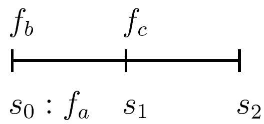
- a assigned to 0: $d_0(b) = 1, d_1(b) = \frac{1}{2}; d_0(c) = 0, d_1(c) = \frac{1}{2}$;
- a assigned to 1: $d_0(b) = 0, d_1(b) = 0; d_0(c) = 0, d_1(c) = 1$;
- a assigned to 2: $d_0(b) = 0, d_1(b) = 0; d_0(c) = 0, d_1(c) = 0$.

Figure 4.3: Example: Shapley value, Consistency

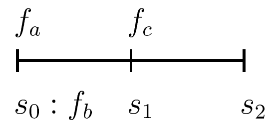
definitions of the consistency principle may exist, the basic idea is always that an allocation rule should be invariant when restricted to subgroups of agents (e.g., the removal of an agent and its share should not affect the allocation of the remaining agents). Let us now illustrate how the notion of consistency might be interpreted within the current context, by considering the slot shares of f_b and f_c in the example above. We saw that, under the Shapley value, flight f_a will be assigned slot s_0 with probability $\frac{1}{2}$. Thus, with probability $\frac{1}{2}$, f_b and f_c (as a group) will be assigned slots s_1 and s_2. Similarly, with probability $\frac{1}{6}$ f_b and f_c (again as a group) will be assigned slots s_0 and s_2, and with probability $\frac{1}{6}$ they will be assigned slots s_0 and s_1. In each of these three cases, we could assign the remaining slots to f_b and f_c according to the Shapley value. The resulting delay distributions for each situation are shown in Figure 4.3.

Subproblems:

a assigned to 0
w.p. $\frac{1}{2}$



b assigned to 0
w.p. $\frac{1}{2}$



Shapley value associate with each subproblem:

- a assigned to 0: $d_0(b) = 1, d_1(b) = \frac{1}{2}; d_0(c) = 0, d_1(c) = \frac{1}{2};$
- b assigned to 0: $d_0(a) = 1, d_1(a) = \frac{1}{2}; d_0(c) = 0, d_1(c) = \frac{1}{2};$

Figure 4.4: Example: Shapley value, Composition

This approach, however, would yield an allocation that is different from the one we obtained by applying the Shapley value to the overall problem in Figure 4.2. Consider for instance the delay of flight f_b. In the overall problem, we have $d_1(b) = \frac{1}{3}$. However, taking the weighted average over its delay shares in the subproblems will yield $d_1(b) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$. The same holds for flight f_c. Thus, the Shapley value does not obey this notion of consistency[3].

Another issue that arises with the use of the Shapley value stems from the so-called Composition principle ([49], [86]). Informally, the composition principle states that the allocation can be decomposed into stages. To illustrate this, consider again the example in Figure 4.2. Suppose now that at first we only want

[3]An alternative consistency axiom can be used to characterize the Shapley value (see [86]). However, this axiom deals with the removal of agents in a way that cannot be viewed as a fixed assignment to a slot (see [49] for a more detailed discussion).

to ration the first slot (a practical reason for doing so is weather uncertainty). Subsequently, we may have to ration the remaining slots among the remaining flights. The composition principle states that this should not affect the overall allocation. That is, the (expected) slot shares should be the same whether we allocate in stages or not. Unfortunately, the Shapley value does not satisfy this concept. To see this, suppose we were to first assign flight f_a or f_b to slot s_0 in the example above. This would lead to two possible cases, which are shown in Figure 4.4. In each of these cases, we could as before apply the Shapley value to determine the resulting delays. Again, this would yield an allocation that is different from the one we obtained by applying the Shapley value to the overall problem.

At the heart of these issues is the interpretation of the airport as a "production technology" that is jointly owned by the airlines. This implies that all flights have equal claims to all the slots, even if the flight cannot use the slot (for instance, in the example flight f_c would have a claim to one-third of slot s_0). As a result, the allocation problem will introduce a bargaining situation in which flights will trade claims on earlier slots they cannot use for shares of later slots. The allocation in the example, for instance, may be explained as the result of a trade in which flight f_c trades its claim on slot s_0 for part of the other flights' share of slot s_1. While the notion that flights have equal claims to all the slots could potentially be a valid approach, our discussion illustrated that its use raises some practical difficulties. Moreover, the idea that claims are traded also raises questions, since it is public knowledge that an airline or flight cannot use the capacity (e.g. the OAG times are known well in advance).

4.3 Axiomatic Slot Allocation

The previous discussion leads us to ask which methods do satisfy these properties, and what principle underlie a flight's claim to the slots. To answer these questions, we formulate the GDP problem as a general allocation problem. We postulate a set of axioms that are more applicable within the context of GDPs and determine the (class of) allocation methods that satisfy these axioms. In other words, whereas the use of the Shapley value assumes a given distribution of slot shares, our objective is to determine a distribution of the shares. It should be emphasized that the models and axioms we introduce are closely based on those proposed in [48] and [51]. However, these approaches consider the allocation of homogeneous demands (e.g. their models correspond to the situation in which all flights would be scheduled to arrive at the start of the GDP). In our case, however, the different arrival times of flights introduce heterogeneous demands. This complication necessitates the use of a more general model, based on the approach outlined by Young (see [86], Appendix A).

Allocating Slots to Flights

As a first step, we consider the situation in which the agents correspond to the individual flights. To define the allocation problem we let $\mathcal{F}$, the set of flights, be the claimants and assume a given capacity vector $c \in \{0,1\}^n$. As before, we assume the existence of a final slot s_n with unbounded capacity. Associated with each flight is its type τ_f which equals its oag time, that is, $\tau_f = oag_f$. We let $\tau \in \mathbb{N}_+^{\mathcal{F}}$ represent the vector of all flight types, and $\tau_F \in \mathbb{N}_+^F$ the vector of types for any subset $F \in \mathcal{F}$.

Associated with each set of flights F and capacities c is a set of feasible and efficient allocations

$$P(F,c) = \{x \in \{0,1\}^{F\times n} : \sum_{f\in\mathcal{F}} x_{f,j} = c_j \quad \text{for all } 0 \le j < n,$$
$$\sum_{j=\tau_f}^{n} x_{f,j} = 1 \quad \text{for all } f \in F\}.$$

Observe that the first constraint implies that all available slots are used, which in general might not be possible. However, for any combination of capacities and flights, an efficient (delay-minimizing) solution will always occupy the same set of slots (see [81]). Consequently, without loss of generality it is always possible to adjust the capacities c such that the constraint will hold.

An *allocation problem* consists of a tuple (τ_F, P), where τ_f represents the types for a given set of flights F and $P \subseteq P(F,c)$ for some capacity profile c. We note that the inclusion of subsets of the feasible set $P(F,c)$ will become clear in the definition of the axioms. A *probabilistic allocation rule* X associates with each allocation problem (τ_F, P) a *random* selection of allocations in the feasible set P. Thus, any allocation can also be represented as a convex combination of the possible assignments, i.e.

$$X(\tau_F, P) = \sum_k \lambda_k x^{(k)}, \lambda_k \ge 0, \sum_k \lambda_k = 1.$$

where $x^{(k)} \in P$ represents a possible assignment of flights to slots. In other words, the allocation rule $X(\tau_F, P)$ selects each assignment $x^{(k)}$ with probability λ_k. Observe that $X(\tau_F, P)_{f,j}$ may be interpreted as the probability that that f is assigned to slot j.

Two fundamental principles of equity are *impartiality* and *consistency*. Impartiality defines the notion that equals should be treated equally, and can be defined as follows (cf. [86]).

Definition 4.3.1. *A probabilistic allocation rule X is* impartial *if for any allocation problem (τ_F, P) and any permutation π of F,*

$$X(\tau_F \circ \pi, P \circ \pi) = X(\tau_F, P) \circ \pi,$$

where we view τ_F as a function from F to $\mathbb{N}_+$, an allocation $x \in P$ as a function from F to $\mathbb{N}_+^{n+1}$, and $X(\tau_F, P)$ as a function from F to $\mathbb{R}_+^{n+1}$. □

This property states that the allocation rule is independent of the indexing of the flights: if two flights are indistinguishable in type and in the feasible set, they will receive the same slot shares. The consistency concept was discussed informally in the previous section. The concept of consistency has a long history, and has been applied in a number of different settings (see [77], [86] p.173 for an overview). Among others, variants of the consistency principle have used in apportionment problems ([7]), cost-sharing and rationing problems ([49]), and bargaining problems ([76]). Here, we follow a definition of consistency proposed by Moulin ([48]), which recognizes the probabilistic nature of the underlying allocation problem. To formalize this concept, we define for a given feasible set P, any $f \in F$, and any slot index $j : 0 \leq j \leq n$ the reduced feasible set $P(f, j)$

$$P(f, j) = \{x \in P : x_{f,j} = 1\},$$

which represents a set of feasible and efficient allocations for the flights in $F - \{f\}$ with slot s_j unavailable. The formal definition is as follows.

Definition 4.3.2. *A probabilistic allocation rule X is* consistent *if for any allocation problem (τ_F, P) and any $f \in F$,*

$$X(\tau_F, P)_{f',j'} = \sum_{j=0}^{n} X(\tau_F, P)_{f,j} X(\tau_{F-\{f\}}, P(f, j))_{f',j'}$$

for all $f' \in F - \{f\}$, $0 \leq j' < n$. □

In other words, the consistency property states that the expected slot shares should be independent of the order in which flights are assigned to slots. While imposing impartiality and consistency has a significant impact on the potential allocation rules, the class of rules that satisfy these axioms is still complex and not easily characterizable. Nevertheless, their impact can be illustrated by considering the case where exactly one slot is available.

Proposition 4.3.3. *Let X be a consistent, impartial allocation rule, let e_j represent a unit capacity vector whose capacity at slot j equals 1, and let τ_F be any demand profile. Then, there exists a set of weights $\lambda_i^j (0 \leq i \leq j)$ and a weak ordering[4] $\succeq_j$ over the OAG times $0 \leq i \leq j$ such that*

$$X(\tau_F, P(F, e_j))_{f,j} = \frac{\lambda_{\tau_f}^j}{\sum_{g \in F} \lambda_{\tau_g}^j} \quad \text{if } \tau_f \succeq_j \tau_{f'} \text{ for all } f' \in F, \text{ and}$$

$$X(\tau_F, P(F, e_j))_{f,j} = 0 \quad \text{otherwise.}$$

Proof. See Appendix ∎

In other words, the flights arriving at slot j are partitioned into priority classes based on OAG times. Within each priority class, the slot is assigned according to a probability based on the weights λ_i^j. Note that by definition of consistency, an impartial and consistent allocation rule can be characterized by a collection of weights and weak orderings $(\lambda_i^j, \succeq_j)_{0 \leq j \leq n}$. It is an open question whether the reverse also holds.

Whereas consistency represents a certain invariance under changes in the number of flights, the composition principle states an invariance under changes in the capacity over time. The composition principle has its origins in so-called

[4]a weak ordering or preordering is an ordering relation $\succeq_P$ that is connected (i.e. $j \succeq_P j'$ or $j' \succeq_P j$ or both) and transitive.

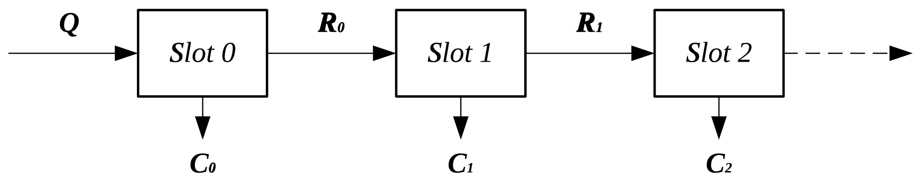


Figure 4.5: Interpretation of Decomposition Axiom

equal sacrifice rule used in taxation problems ([85]), where the principle represents an invariance under the order in which taxes are levied. Here, composition represents the idea that the allocation can be decomposed into stages without affecting the overall assignments. This definition closely corresponds to the definition given in [47] and in [48]. The formal definition is as follows.

Definition 4.3.4. *For any capacity vector c and any period t : $(0 \leq t < n)$, define $c^t = (c_0, \ldots, c_t, 0, \ldots, 0)$. Then, a probabilistic allocation rule X satisfies the* composition *property if for any vector of types τ_F, any capacity profile c and any time period t,*

$$x(\tau_F, P(F, c)) = \sum_k \lambda_k x(\tau_F, R_t(F, c, x^{(k)})),$$

where $x(\tau_F, P(F, c^t)) = \sum_k \lambda_k x^{(k)}$ and

$$R_t(F, c, x^{(k)}) = \{x \in P(F, c) : x_{f,j} = x_{f,j}^{(k)} \quad \textit{for all } f \in F, 0 \leq j \leq t\}.$$

□

In other words, composition states the expected slot shares do not change if we first allocate the slots up to period t, and subsequently assign the remaining slots. A consequence of decomposition and consistency is that the allocation problem can be divided into a sequence of allocation problems as shown in Figure 4.5. That is, each slot may be viewed as a station which assigns its available capacity among the incoming flights and sends the remaining flights to the subsequent station.

The final axiom we impose defines a certain regularity on the manner in which each station allocates its capacity. Specifically, the idea is that is that cost-sharing methods should be time-independent, that is, if identical (and feasible) demand profiles were to arrive at two different stations (slots), the capacities should be allocated in the same way.

Definition 4.3.5. *Let e_j represent the unit capacity vector whose capacity at slot j equals 1. Let j_1, j_2 be two slots such that $0 \leq j_1 < j_2 < n-1$, and let τ_F represent a set of types such that $\tau_f \leq j_1$ for all $f \in F$. Then, a probabilistic allocation rule X is* time independent *if*

$$X(\tau_F, P(F, e_{j_1}))_{f,j_1} = X(\tau_F, P(F, e_{j_2}))_{f,j_2}.$$

□

The combination of the above mentioned axioms restricts the allocation rules that can be used to allocate slots to flights during a GDP. To characterize the allocation rules that satisfy the combination of these axioms, we let Q represent a *priority standard*, that is, a weak ordering of the OAG times $0, \ldots, n-1$. The use of priority standards in allocative situations is discussed in more detail by Young ([86]). More precisely, a priority standard imposes an ordering on flights that arrive at different times, which can be used to allocate slots to flights. We note that a priority standard is not necessarily equal to the natural ordering imposed by OAG times, e.g. the priority of a flight arriving at time 4 could be less than, equal to, or greater than the priority of a flight arriving at time 6. Intuitively, a *priority method* assigns slots according to a priority standard Q. That is, a priority method sequentially assigns the slots and, at each step, the current slot is randomly assigned to one of the remaining eligible flights that has the highest priority according to its type τ_f under P. To formalize this notion, we introduce the following two definitions.

Definition 4.3.6. *For any capacity profile c and set of flights F, a solution $x \in P(F, C)$ is* equitable *with respect to a priority standard Q if for any two flights $f, f' \in F$ such that $x_{f,j} = 1$, $x_{f',j'} = 1$ and $j < j'$, then $\tau_f \succeq_Q \tau_{f'}$ or $\tau_{f'} > j$.* □

Definition 4.3.7. *For any capacity profile c and set of flights F, the* priority method *Q(F,c) based on Q consists of all solutions $x \in P(F, c)$ that are equitable with respect to Q.* □

These definitions allow us to state the following Theorem.

Theorem 4.3.8. *Let c be any capacity profile and F be any set of flights. Then, for any probabilistic allocation rule X that is impartial, consistent, time independent, and satisfies composition, there is a priority standard Q such that*

$$X(\tau_F, P(F, c)) = \sum_{x \in Q(F,c)} \frac{1}{|Q(F, c)|} x.$$

Proof. See Appendix. □

In other words, a probabilistic allocation rule randomly selects one of the allocations in a priority method. Thus, the combination of a local decomposition in stations (which allocate slots in the same way) together with a global consistency property leads to allocation rules that are priority methods.

Allocating Slots to Airlines

Theorem 4.3.8 characterizes allocation rules in which slots are assigned to individual flights. To generalize the result to the case where slots are assigned to airlines we define, for each airline $a \in \mathcal{A}$, a type vector $\tau_a \in \mathbb{N}_+^n$ such that $\tau_{a,j} = |\{f \in \mathcal{F}_a : oag_f = j\}|$. Again, we associate with each set of airlines A and capacity profile c a set of feasible and efficient allocations

$$\begin{aligned} P(A,c) = \{x \in \{0,1\}^{A\times n} : &\sum_{a\in A} x_{a,j} = c_j \quad \text{for all } 0 \le j < n, \\ &\sum_{k=0}^{j} x_{a,j} \le \sum_{k=0}^{j} \tau_{a,j} \quad \text{for all } a \in A, 0 \le j < n\} \\ &\sum_{k=0}^{n} x_{a,j} = \sum_{k=0}^{n} \tau_{a,j} \quad \text{for all } a \in A\}. \end{aligned}$$

Now, a probabilistic allocation rule X associates with each allocation problem (τ_A, P) a random allocation of slots to airlines in the feasible set.

The previously defined axioms readily generalize to the case in which slots are assigned to airlines. Impartiality can be restated by requiring that if two flights are indistinguishable in type and in the feasible set, they will receive the same slot shares. The consistency axiom can be generalized by requiring that all slot shares should remain invariant after assigning an airline its (random) share. Similarly, the composition and time independence axioms can be adjusted by framing the requirements in terms of allocations to airlines.

In addition to these axioms we also impose collusion-proofness, which may be defined as follows.

Definition 4.3.9. *Let c represent any demand profile, and let $A \subseteq \mathcal{A}$ any set of airlines. For any airline $a \in A$, define a_1, a_2 and their respective types τ_{a_1}, τ_{a_2} such that $\tau_{a_1} + \tau_{a_2} = \tau_a$, and let $A' = A - \{a\} + \{a_1, a_2\}$. Then, a probabilistic* *allocation rule is collusion-proof if*

$$X(\tau_A, P(A,c))_{a,j} = X(\tau_{A'}, P(A',c))_{a_1,j} + X(\tau_{A'}, P(A',c))_{a_2,j},$$

for all $0 \leq j < n$. $\square$

Informally, the idea behind the collusion-proofness property is that no airline or group of airlines should have an advantage or disadvantage from grouping its flights. While our model assumes all airlines and demands are known well in advance of the GDP, the underlying idea still has some appeal within the context of GDPs: in many cases an large carrier will manage operations for one or several smaller carriers, and it would be undesirable if this affected their overall allocation.

The following theorem immediately follows from Theorem 4.3.8 and the definition of collusion-proofness.

Theorem 4.3.10. *Let* c *be any capacity profile, and* $A \subseteq \mathcal{A}$ *be any set of airlines. Then, for any probabilistic allocation rule* X *that is impartial, consistent, time independent, collusion-proof and satisfies composition, there is a priority standard* Q *such that*

$$X(A, P(A,c))_{a,j} = \sum_{f \in F_a} \sum_{x \in Q(F,c)} \frac{1}{|Q(F,c)|} x_{f,j}.$$

$\square$

In other words, each airline will receive the sum of the shares its flights would obtain under a priority method. The resulting allocation rules may be viewed as a certain proportional scheme: whereas for flights a slot is randomly assigned to one the flights of highest priority, in the case of airlines a slot is assigned to an airline with a probability that is proportional to the number of its remaining flights in the highest priority class.

Interpretation

Theorems 4.3.8 and 4.3.10 strongly suggest the use of priority methods to allocate slots, given their appealing structural properties within the context of GDPs. As such, the results of this section provide a strong foundation for the RBS procedure, which corresponds to a priority standard Q where $i \succeq_Q j$ iff $0 \le i < j < n$. At the same time, however, the Theorems state that any priority standard yields these structural properties which could indicate a number of alternate possibilities. One possibility in particular would be to give all flights equal priority, that is, a priority method with the standard Q in which $i \succeq_P j$ for all $0 \le i, j < n$. This is in some sense the "opposite" of RBS, as there are no strict priorities. In this case, the resulting allocation method corresponds to the proportional random assignment scheme shown in Figure 4.6. The use of this priority standard is actually similar to the principle underlying the Shapley value, in which each flight was entitled an equal share of each slot. The difference is that in the proportional random assignment method, each flight is entitled to an equal share of each slot *it can use* (i.e. that is later than its scheduled arrival time). In the beginning of this chapter we discussed how CDM has initiated a relationship between an airline's scheduled demand and its rights to airport capacity at the day of operation; that is, an airline's flight schedule could be interpreted as a claim on the arrival capacity available during a GDP. Under RBS, for instance, a scheduled arrival may be interpreted as a service priority. Consequently, the flight schedule defines for each airline a *priority list*, which it can associate with the actual flights during the day of operation. The underlying idea here is that a period of time is allocated to carry out activities and that the time required for each activity may vary. However, the implicit assumption

Initialization Let $\mathcal{F}'_a := \{f_i \in \mathcal{F}_a : oag_i = 0\}$ and $j := 0$;

While $j < n$ **do** :

I. If $c_j = 0$, do nothing. Otherwise, randomly select an airline a' with probability proportional to $|\mathcal{F}'_a|$;

II. Assign the earliest flight $f' \in \mathcal{F}'_a$ to s_j;
Let $\mathcal{F}'_a := \mathcal{F}'_a - \{f'\}$;

III. Let $j := j + 1$;

IV. Let $\mathcal{F}'_a := \mathcal{F}'_a + \{f_i \in \mathcal{F}_a : oag_i = j\}$;

Figure 4.6: Proportional Random Assignment Mechanism

is that it will always be possible to carry out all the activities even when all activities require their maximum time. Another interpretation, however, could be that a period of time (at the airport) is allocated to carry out activities and that participants have access to portions of that time based on the extent of their planned activities. Now suppose that the extent of the time available varies so that it is not possible for all participants to carry out all of their activities. This corresponds to the proportional random assignment mechanism above, in that time is assigned in proportion to an airline's (remaining) demands. A such, under the preordering in which all arrival times have equal priority, the claim associated with each arrival corresponds to the right to land a flight at a time greater than or equal to the OAG arrival time. Thus, based on that flight the airline has equal rights to all slots available after that time.

4.4 Empirical Analysis

In this section, we empirically analyze the distribution of slots and delays among airlines during GDPs, using historical data from actual GDPs. Our analysis considers the difference in delays between Ration-By-Schedule and the Proportional Random Assignment Mechanism.

We studied the difference between the delays airlines would incur in the ration-by-schedule and in the proportional random assignment mechanism. For different airports, we considered a number of actual GDPs during the period January-May 2001. For each of these GDPs we determined the delay each flight would be assigned under RBS, and the *average* delay each flight would be assigned under the proportional random assignment mechanism. Subsequently, we calculated the average delay for each airline. The empirical results are shown in Figures 4.7, 4.8, and 4.9. The graphs in these figures represent, for a selected number of airlines, the difference between an airline's average delay under the proportional random assignment mechanism and under RBS (a negative number means the airline would have been allocated less delay under the proportional random assignment mechanism). The results indicate that, on the aggregate, there appears to be little difference in the delays incurred by airlines under the proportional random assignment mechanism and under RBS. While substantial differences may occur during any given GDP, there appear to be no systematic biases, and generally speaking these differences decrease for airlines with larger numbers of flights during a GDP. The lack of difference between RBS and proportional random assignment delays is somewhat surprising, given the imbalances commonly caused by the practice of scheduling banks of flights. Suppose, for instance, that airline A has a bank of flights early in the schedule, while

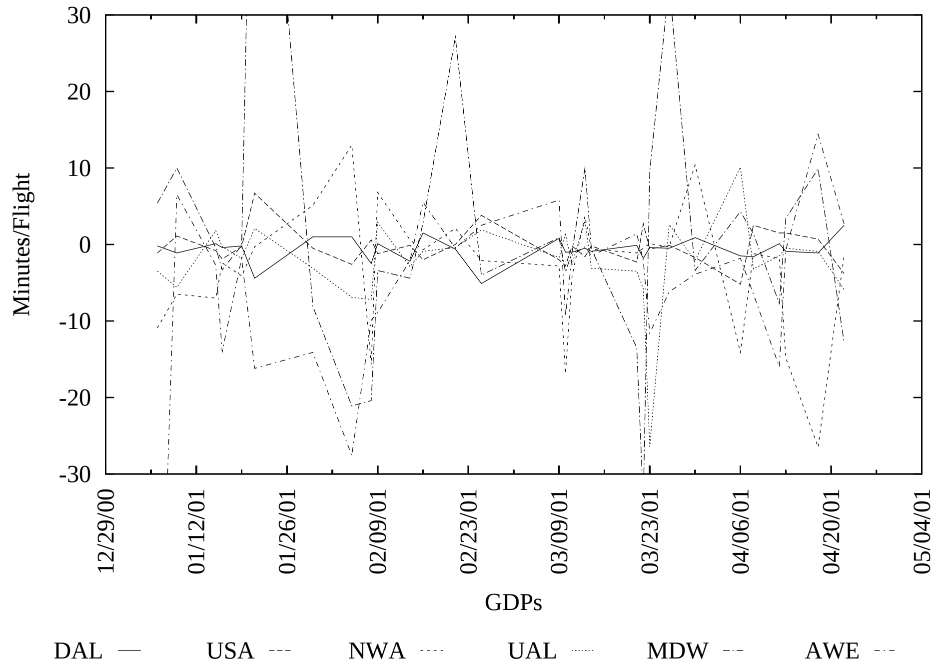


Figure 4.7: Delay Comparison : Logan Airport, Boston

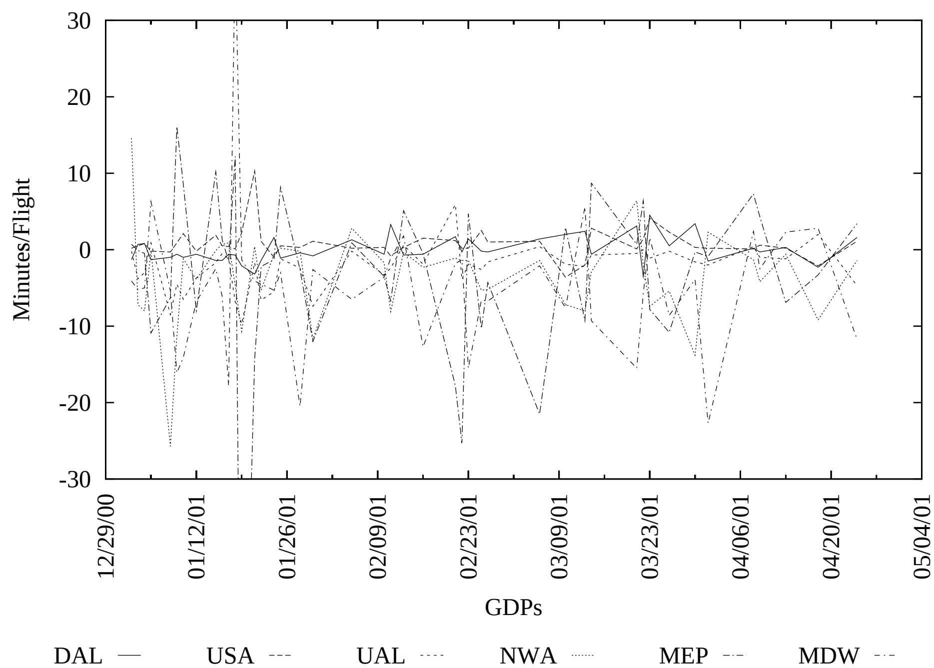


Figure 4.8: Delay Comparison : LaGuardia Airport, New York

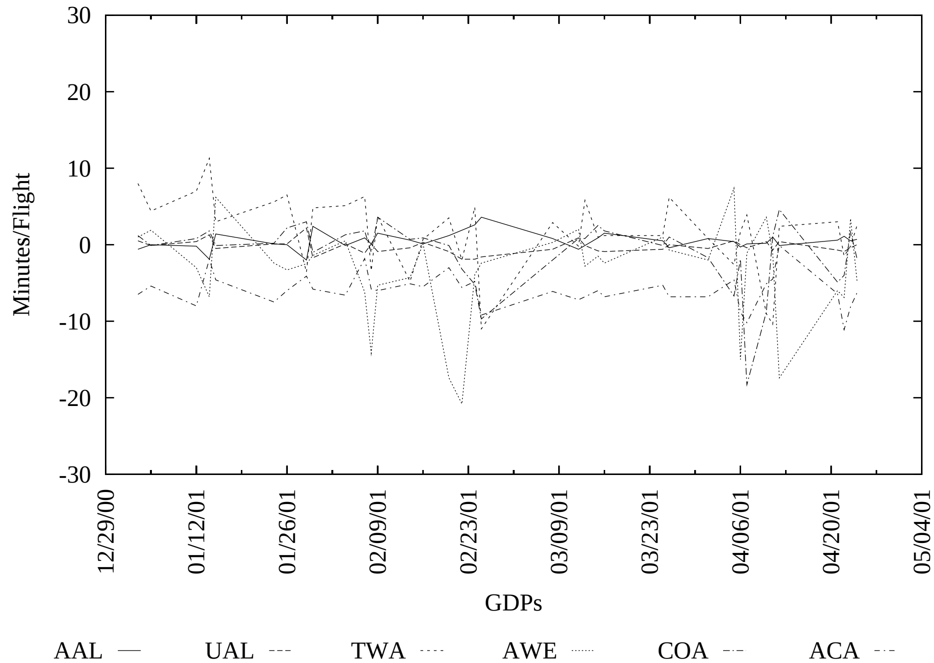


Figure 4.9: Delay Comparison : Logan Airport, Boston

airline B has a bank late in the schedule. Then, one might expect that airline B would receive significantly less delay under the proportional random assignment method than under RBS.

To further illustrate the difference between RBS and the proportional random assignment method, we therefore compared their differences in delay in the three scenarios shown in Figure 4.10. These three scenarios depict different classes of OAG schedules, parameterized by a single variable z : $(0 \leq z \leq 60)$. The potentially largest imbalance in the schedule occurs in Scenario 1; scenarios 2 and 3 represent situations with successively smaller imbalances. Furthermore, the imbalance in each scenario is largest with parameter values $z = 0$ and $z = 60$. If $z = 30$, the schedules for both airlines are identical. Figure 4.11 shows, for each scenario, the differences between RBS and proportional random assignment

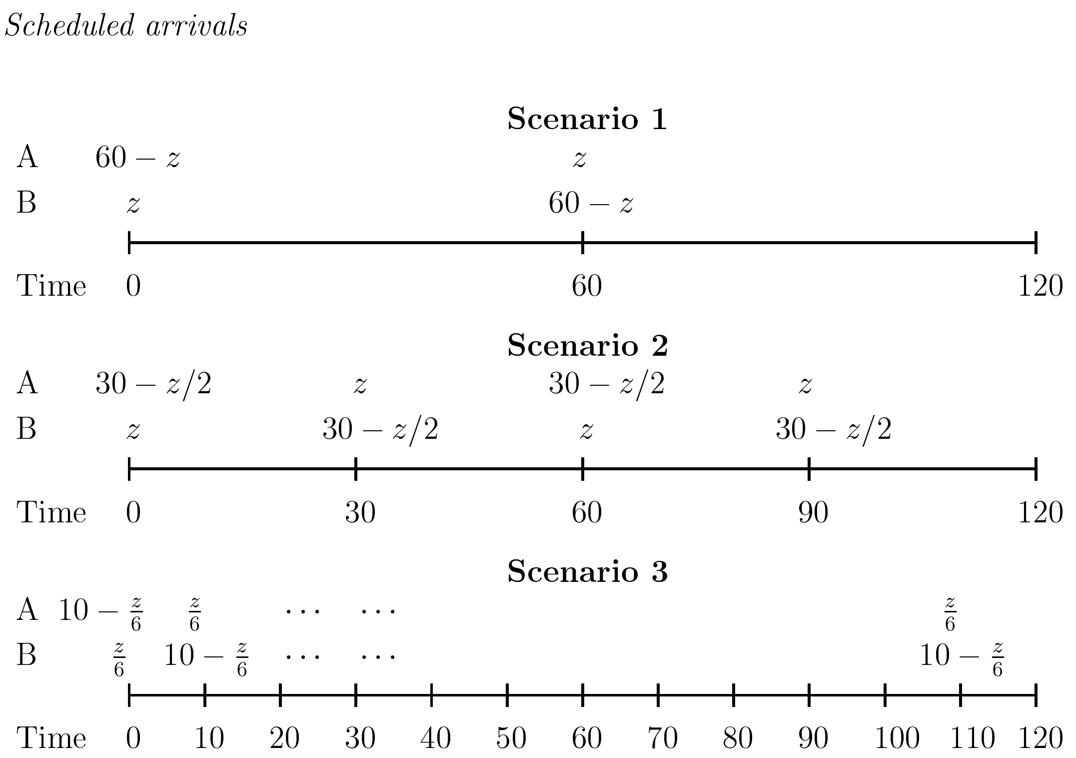


Figure 4.10: Example: OAG Schedule Scenarios

delays (per flight) as a function of the parameter z for airline A. To determine the delays, we assumed that the arrival capacity during the GDP was reduced from 60 to 30 flights per hour. We note that for parameter values $0 \leq z < 30$, airline A would receive more delay under RBS while for values $30 \leq z < 60$ the RBS delays would be less than the proportional random assignment delay. The results show that reductions in the imbalance (e.g. going from scenario 1 to scenario 3) quickly reduce the differences in delays. While imbalances in Scenario 1 can lead to significant differences in the delay per flight, the differences are substantially smaller in scenario 3. As such, this example may partly explain the empirical results observed in Figures 4.7, 4.8, and 4.9.

Finally, it should be noted that the use of the probabilistic allocation schemes

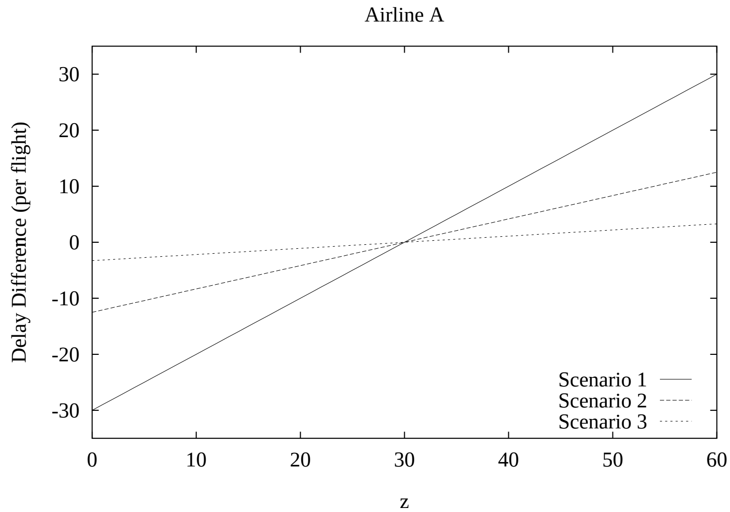


Figure 4.11: Delay Differences by Scenario

may introduce a substantial amount of variance in the delay assigned to each airline. For instance, there is a positive probability that an airline will be assigned the last slots in a GDP. To illustrate this, we ran 500 replications of the proportional random assignment mechanism for a *single* GDP. Figure 4.12 shows the distribution of delays (here, the average error represents the per flight devi-

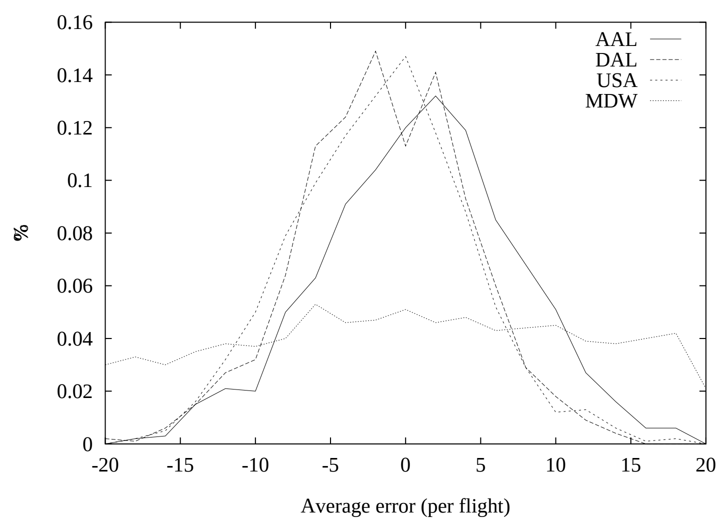


Figure 4.12: Delay Distribution using Proportional Random Assignment

ation w.r.t. the expected delay) over these replications for a selected number of airlines. These delay distributions clearly show, during any given GDP, airlines might experience significant variance in their delays. It is unlikely that such levels of variance would be acceptable to the airlines.

4.5 Discussion

This chapter introduced a formal approach to the allocation of arrival slots during GDPs. The basis for our analysis was the CDM-initiated notion that slots are assigned to airlines based on claims derived from the original flight schedules. As a first step, we introduced both multi-objective optimization methods and methods based on concepts from cooperative game theory, i.e. the Shapley value. Undesirable structural properties, however, led us to pursue a more direct approach. We postulated a set of intuitively desirable properties within the context of GDPs, and derived the class of allocation methods that satisfied them, i.e. those that are characterized by a preordering of the arrival times.

Within this class, we identified two methods: Ration-By-Schedule and the so-called proportional assignment method. While these methods appear to give surprisingly similar results in actual GDPs, their underlying philosophies are fundamentally different.

Chapter 5

Fair Slot Allocation: Equity As Near May Be

The previous chapter introduced probabilistic methods for the allocation of slots during a GDP. These methods specified fair slot shares for each airline, based on claims derived from the original flight schedules. This chapter considers methods that aim to approximate these shares in situations where the "ideal" may not be attainable. The motivation for using such methods is twofold. First, we could use these methods when the probabilistic allocation method would have an unacceptably high level of variance (e.g. if we wanted to use the proportional random assignment method). A second and more important reason in practice, however, is due to the dynamic nature of GDPs. For example, flight cancellations and delays may make it impossible to achieve the ideal share (e.g. we may view the Compression Algorithm as approximating fair slot shares).

The organization of this chapter is as follows. First, we discuss well-known methods for minimizing the deviation from an ideal share, and relate them to the actual situation during a GDP. Subsequently, we discuss how these methods can be used to manage the various dynamic changes during a GDP. Finally, we study the impact of using alternate standards of fairness.

5.1 Background

The problem of approximating a given fair or "ideal" share arises in a number of situations, most notably perhaps in apportionment and balanced just-in-time scheduling problems. This section provides an overview of these problems, introduces methods used to solve these problems, and discusses their relationship to the allocation of slots during a GDP. Subsequently, we outline the issues that arise when applying these methods to GDPs under CDM.

5.1.1 Apportionment Problems

Apportionment problems arise in situations where a set of homogeneous indivisible objects must be assigned to a group of claimants in proportion to their respective claims. Because the objects are indivisible, it is generally impossible to give each claimant his exact proportional share (his "quota"). Therefore, the question is how to distribute the objects such that each claimant's share is "as close as possible" to his quota. The classical application is the distribution of legislative seats, e.g. when seats in the U.S. House of Representatives are to be distributed among states in accordance with the proportions of their respective populations (see [7]).

Within the context of GDPs, apportionment problems are analogous with a coarse-grained approach to the allocation of slots. This is illustrated in Figure 5.1, which depicts a single-period GDP (e.g. one hour) in which C available slots ("seats") are to be distributed among airlines ("states") in accordance to

Airlines	*Flights*	
a	$f_{a,1}, \dots, f_{a,n_a}$	
b	$f_{b,1}, \dots, f_{b,n_b}$	*Delayed:* $n - C$
$\vdots$	$\vdots$	
Period	$\vdash\!\!\!-\!\!\!-\!\!\!-\!\!\!\dashv$	*Capacity:* C

Figure 5.1: Slot allocation by apportionment

their respective demands ("populations") in the flight schedule. The problem can be formalized as follows: given a set of airlines $\mathcal{A}$, their respective numbers of flights n_a with $n = \sum_{a \in \mathcal{A}} n_a$ and a capacity C, we have to find an allocation $x \in \mathbb{N}_+^{\mathcal{A}}$ such that $\sum_{a \in \mathcal{A}} x_a = C$ and the differences between all the allocations x_a and their quota $q_a = Cn_a/n$ are as small as possible. The key, of course, is the measure of deviation.

One common approach, known as Hamilton's method, operates by first assigning each airline $\lfloor q_a \rfloor$ slots (its lower quota) and the remaining slots in descending order of the fractional parts $q_a - \lfloor q_a \rfloor$. Solutions obtained by this method optimize the objective function $\min \max_{a \in \mathcal{A}} |x_a - q_a|$, as well as other objectives (see [86]). A desirable property of the solutions, which follows by construction, is that they *satisfy quota*, i.e. $\lfloor q_a \rfloor \leq x_a \leq \lceil q_a \rceil$ for all $a \in \mathcal{A}$. An undesirable property, however, is that the procedure is not *monotone*, i.e. an increase in the capacity C could lead to a decrease in the slots assigned to an airline (this is also known as the famous "Alabama paradox", cf. [86]).

This questionable feature has led to a number of other approaches, the most prominent being the so-called *divisor* methods. In a divisor method, slots are allocated iteratively to airlines: at each iteration a slot is assigned to the airline whose value of the quotient $n_a/d(x_a)$ is the highest, where x_a is the number of slots already assigned to a and $x_a \leq d(x_a) \leq x_a + 1$. A particularly attractive procedure results from letting $d(x_a) = x_a + \frac{1}{2}$, this is called Webster's Method. Solutions obtained by Webster's method minimize the function $\sum_{a \in \mathcal{A}} (x_a - q_a)^2/n_a$. Moreover, in addition to the monotonicity property, this method also has a number of other desirable features, such as consistency and unbiasedness (see [86] for a discussion of these properties).

5.1.2 Balanced Just-In-Time Scheduling Problems

A closely related, but somewhat finer-grained, approach to the allocation of slots during a GDP can be obtained by an analogy with the so-called Product Rate Variation (PRV) problem. The PRV problem arises in the determination of the sequence schedule for producing different products on a mixed-model assembly line, and has been studied extensively ([6], [13], [38], [39], [44], [69]). In certain just-in-time production systems, it is desirable that the quantity of each part used in the assembly process per unit time is kept as constant as possible; this is called levelling or balancing the schedule. Under certain assumptions (see [44]), this may be achieved by minimizing the variation in the rate at which successive units ("flights") of different product types ("airlines") are produced in the line.

An instance of the PRV problem is given by a set $\mathcal{A}$ of different product types ("airlines"), and a vector $n_a (a \in \mathcal{A})$ which represents the demand for each product type ("flights"). The production of each unit requires one unit of time. Given the total demand $n = \sum_{a \in \mathcal{A}} n_a$, an "ideal" production rate for product type a can be defined as $r_a = n_a/n$. The idea behind the ideal production rate is that at each instant we would like the production of a to

IP FORMULATION:

Min $G(x, r)$

subject to:

$$\sum_{a \in \mathcal{A}} x_{a,k} = k \quad \text{for all } k \in 1, \ldots, D$$

$$x_{a,k} \leq x_{a,k+1} \quad \text{for all } k \in 1, \ldots, n-1, a \in \mathcal{A}$$

$$x_{a,n} = n_a \quad \text{for all } a \in \mathcal{A}$$

$$x_{a,k} \geq 0 (integer)$$

Figure 5.2: IP formulation of the PRV problem

be in proportion to r_a, which would yield a perfectly levelled schedule. Such a schedule, however, is never attainable, and the objective of the PRV problem is to keep the actual production of each product somehow "as close as possible" to the ideal rate. Clearly, the PRV problem is closely related to the apportionment problem; however, whereas the apportionment allocates a fixed quantity C, the PRV problem seeks an apportionment for all quantities C between 1 and n (see [6], [13] for a discussion of their relationship). In particular, monotone methods of apportionment could also be used for the PRV problem (i.e., if we used a monotone method to solve the apportionment problems for all quantities between 1 and n, the resulting allocations would define a feasible production schedule).

The PRV problem can be formulated as an integer programming problem, as shown in Figure 5.2. In this formulation, $x_{i,k}$ represents the number of items of product i produced by time k. Consequently, the first constraint states that k units have to be produced during the first k periods, and the second constraint states that a product's cumulative production cannot decrease. The third constraint states that all demands should be satisfied after the final period. Obviously, the key to the formulation lies in the specification of the objective function $G(x,d)$, which measures the aggregate deviation from the ideal production rates. As in the case of the apportionment problem, however, a number of different possible measures have been proposed. One possibility is to minimize the total deviation ([39],[44]), which could be achieved using the objective function $G(x,r) = \sum_{a,k}(x_{a,k} - kr_a)^2$. Another possibility is to minimize the maximum deviation from the ideal production rates ([69]), using the objective function $G(x,r) = \max_{a,k}|x_{a,k} - kr_a|$. With either objective, the resulting optimization problem will be a non-linear integer program; in general, however, these can be solved efficiently (this will be discussed later).

A somewhat different approach to the PRV problem ([34]) worth mentioning is based on the notion of an ideal position (or due date) $p_{a,k}$ for the k-th unit of product type a, which is defined as $p_{a,k} = (k - \frac{1}{2})/r_a$. Under this approach the objective is to minimize the deviations between the ideal position $p_{a,k}$ and the actual position $t_{a,k}$ at which the the k-th unit of a is produced. This can be done (see [34]) by applying the earliest due date rule, using $p_{a,k}$ as the due dates. It is interesting to note that this method is actually equivalent to Webster's method of apportionment (see [13]).

5.1.3 Approach

The apportionment problem and in particular the PRV problem are closely related to the allocation of slots during a GDP. The primary difference, however, is that during a GDP not all flights are present initially; in other words, the demands associated with each airline may have release times. This is illustrated

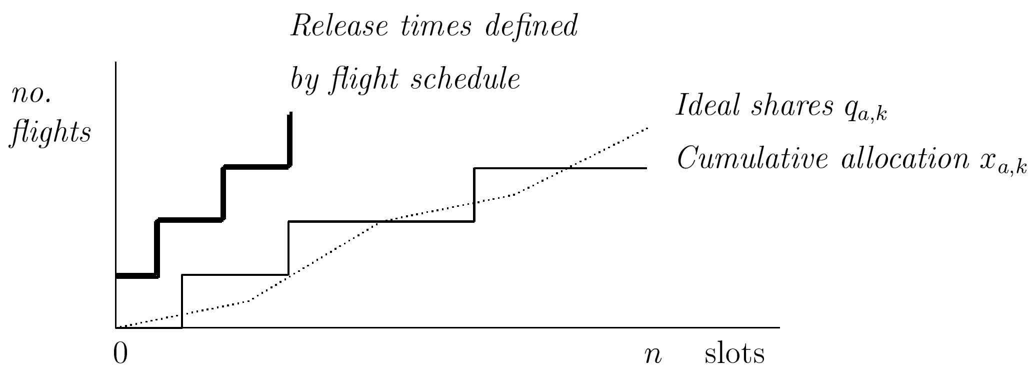


Figure 5.3: Share deviation in GDPs

in Figure 5.3, where the thick line may be interpreted as the cumulative demand of an airline at each time instance.

As a result, the proportional rates/shares used in the apportionment and PRV problems may no longer be applicable. This, however, may be addressed using approaches based on the results from the previous chapter, where we derived fair slot shares under the assumption that slots were divisible (e.g. by allowing random assignments). This corresponds to the definition of quota in the apportionment and PRV problems, which were similarly based on relaxing the indivisibility assumption. Thus, given an allocation rule X (e.g. corresponding to RBS or the proportional random assignment), we could define quota as $q_{a,k} = \sum_{j=0}^{k} X_{a,j}$.

At first sight, the approximation of fair shares therefore appears to be applicable primarily if proportional random assignment is used to define fair shares, since RBS already yields an assignment that (nearly) corresponds to the fair shares. As stated before however, an additional reason for studying this model is the dynamic nature of GDPs: during the course of a GDP, flights may be cancelled, delayed, etc. These changes impose additional constraints on the possible allocations, and make achievement of ideal shares impossible even under RBS.

5.2 Managing Flight Cancellations and Delays

During the course of a GDP flights are frequently cancelled and/or delayed, leading to suboptimal utilization of the airport's arrival capacity. This section describes an approach that deals with flight cancellations and delays based on the idea of approximating fair airline shares. As such, the procedure we introduce may be viewed as an alternative to the Compression Algorithm that is currently used under CDM. However, whereas Compression is based on the notion of an inter-airline slot exchange, the procedure discussed here may simply be viewed as a form of rerationing. Consequently, this procedure unifies both RBS and Compression, leading to a single resource allocation mechanism to be used during GDPs.

The general concept assumes we have defined fair shares $q_{a,k}$ for each airline that are independent of the actual allocation and remain *constant* for the duration of the program. In the remainder of this Chapter, we assume these shares are obtained using the first-scheduled, first-served principle (as in RBS). Whereas fair shares remain constant, dynamic changes may occur to each airline's input data. An airline's input data can be represented by (1) the set of cancelled flights F^C and (2) for each flight f an earliest arrival time e_f. We note that e_f in general can be equal to or later than the flight's oag time, due to upstream delays, crew or mechanical problems, etc. Together, these parameters

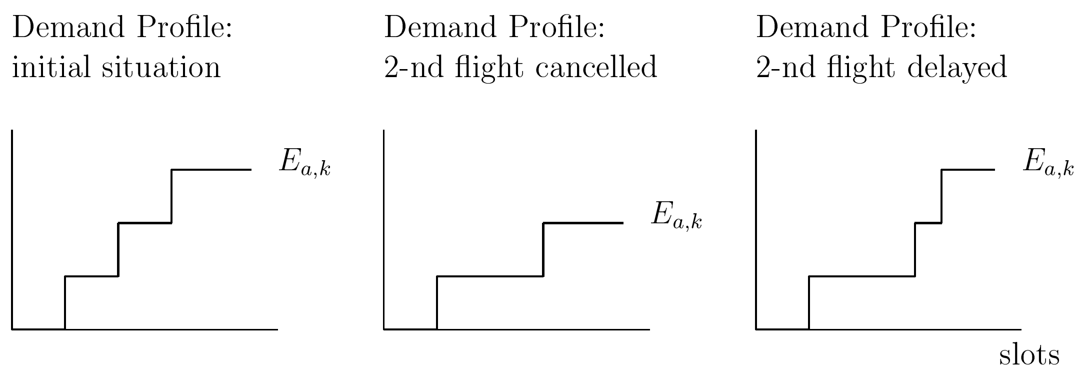


Figure 5.4: Demand changes from flight cancellations and delays

can be used to define each airline's cumulative demand (or release time) profile $\{E_{a,k}\}_{k=1}^{n_a}$, where

$$E_{a,k} = |\{f \in F_a/F^C : e_f \leq k\}|.$$

Thus, any change in an airline's input data can be interpreted as a change to its cumulative demand profile. This is illustrated in Figure 5.4, which shows the effects of flight cancellations and flight delays.

Generally speaking, the GDP process would operate as follows. The FAA continuously monitors airline updates and adjusts each airline's demand profile accordingly. If, based on these demand profiles, the current allocation were infeasible or suboptimal, the rationing procedure described in the next Section would be executed. The procedure would also be used to initiate the GDP (this would be analogous to first executing RBS followed by Compression, which is currently common practice).

IP FORMULATION:

$$\text{Min} \quad \textstyle\sum_{a,j}(x_{a,j} - q_{a,j})^2 \quad \text{or} \quad \max_{a,j} |x_{a,j} - q_{a,j}|$$

subject to:

$$\begin{array}{ll}
\sum_{a \in \mathcal{A}} s_{a,j} = c_j & \text{for all } j \in 0, \ldots, n-1 \\
x_{a,0} = s_{a,0} & \text{for all } a \in \mathcal{A} \\
x_{a,j+1} = s_{a,j+1} + x_{a,j} & \text{for all } a \in \mathcal{A}, j \in 0, \ldots, n-1 \\
x_{a,j} \leq E_{a,j} & \text{for all } a \in \mathcal{A}, j \in 0, \ldots, n-1 \\
x_{a,j}, s_{a,j} \geq 0 (integer) &
\end{array}$$

Figure 5.5: IP formulation of the slot allocation problem

5.2.1 Model Formulation

The resulting allocation problem is similar to the PRV problem, with the only added complication coming from the bounds imposed by the cumulative demand profiles. As such, we can use any approaches for the PRV problem to allocate slots during a GDP. One possibility is to minimize the total or maximum deviation. This leads to the formulation shown in Figure 5.5, where the cumulative demand profile bounds are incorporated by additional constraints.

Here, $q_{a,j}$ and $E_{a,j}$ are as defined before, and c represents the capacity vector. Observe that, as in Chapter 4, we assume that the capacities are such that all available slots will be used (i.e. the first constraint is posed as an equality). The variable $s_{a,j}$ equals 1 if slot j is assigned to airline a, and 0 otherwise. Again, $x_{a,j}$ represents the cumulative number of slots assigned to a by time j.

With either objective, the formulation in Figure 5.5 results in a non-linear integer program. Both cases, however, can be solved efficiently. With the total deviation objective function, the resulting model can be reformulated as a network flow problem (for a similar approach, see [38], [39]). To illustrate this, we redefine $x_{a,j}$ as

$$x_{a,j} = q_{a,j} - \sum_{l=0}^{q_{a,j}-1} u_{a,j,l} + \sum_{q_{a,j}+1}^{E_{a,j}} o_{a,j,l},$$

where $0 \leq u_{a,j,l}, o_{a,j,l} \leq 1$. Under this formulation $u_{a,j,l} = 1$ iff at most l slots are assigned to a by time j, and $o_{a,j,l} = 1$ iff at least l slots are assigned to a by time j. Observe that if $x_{a,j}$ is substituted out, this reformulation preserves the underlying network structure of the constraints. A linear objective function can now be obtained by introducing appropriate coefficients for the variables $u_{a,j,l}$ and $o_{a,j,l}$. With each variable $u_{a,j,l}$, we associate a coefficient $v_{a,j,l}$, which is defined as

$$v_{a,j,l} = (l - k_{a,j})^2 - (l + 1 - k_{a,j})^2.$$

Observe that $v_{a,j,l} > 0$, and that $v_{a,j,l-1} > v_{a,j,l}$. With each variable $o_{a,j,l}$, we associate a coefficient $w_{a,j,l}$, which is defined as

$$w_{a,j,l} = (l - k_{a,j})^2 - (l - 1 - k_{a,j})^2.$$

Again, $v_{a,j,l} > 0$, and $v_{a,j,l+1} > v_{a,j,l}$. The resulting objective function will therefore be

$$G(x, d) = \sum_{a,j} \sum_{l=0}^{k_{a,j}-1} v_{a,j,l} u_{a,j,l} + \sum_{a,j} \sum_{k_{a,j}+1}^{Q_{a,j}} w_{a,j,l} o_{a,j,l},$$

and it is easy to see that the resulting network flow problem yields optimal solutions that are also optimal for the original problem.

In case of the maximum deviation objective, the problem can be solved by a sequence of network flow problems (see also [69]). To illustrate this, let us consider the question of deciding whether there exists a solution x to the constraints in Figure 5.5 for which

$$G(x,i) = \max_{a,j} |x_{a,j} - q_{a,j}| \leq B.$$

Of course, a solution x will only satisfy this condition iff

$$\max_{a,j} |x_{a,j} - q_{a,j}| \leq B \quad \text{for all } a \in \mathcal{A}, j \in 0, \ldots, n-1.$$

This, however, is equivalent to the conditions

$$x_{a,j} \leq \lfloor q_{a,j} + B \rfloor, \quad x_{a,j} \geq \lceil q_{a,j} - B \rceil \quad \text{for all } a \in \mathcal{A}, j \in 0, \ldots, n-1. \tag{5.1}$$

Consequently, a solution for which $G(x,i) \leq B$ exists iff it satisfies both the constraints in Figure 5.5 and the constraints in 5.1. Thus, the decision problem reduces to the problem of finding a feasible flow, which can be done efficiently. In fact, the special structure of the constraint set can be exploited to achieve a highly efficient procedure (see [69]). Given such a procedure, the overall problem can be solved by performing a bisection search procedure over B.

Another possibility is to use an approach that is based on the notion of an ideal position for the k-th flight of each airline a (that is, the approach discussed in [34]). In this case, the objective is to minimize the deviation between an (appropriately defined) ideal position $p_{a,k}$ for the k-th flight of airline a and the actual position of a's k-th flight. If the underlying fair shares are based on the RBS procedure, the ideal positions $p_{a,k}$ can be defined as follows

$$p_{a,k} = \min_{j \geq 0 : q_{a,j} \geq k} j,$$

that is, the ideal position for airline a's k-th flight corresponds to its k-th slot in the RBS schedule. Note that by definition of the RBS shares, each slot defines an ideal position for a single airline. Given the current demand profiles $E_{a,k}$ for

IP FORMULATION:

$$\text{Min} \quad \sum_{a,k,j}(j - p_{a,k})^2 x_{a,k,j}$$

subject to:

$$\sum_{a\in\mathcal{A},k:e_{a,k}\leq j} x_{a,k,j} = c_j \qquad \text{for all } j \in 0,\ldots,n-1$$

$$\sum_{j:e_{a,j}\leq j\leq n} x_{a,k,j} = 1 \qquad \text{for all } a \in \mathcal{A}, k \in 1,\ldots,E_{a,n}$$

$$x_{a,k,j} \geq 0 (integer)$$

Figure 5.6: Alternative IP formulation of slot allocation problem

each airline, we can furthermore define the earliest arrival time $e_{a,k}$ of the k-th flight as

$$e_{a,k} = \min_{j\geq 0:E_{a,j}\geq k} j.$$

The resulting IP formulation is shown in Figure 5.6, where $x_{a,k,j} = 1$ if a's k-th flight is assigned to slot j, and 0 otherwise.

The formulation in Figure 5.6 corresponds to an assignment problem, and can therefore be solved efficiently. However, a simpler procedure (similar to the earliest due date algorithm) that finds optimal solutions exists, and is shown in Figure 5.7. In this procedure, each airline is assigned a set of (remaining) priorities corresponding to its ideal positions. The procedure repeatedly assigns the next available slot to the airline which has the highest remaining priority among all airlines that can use the slot. The correctness of the procedure is insured by the following theorem.

Theorem 5.2.1. *A solution x obtained by the greedy algorithm shown in Figure 5.7 is an optimal solution for the IP formulation shown in Figure 5.6.*

Proof. See Appendix. □

Init :

Let $P_a := \bigcup_{k=1}^{E_{a,n}} \{p_{a,k}\}$ for all $a \in \mathcal{A}$

Let $x_{a,k,j} := 0, k_a := 1$ for all $a \in \mathcal{A}, j \in 0, \ldots, n-1$

For $j \in 0, \ldots, n-1 : c_j = 1$ **Do**

Let $A' := \{a \in \mathcal{A} : \sum_{k=1}^{j-1} \sum_{l=1}^{j-1} x_{a,k,l} < E_{a,j}\}$

Let $p_a := \min_{p \in P_a} p$

Let $a' := \arg\min_{a \in A'} p_a$

Let $x_{a',k_a,j} := 1, P_{a'} := P_{a'} - \{p_{a'}\}, k_a := k_a + 1$

Od

Figure 5.7: Greedy Algorithm for slot allocation problem

5.2.2 Comparison

The "total deviation" model and the "ideal position" model define two possible approaches to the management of flights cancellations and delays during a GDP. These procedures could be executed periodically, whenever changes in airline demand profiles (due to cancellations and delays) cause the current schedule to be infeasible or suboptimal. This section illustrates their differences, and compares their resulting allocations with the Compression procedure.

Total Deviation vs. Ideal Position

Intuitively, the difference between the total deviation approach and the model using ideal positions can be illustrated by the example shown in Figure 5.8. The example shows an initial schedule in which all flights are assigned their ideal position, but where the first three flights are subsequently delayed. Under the total (or maximum) deviation model, slot 3 would be assigned to airline b, Initial Assignment:

$f_{a,1}$	$f_{b,1}$	$f_{b,2}$	$f_{c,1}$	$f_{c,2}$	$f_{c,3}$
0	1	2	3	4	5

Demand profile changes:

$$e_{a,1}, e_{b,1}, e_{b,2} = 3$$

Total Deviation Model Assignment:

$f_{c,1}$	$f_{c,2}$	$f_{c,3}$	$f_{b,1}$	$f_{a,1}$	$f_{b,2}$
0	1	2	3	4	5

Ideal Position Model Assignment:

$f_{c,1}$	$f_{c,2}$	$f_{c,3}$	$f_{a,1}$	$f_{b,1}$	$f_{b,2}$
0	1	2	3	4	5

Figure 5.8: Comparison: Ideal Position vs. Total Deviation

whereas the ideal position model would allocate slot 3 to airline a. Intuitively, the reason is that the total deviation model favors the airline with the highest number of flights that can use the slot, while the ideal position model favors the airline with the earliest flight that can use the slot.

A more general difference between the methods stems from the so-called monotonicity condition. The monotonicity condition states that if an allocation x is optimal with respect to the first k slots, there exists a solution y that is optimal with respect to the first $k+1$ slots such that $y \geq x$. It can be shown (see [13]) that the total and maximum deviation models need not the satisfy the monotonicity condition; the model based on ideal positions, however, is monotone by construction.

Overall, it appears therefore that the model based on ideal positions may be a more applicable approach within the context of allocating slots. Allocations can be obtained using a basic greedy procedure, and the general approach is analogous to the priorities that were the basis for determining the fair shares.

Relationship to Compression Procedure

The total deviation model and the ideal position model provide alternatives to the Compressionprocedure. The greedy procedure associated with the ideal position model, in particular, is closely related to the Compressionalgorithm: both procedures repeatedly assign slots according to a priority ordering. Yet in spite of their similarities, there are also a number of key differences. The first difference is in the basis for the priorities (ideal positions). Under compression, these are based on the current assignment, whereas under the greedy procedure they remain constant (based on the original schedule) throughout the duration of the GDP. Another difference is the order in which slots are assigned to flights: the greedy procedure assign the slots in sequence, while the Compressionalgorithm repeatedly assigns the slot that has been vacated by a flight movement. The impact of this difference is illustrated in the example in Figure 5.9.

The example starts with an initial schedule in which all flights are assigned their ideal position, but where the first two flights have been cancelled. The Compression Algorithm will first move up flight $f_{c,1}$ to time 0 and flight $f_{a,2}$ to time 2. Subsequently, flight $f_{b,2}$ will be moved up to time 1 and flight $f_{d,1}$ to time 3. The greedy procedure, on the other hand, will allocate the four slots sequentially, leading to a different assignment. A final difference between these procedures is that the Compressionalgorithm only moves up flights (that is, it insures that flights do not lose their current positions). In contrast, the greedy

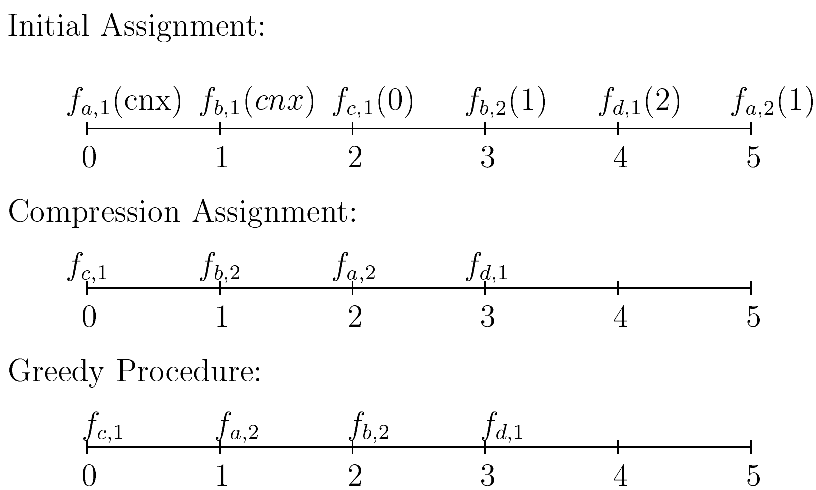


Figure 5.9: Comparison: Compression vs. Greedy Procedure

procedure does not explicitly take the current assignment into account. This, however, is only an apparent difference, caused by peculiarities in the implementation of the Compressionalgorithm. Specifically, in cases where a flight is delayed and the slot it has been assigned to cannot be used, the Compression-algorithm may maintain an infeasible solution by creating a new "slot" for the delayed flight.

We compared the greedy procedure with the Compression Algorithm using four scenarios derived from real-world GDPs. Three of the data sets that were used represented GDPs at Newark International Airport (EWR), while one of the data sets considered a GDP at Los Angeles International Airport (LAX). The data gathered for each Compressionscenario consisted of the flights and slots in the GDP, the initial assignment of flights to slots, the earliest arrival times for each flight, and the set of flights that were cancelled. The four scenarios are summarized in Table 5.1.

Table 5.1: Problem Characteristics

	EWR 01/01/96(1)	EWR 01/01/96(2)	EWR 01/02/96	LAX 01/01/97
Number of Flights	73	94	54	62
Number of Cancellations	12	21	6	10

For each of the scenarios we ran both the greedy procedure and the Compression Algorithm. The results are shown in Tables 5.2 through 5.5. The tables show for both procedures the absolute and the relative delay savings for each airline (delay savings are measured in minutes). In addition, the tables show for each airline the *baseline savings*, that is, the reduction in delay each airline would have been able to achieve by itself. Baseline savings provide a convenient basis for comparison of delay reduction on an airline-by-airline basis.

Table 5.2: Delay reduction for Scenario EWR, 01/01/96(1)

Airlines	Comp Absolute	Comp Relative	Opt Absolute	Opt Relative	Baseline Absolute	Baseline Relative
COA	402	46.53	406	46.99	281	57.00
UAL	200	23.15	195	22.57	142	28.80
TWA	17	1.97	17	1.97	0	0.0
AAL	123	14.24	126	14.58	70	14.20
ACA	2	0.23	0	0.00	0	0.0
USA	38	4.40	38	4.40	0	0.0
BSK	2	0.23	2	0.23	0	0.0
NWA	19	2.20	19	2.20	0	0.0
AWE	14	1.62	14	1.62	0	0.0
DAL	19	2.20	19	2.20	0	0.0
KMR	3	0.35	3	0.35	0	0.0
CAA	0	0.0	0	0.00	0	0.0
LOT	2	0.23	2	0.23	0	0.0
SJI	10	1.16	10	1.16	0	0.0
COM	13	1.50	13	1.50	0	0.0
TOTAL	864	100.00	864	100.00	0	0.0

Table 5.3: Delay reduction for Scenario EWR, 01/01/96(2)

Airlines	Comp Absolute	Comp Relative	Opt Absolute	Opt Relative	Baseline Absolute	Baseline Relative
FDX	0	0.0	0	0.00	0	0.0
COA	521	50.63	524	50.92	420	75.0
NWA	79	7.68	77	7.48	0	0.0
ACA	4	0.39	4	0.39	0	0.0
UAL	171	16.62	168	16.33	68	12.14
AAL	81	7.87	81	7.87	72	12.86
USA	84	8.16	87	8.45	0	0.0
DAL	29	2.82	29	2.82	0	0.0
DLH	0	0.0	0	0.00	0	0.0
TWA	2	0.19	2	0.19	0	0.0
BSK	6	0.58	5	0.49	0	0.0
AWE	6	0.58	6	0.58	0	0.0
BAW	6	0.58	6	0.58	0	0.0
KMR	16	1.55	16	1.55	0	0.0
LOT	24	2.33	24	2.33	0	0.0
TOTAL	1029	100.00	1029	100.00	560	100.00

Table 5.4: Delay reduction for Scenario EWR, 01/02/96

Airlines	Comp Absolute	Comp Relative	Opt Absolute	Opt Relative	Baseline Absolute	Baseline Relative
COA	231	64.71	270	75.63	167	85.20
ACA	40	11.20	10	2.80	0	0.0
SJI	3	0.84	3	0.84	0	0.0
COM	2	0.56	2	0.56	0	0.0
N4I	2	0.56	2	0.56	0	0.0
UAL	60	16.81	60	16.81	29	14.80
MXA	2	0.56	0	0.00	0	0.0
NWA	5	1.40	0	0.00	0	0.0
VIR	3	0.84	3	0.84	0	0.0
TWA	3	0.84	3	0.84	0	0.0
PAL	2	0.56	0	0.00	0	0.0
AJM	1	0.28	1	0.28	0	0.0
USA	1	0.28	1	0.28	0	0.0
AAL	1	0.28	1	0.28	0	0.0
CAA	1	0.28	1	0.28	0	0.0
TOTAL	357	100.00	357	100.00	196	100.0

Table 5.5: Delay reduction for Scenario LAX, 01/01/97

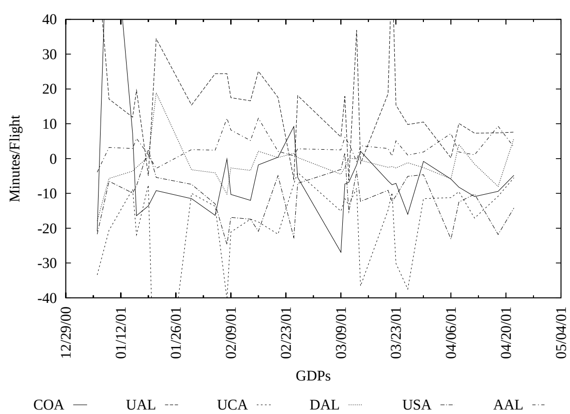
Airlines	Comp Absolute	Comp Relative	Opt Absolute	Opt Relative	Baseline Absolute	Baseline Relative
UAL	153	42.62	142	39.55	127	53.59
AAL	72	20.06	66	18.38	70	29.54
SWA	25	6.96	32	8.91	18	7.59
TWA	38	10.58	38	10.58	0	0.0
ASA	6	1.67	6	1.67	0	0.0
SER	0	0.0	0	0.00	0	0.0
DAL	8	2.23	8	2.23	0	0.0
FDX	4	1.11	4	1.11	0	0.0
RKT	2	0.56	4	1.11	0	0.0
ROA	9	2.51	9	2.51	0	0.0
AMX	2	0.56	8	2.23	0	0.0
ANZ	2	0.56	4	1.11	0	0.0
AWE	0	0.0	0	0.00	0	0.0
USA	24	6.69	24	6.69	22	9.28
COA	2	0.56	2	0.56	0	0.0
NWA	6	1.67	6	1.67	0	0.0
FFT	6	1.67	6	1.67	0	0.0
TOTAL	359	100.00	359	100.00	237	100.00

An airline will always achieve this amount of delay savings, and the fact that more total savings are possible is exactly due to inter-airline reallocation of slots. The results in Tables 5.2 through 5.5 indicate that the greedy procedure results in flight-slot assignments that are very similar to those obtained by the Compression Algorithm.

5.3 Managing Flight Exemptions

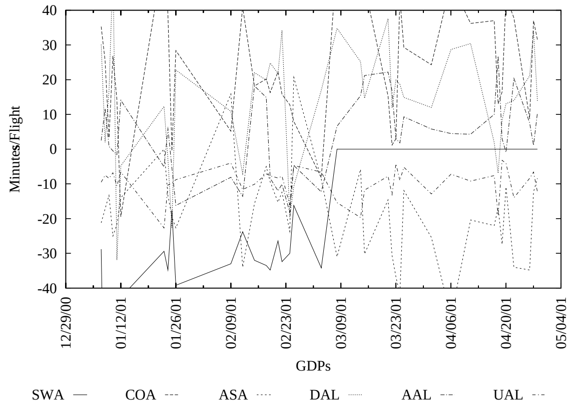
In addition to flight cancellations and delays, the numerous flight exemptions that may occur during actual GDPs may also have a significant impact on the allocation of slots. Flights may be exempted during a GDP for various reasons: a flight may have departed already, in which case it cannot be assigned ground delay, and in some cases flights from certain departure airports (or centers) are explicitly exempted. Typically, this is done for long-haul flights, so as to prevent potentially unrecoverable delays that might be caused by the uncertainty in the weather predictions. Currently, flight exemptions are managed on a somewhat ad-hoc basis: exempted flights are assigned slots first, followed by the allocation of the remaining slots to the non-exempted flights. The manner in which exemptions are managed, however, can have a significant impact on the distribution of delays among airlines. To illustrate this, we analyzed the impact of flight exemptions using historical data. For a number of GDPs, we determined the delays for each airline with and without the exemptions that occurred during that day. The results are shown in Figures 5.10, 5.11, and 5.12.

The graphs in these Figures represent, for a selected number of airlines, the difference between an airline's average delay under RBS without exemptions and under RBS with exemptions included (a negative number means the airline



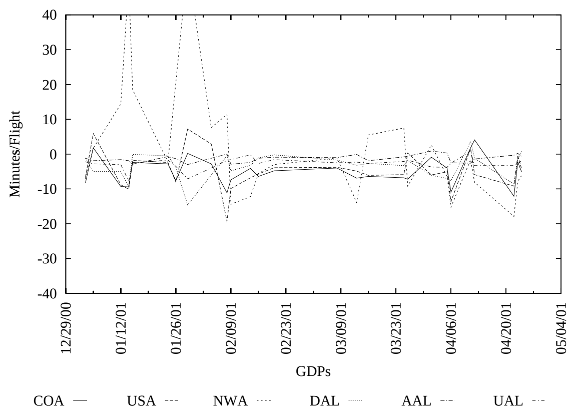
flts/gdp	No Exempt delay/flt	Actual delay/flt
COA(10.5)	74.4	79.2
UAL(18.4)	73.6	58.2
UCA(19.3)	69.9	88.2
DAL(47.3)	71	72.8
USA(66)	71.7	83
AAL(66.7)	71.7	68.3

Figure 5.10: Exemption Impact : Logan Airport, Boston



flts/gdp	No exempt delay/flt	Actual delay/flt
SWA(4.3)	79.2	112.8
COA(4.8)	60.9	38
ASA(7.3)	55.8	37
DAL(8.2)	55.5	42.5
AAL(13.4)	63.9	61.7
UAL(118)	58.2	67.6

Figure 5.11: Exemption Impact: San Francisco Airport



flts/gdp	No exempt delay/flt	RBS delay/flt
COA(9.4)	40.1	45.3
USA(10)	42.1	47
NWA(11.6)	39.6	37.9
DAL(12)	43.7	47.3
AAL(224.6)	41.4	42.7
UAL(250.4)	41.5	44.2

Figure 5.12: Exemption Impact: O'Hare Airport, Chicago

would have been allocated less delay if exemptions were not taken into account). The adjacent table in each Figure shows the average delay per flight for each airline, aggregated over all GDPs.

The results clearly show that exemptions may have a significant impact on the distribution of delays. Moreover, they illustrate that exemptions may introduce a systematic bias in favor or against certain airlines. At Boston's Logan airport, for example, USA and UCA (a small commuter airline) appear to have a systematic disadvantage; the reason for this likely is that most of their flights are short-haul (these flights are rarely exempted). Similar results also hold for San Francisco airport. At Chicago O'Hare airport, the differences are less pronounced; here, however, it appears that the larger airlines are at a systematic disadvantage. In the remainder of this section we propose allocation methods that incorporate flight exemptions, and analyze their impact on the overall distributions of the delays among airlines.

5.3.1 Model Formulation

To incorporate exempted flights, we assume as before that the capacities c are given, and that the quota $q_{a,k}$ and demand profiles $E_{a,k}$ are known. In addition, however, we now have a set $F^e \subseteq F$ of exempted flights. Each of these flights has a current estimated time of arrival eta_f. Since exempted flights may be airborne, each flight will have to be assigned to the slot corresponding to its estimated time of arrival. In the remainder of this Section we assume that this is always possible, that is, $|\{f \in F^e : eta_f = j\}| \leq 1$ for all periods j.

Under current procedures, exempted flights are assigned first, and RBS is used to assign the remaining flights to the remaining slots. Here, we take a different approach: we minimize, as before, the overall deviation from the ideal airline shares, but take into account exemptions by imposing the additional constraints that each airline is assigned the slots corresponding to its exempted flights. As such, exempted flights no longer have a strict priority. For the total deviation model, this can be done by adding the additional constraints

$$s_{a,eta_f} = 1 \quad \text{for all } a \in \mathcal{A},\ f \in \mathcal{F}_a \cap F^e$$

to the formulation shown in Figure 5.5. We note that this will preserve the network structure of the resulting formulation, since we only fix the values of certain variables.

For the approach based on ideal positions, shown in Figure 5.6, exempted flights may be incorporated by adding the constraints

$$\sum_{k=0}^{E^{a,n}} x_{a,k,eta_f} = 1 \quad \text{for all } a \in \mathcal{A},\ f \in \mathcal{F}_a \cap F^e. \tag{5.2}$$

This constraint states that one of airline a's flights (i.e. its first flight, second flight, etc.) should be assigned to the slots corresponding its exempted flights.

It is important to note, however, that with the added constraints the greedy procedure used in the previous Section to solve the ideal position model (or a variant thereof) does not necessarily give an optimal solution. Nevertheless, the use of a greedy procedure has an intuitive appeal within the context of GDPs, and in the remainder of this section we outline a possible allocation procedure that accounts for exempted flights.

To motivate this procedure, we first consider a model in which the constraints imposed by the exempted flights are relaxed. This would allow us to use the same approach as in the previous Section: where the cumulative demand profiles imposed upper bounds on the slots assigned to an airline by a period, flight exemptions would impose lower bounds. Formally, we define these lower bounds as

$$L_{a,j} = |\{f \in F_a \cap F^e : eta_f \leq j\}| \quad \text{for all } a \in \mathcal{A},\ j \in 0, \ldots, n-1.$$

In other words, $L_{a,j}$ represents the number of flights that should have been assigned to a by time j. It is fairly straightforward to incorporate these lower bounds into the models proposed in the previous Section. For the approach based

Init :

Let $P_a := \bigcup_{k=1}^{E_{a,n}} \{p_{a,k}\}$ for all $a \in \mathcal{A}$

Let $x_{a,k,j} := 0, k_a := 1$ for all $a \in \mathcal{A}, j \in 0, \ldots, n-1$

For $j \in 0, \ldots, n-1$ **Do**

Let $A' := \{a \in \mathcal{A} : \sum_{k=1}^{j-1} s_{a,j} < E_{a,j}\}$

Let $p_a := \min_{p \in P_a} p$

if $k'_a < L_{a',j}$ for some $a' \in \mathcal{A}$ **then**

Let $x_{a',k_a,j} := 1, P_{a'} := P_{a'} - p_{a'}, k_a := k_a + 1$

else

Let $a' := \arg\min_{a \in A'} p_a$

Let $x_{a',k_a,j} := 1, P_{a'} := P_{a'} - p_{a'}, k_a := k_a + 1$

Od

Figure 5.13: Modified Greedy Algorithm

on ideal positions, which is shown in Figure 5.6, we would add the constraints

$$x_{a,k,j} = 0 \quad \text{for all } a, k, j \text{ such that } L_{a,j} > k. \tag{5.3}$$

The resulting optimization model can be solved using the modified greedy procedure shown in Figure 5.13. Intuitively, the greedy procedure proceeds as before, except when a flight is due: in that case, the slot is assigned to the corresponding airline. The correctness of the procedure is shown by the following theorem.

Theorem 5.3.1. *A solution x obtained by the greedy algorithm shown in Figure 5.7 is an optimal solution for the IP formulation shown in Figure 5.6 with constraints 5.3 added.*

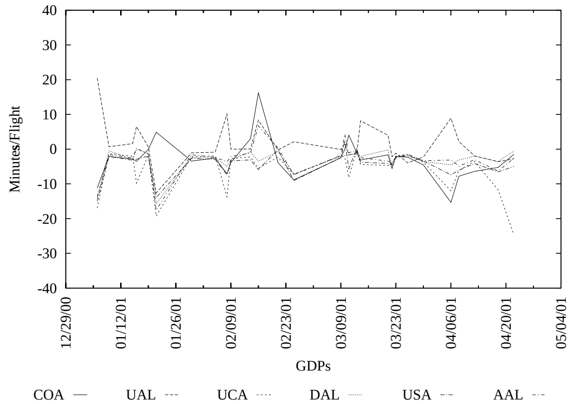
Proof. See Appendix. □

Given that we relaxed the constraints imposed by the exempted flights, the use of lower bounds may not always yield feasible solutions. To illustrate this, consider an airline that has two flights f_1 and f_2 with associated earliest arrival times $e_1 = 0$ and $e_2 = 4$. Flight f_2 has been exempted and has $eta_2 = 5$. Thus, we have $E_{a,j} = 1$ if $j < 4$ and $E_{a,j} = 2$ otherwise. Similarly, we have $L_{a,j} = 0$ if $j < 5$ and $L_{a,j} = 1$ otherwise. According to these bounds, it would be possible that a was assigned its first flight at time 3 and its second flight at time 6, which may not be feasible for flight f_2. In other words, the lower bounds become flight dependent. Nevertheless, the procedure indicates a potential alternative approach: we proceed with the greedy procedure as before, except when an exempted flight needs to be assigned (as opposed to a flight that is due). In that case, the slot is assigned to the corresponding airline.

5.3.2 Comparison

The empirical analysis at the start of this section indicated that the current procedures for managing exemptions may introduce systematic biases against some carriers. Here, we analyze the extent to which the optimization models are able to mitigate these biases. Moreover, we study how these optimization models impact the distribution of delays within an airline, as well as their impact on the delays by aircraft size.

First, we compared the delay obtained under RBS and the delay that would have been obtained using the optimization model based on ideal positions (with the constraints in equation 5.2 incorporated). The results for Logan airport are shown in Figure 5.14. Figure 5.14 represents, for a selected number of airlines, the difference between an airline's average delay under RBS without exemptions

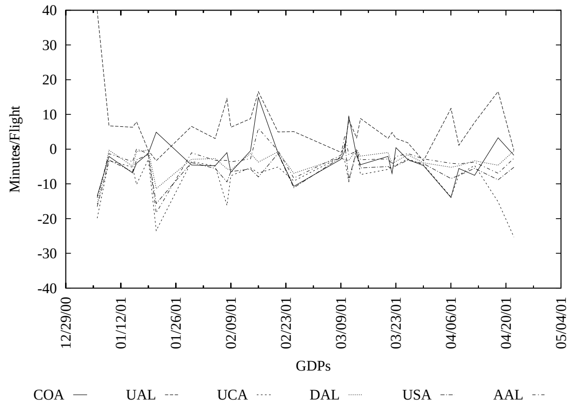


flts/gdp	Opt diff/flt	Actual diff/flt
COA(10.5)	-2.8	-6.1
UAL(18.4)	1.2	19.4
UCA(19.3)	-5.4	-18.3
DAL(47.3)	-3.3	-1.3
USA(66)	-3.4	-11.7
AAL(66.7)	-2.9	4.4

Figure 5.14: Optimization Model results : Logan Airport, Boston

and under the optimization model (a negative number means the airline would have been allocated less delay if exemptions were not taken into account). It is instructive to compare Figure 5.14 with Figure 5.10, which shows the difference in delay between RBS and the current procedures. Clearly, the optimization model has a significant impact and is able to reduce the biases substantially. This is further illustrated in the table shown in Figure 5.14, which shows the differences in delay for both the current procedures and for the optimization model.

In addition, we also analyzed the delay changes that would have been obtained using the modified greedy procedure. While, as we discussed, this procedure will not necessarily achieve optimal solutions (i.e. minimize the deviation from optimal position), we still believe that the simple and intuitive nature of the procedure might make it a potentially attractive alternative. The table and graph in Figure 5.15 show the delay differences that would have been obtained using the greedy procedure as opposed to the optimization model, using the same data from GDPs at Logan airport. As expected, the results are not as



flts/gdp	Greedy diff/flt	Actual diff/flt
COA(10.5)	-2.9	-6.1
UAL(18.4)	6.1	19.4
UCA(19.3)	-7.4	-18.3
DAL(47.3)	-3.7	-1.3
USA(66)	-5.9	-11.7
AAL(66.7)	-3.2	4.4

Figure 5.15: Greedy Procedure results : Logan Airport, Boston

pronounced as those using the optimization model. Nevertheless, the greedy procedure still yields allocations that are substantially closer to the ideal RBS share than those that would have been obtained using the current procedures.

While these results clearly indicate the potential to mitigate the exemption bias, the use of optimization-based allocation methods can also have a significant impact on the the distribution of delays within an airline. Potentially, such changes in the delay distribution could have detrimental effects: if, for example, the resulting allocations would significantly increase the percentage of flights with excessive delays, it might be difficult for an airline to recover (part of) its schedule with flight cancellations and/or substitutions. The optimization model's impact on the distribution of delays is shown in Figure 5.16. Figure 5.16 depicts, for a selected number of airlines, both the distribution of delays that was obtained under RBS and the distribution of delays that would have been obtained by the optimization-based approach. To determine these distributions, we used the same GDPs at Boston's Logan airport as in the previous experiments. In each of the graphs, the solid lines represent the distribution of delays under RBS,

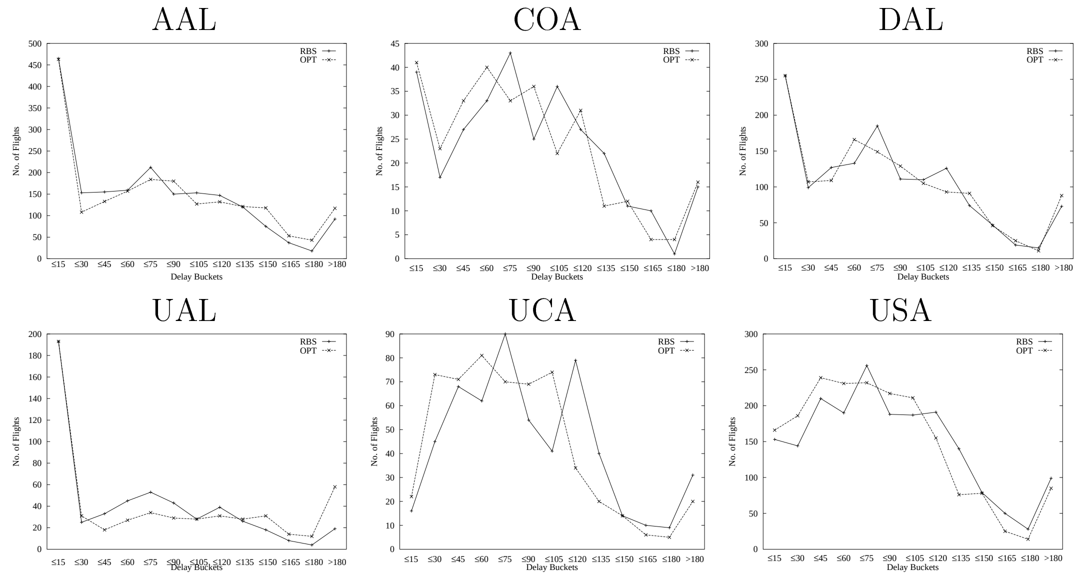


Figure 5.16: Delay Distribution Impact: Logan Airport, Boston

whereas the dashed lines represent the distribution of delays that would have been obtained using the optimization-based approach. The results in Figure 5.16 indicate that, on the aggregate, the use of the optimization model appears to have a relatively minor influence on the distribution of delays. The impact is most severe for United Airlines (UAL), which would see a sizeable increase in large delays ($\geq$ 180 minutes). Intuitively, the reason for this is that UAL has a high percentage of flight exemptions (approximately 60% of its flights are exempt), while it has a relatively small number of flights in a GDP (approximately 18 flights per GDP). As such, there will be little opportunity to shift the delay "benefits" absorbed by the exempt flights to its non-exempt flights. Further evidence of the robustness of the delay changes is found in Figure 5.17, which depicts the distribution of delay changes (that is, a flight's delay under the optimization model minus its delay under RBS) for the flights of the same set of airlines. The results are again aggregated over the same set of GDPs at Boston's

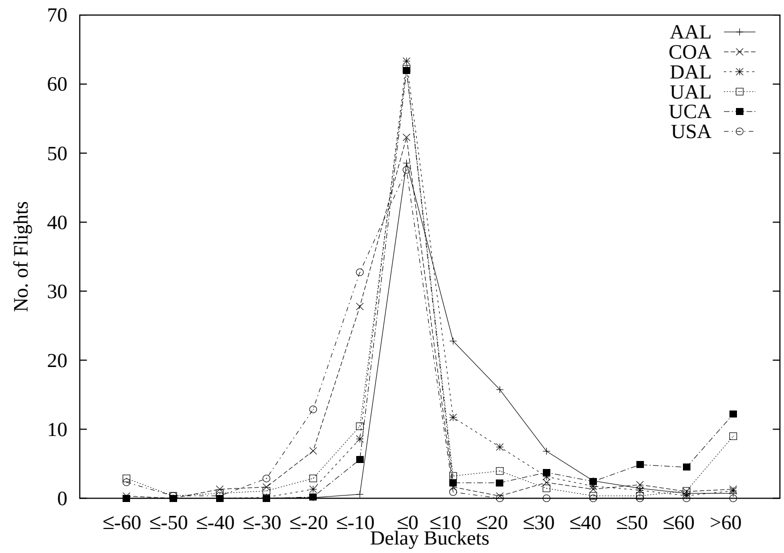


Figure 5.17: Distribution of Delay Changes by Airline: Logan Airport, Boston

Logan airport. In particular, Figure 5.17 illustrates that approximately 95% of the flights would receive a delay increase of at most 30 minutes if the optimization model was used.

Finally, we also analyzed the changes in delay over different classes of aircraft sizes. Figure 5.18 represents the flight delay changes for three FAA-designated classes of aircraft: Small, Large, and Heavy (this classification is based on an aircraft's pounds of wake vortex, as its primarily used for flight separation). On average, the optimization-based approach will reduce the delay of aircraft in the class Small by 9.1 minutes per flight, while delay of Large and Heavy aircraft increases by 1.6 resp. 0.5 minutes per flight. From a system-wide perspective this reduction in delays for smaller aircraft is less desirable; however, given that Small aircraft constitute a small fraction of all traffic(13% of all flights) the impact on the delay of Large and Heavy aircraft appears to be relatively low.

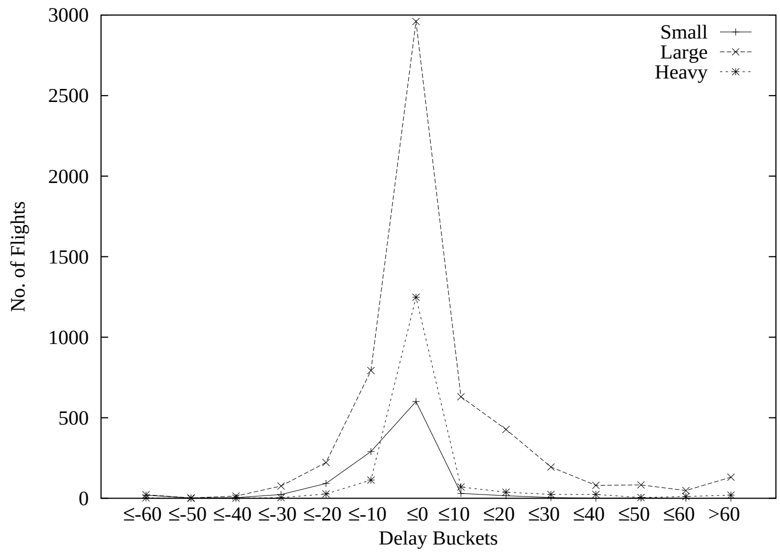


Figure 5.18: Distribution of Delay Changes by Aircraft Size: Logan Airport, Boston

5.4 Using Alternate Fairness Standards

So far, the models discussed in this chapter used the priority standard from RBS (i.e. first-scheduled, first-served) to determine the fair slot shares for each airline. The use of these models was therefore motivated by GDP dynamics that prevented airlines from realizing these shares. This section, however, briefly analyzes the impact of using alternative standards of fairness (and the fair shares resulting from these standards), based on the approaches discussed in Chapter 4. As stated before, the optimization models in this case could be used to prevent the unacceptable high levels of variance that may result from using a probabilistic allocation method.

This section considers two alternative standards of fairness, which use fair share definitions based on the proportional random assignment mechanism and on a more basic notion of airline proportionality. The proportional random assignment mechanism was introduced as an alternative to RBS in Chapter 4; whereas RBS is based on the notion of first-scheduled, first-served, the proportional random assignment mechanism used the concept of "equal access to usable slots" (that is, all flights that can use a slot have an equal entitlement to it). This concept of fairness could be achieved with the allocation scheme shown in Figure 4.6; consequently the expected share $X_{a,j}$ of slot j for airline a can be calculated recursively as

$$X_{a,j} = \frac{E_{a,j} - \sum_{k=0}^{j-1} X_{a,k}}{\sum_{a \in \mathcal{A}} E_{a,j} - \sum_{k=0}^{j-1} X_{a,k}}, \quad a \in \mathcal{A}, j = 0, \ldots, n-1,$$

where $E_{a,j}$ represents the cumulative demand from airline a up to period j (as defined in Section 5.2). In other words, the proportional random assignment mechanism will assign shares of slot j in proportion to each airline's current unsatisfied demand. The use of the proportional random assignment scheme as the basis of fairness would yield the following definition of quota:

$$q_{a,k} = \sum_{j=0}^{k} X_{a,j}, \quad k = 0, \ldots, n-1.$$

Given this definition of quota, we could in principle use any of the optimization models proposed in this chapter.In this case, however, the definition of ideal positions (and therefore the use of the greedy procedure) would introduce several complications, since the rate of increase in the cumulative slot shares need not be constant. As a result, the allocations obtained by a greedy procedure would not necessarily yield the monotonicity property discussed before. For this reason, we have used the total deviation model in our experimental results.

In addition, we also considered the more basic fairness standard in which each flight has an equal entitlement to all available slots in a GDP; in other words, this standard does not take into account the scheduled arrival times of flights and each airline is entitled to a share of the arrival slots that is proportional to its flights. As such, it follows that the quota for this approach can be defined as

$$q_{a,k} = (k+1)\frac{|F_a|}{|\sum_{a' \in \mathcal{A}} F_{a'}|}, \quad a \in \mathcal{A}.$$

The idea that flights are equally entitled to all slots is in fact similar to the basic assumptions underlying the Shapley value, as was discussed in Chapter 4. In this case, we can again use the model based on ideal positions, which are defined as

$$p_{a,k} = n(k - \frac{1}{2})/|F_a|, \quad a \in \mathcal{A}, k = 1, \ldots, |F_a|.$$

which equivalent to the definition of ideal positions in the PRV problem. Subsequently, the allocation of slots with this standard of fairness can be achieved using the greedy procedure.

5.4.1 Empirical Results

This section analyzes the impact of using the alternate fairness standards described above. For both fairness standards, we compared the allocation that would have been obtained using the resulting optimization models with the allocations obtained by RBS. The analysis was performed for the same of GDPs at Boston's Logan airport as before.

The impact of using proportional random assignment to determine fair shares is shown in Figure 5.19. Figure 5.19 shows, for a selected number of airlines, both the distribution of delays that was obtained under RBS and the distribution of delays that would have been obtained using the optimization model with fair shares derived from proportional random assignment. In each of the graphs, the solid lines represent the distribution of delays under RBS, whereas the dashed

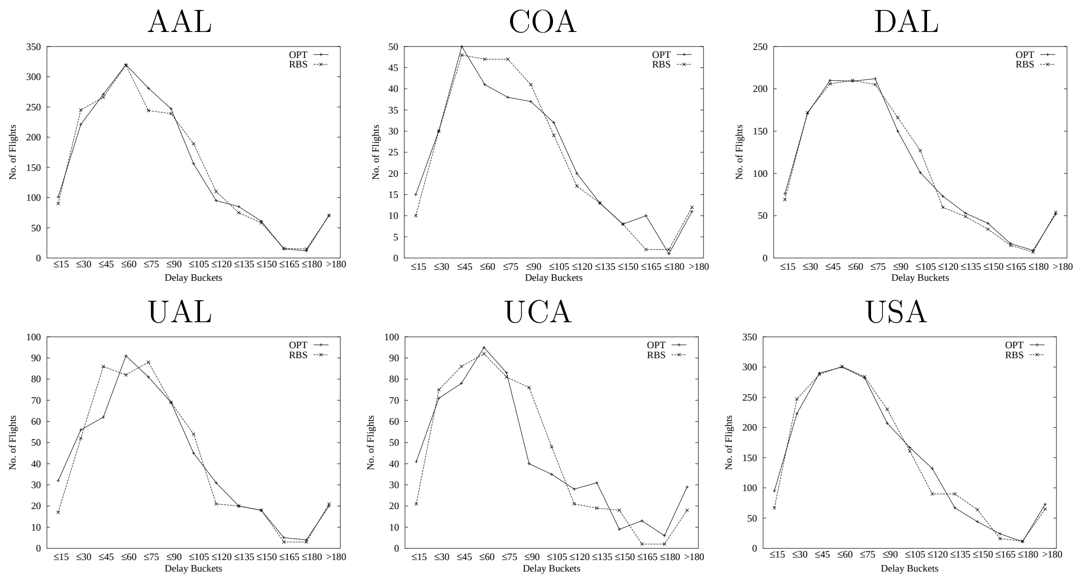


Figure 5.19: RBS vs. Proportional Random Assignment Approximation

lines represent the distribution of delays that would have been obtained using the optimization-based approach. The results in Figure 5.19 indicate that, on the aggregate, both approaches yield similar results, which is perhaps not surprising given the empirical comparison we performed in Chapter 4.

The impact of using the basic proportionality principle to determine fair shares is shown in Figure 5.20. Again, we determined both the distribution of delays that was obtained under RBS and the distribution of delays that would have been obtained using the optimization model with proportional shares. The graphs shown in Figure 5.20 show an interesting result, in that the use of proportional shares increases the on-time performance of flights (that is, flights with a delay of 15 minutes or less) in the resulting allocations. Overall, we found that RBS yields allocations in which approximately 3.8% of flights arrive on-time, while the optimization models would yield allocations in which on-time performance is approximately 15.7%. This increase in on-time performance however

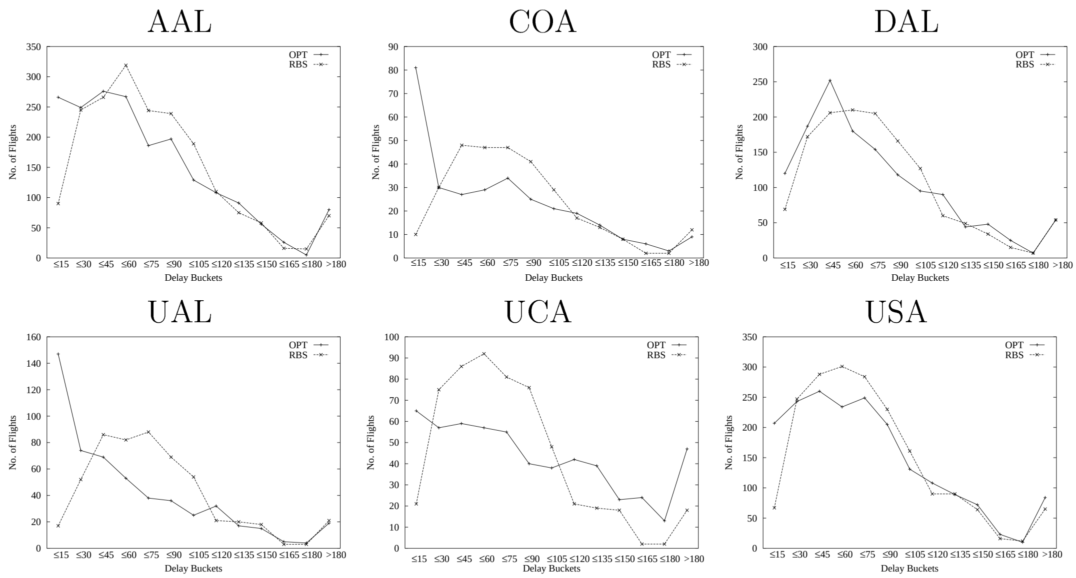


Figure 5.20: RBS vs. Proportional Approximation

is offset by an increase in the flights that have delays of 2 hours or more. This is illustrated in Figure 5.21, which compares the distributions of delays over all the flights for the same set of GDPs. Observe that for all 15-minute intervals with delays of 2 hours or more, the optimization model will have a higher number of flights than RBS. It is interesting to note that the increase in on-time performance appears to have come primarily at the expense of General Aviation (GA)

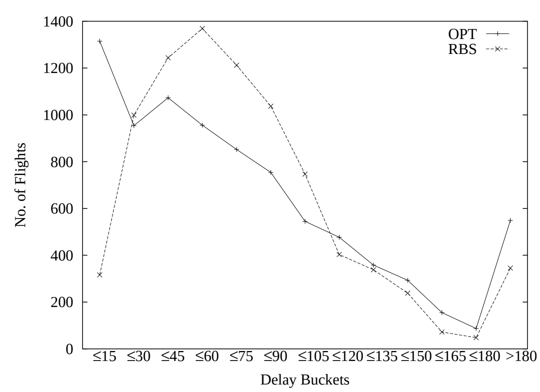


Figure 5.21: RBS vs. Proportional Approximation: All Flights

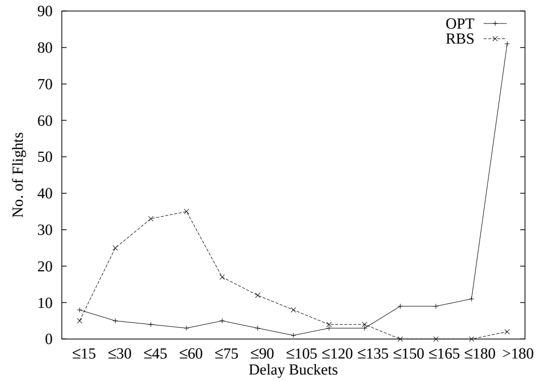


Figure 5.22: RBS vs. Proportional Approximation: General Aviation Flights

flights. This is illustrated in Figure 5.22, which compares the distributions of delays under RBS and under the optimization-based approach over all the GA flights for the same set of GDPs.

The discrepancy in the delay distributions for GA flights could potentially be the result of strategic behavior ("gaming") by GA operators. Since GA flights do not have a scheduled arrival time, RBS uses the arrival times submitted by GA operators to determine their priority. The changes in delay shown in Figure 5.22, however, indicate that an unusually large number of GA flights tend to arrive in the early stages of a GDP. This follows from the fact that both the optimization model and RBS can be interpreted as priority schemes (RBS is based on the scheduled arrival time, whereas the proportional shares are based the ideal positions). Thus, the results in Figure 5.22 imply that the relative priority of GA flights will generally be significantly lower under the proportional scheme. However, it is easy to verify that if an airline has only 1 flight (i.e. a GA flight), its priority according to the proportional scheme should fall somewhere in the middle (that is, approximately half the other priorities would be lower and half the other priorities are higher). On the other hand, under RBS a flights priority would be in the middle if it was scheduled to arrive in the middle of the GDP period. Consequently, a disproportionate number of GA flights currently appears to arrive in the first half of GDP (based on planned arrival times submitted by the GA operators).

5.5 Discussion

This Chapter introduced methods to approximate fair slot shares in situations where the “ideal” allocation might not be attainable. Within the context of GDPs, these methods were primarily motivated by the impact of program dynamics (i.e. flight cancellations/delays and flight exemptions). First, we addressed the impact of flight cancellations and delays, and proposed optimization procedures based on closely related models used in apportionment and jit scheduling problems. The resulting models yielded an intuitive greedy procedure, which we showed is very similar to the Compression Algorithm that is currently used. Subsequently, we discussed how these methods could be extended to manage flight exemptions. While the greedy procedure might no longer be applicable in this case, empirical results clearly showed the potential to reduce systematic biases inherent in the current procedures. Finally, we considered the use of these optimization models to implement alternate standards of fairness. Empirical results showed that using proportional shares (independent of the of scheduled arrival times) could have a significant impact, which could potentially be the consequence of strategic behavior from GA operators.

Chapter 6

Slot Trading during Ground Delay Programs

The previous chapters describe an approach to the allocation of slots during a GDP that is based on priorities, and where a single allocation scheme is executed periodically in response to dynamic changes that may occur. Under this interpretation, the role of the airlines in the (re)allocation process is limited to the provision of schedule updates (i.e. flight cancellations and delays). Consequently, changes in airline preferences are only considered internally by flight substitutions and/or cancellations (which, under this interpretation, may be viewed as an internal reassignment of priorities to flights).

This chapter, in contrast, follows an approach in which airlines "own" a given set of slots (as opposed to priorities), which is closer to the currently established interpretation under CDM. We study the potential benefits of more active airline involvement in the allocation process, by considering a system in which airlines can actively pursue schedule improvements by proposing trades. Under this approach, the FAA acts as a mediator coordinating the resulting exchange of slots. The organization of this chapter is as follows. First, we give a brief

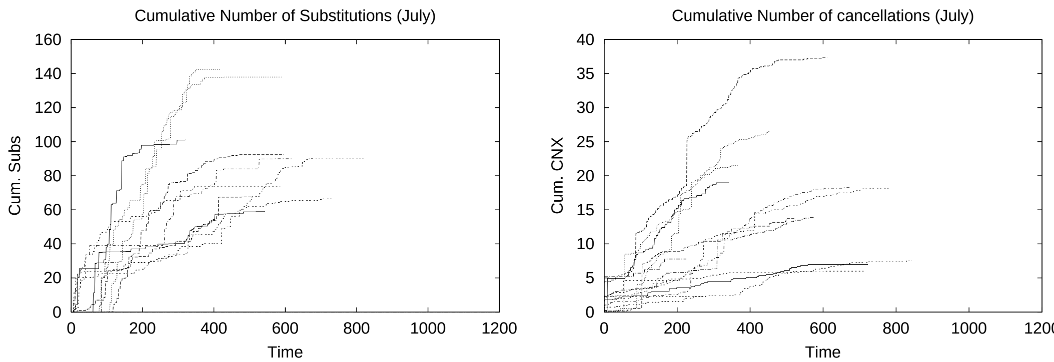


Figure 6.1: Airline GDP behavior at O'Hare Airport, July 2000

motivation of our approach and discuss the relationship to current CDM efforts. Subsequently, we discuss potential slot trading mechanisms, and analyze their benefits under different models of airline decision-making.

6.1 Introduction

From an airline standpoint, the ability to substitute flight-slot assignments is clearly the single most important aspect of a GDP. As discussed in Chapters 2 and 3, this allows an airline to mitigate the disruptions to its flight schedule, and address the potential downstream effects of ground delays. A clear indication of their importance follows by considering Figure 6.1, which shows flight substitution and cancellation patterns using empirical results from actual GDPs at O'Hare airport during July 2000.

The leftmost graph in Figure 6.1 represents the cumulative number of flights (as a percentage of the total number of flights that have been allocated a slot) that are substituted during the course of a GDP day (note that percentages can be greater than 100 since a single flight can be involved in multiple substitutions).

Time 0 corresponds to the first time instance at which each flight was first allocated a slot, and each curve corresponds to one GDP day. It should be noted that flight substitutions due to the Compression Algorithm or GDP revisions were not included; substitutions of cancelled flights were not included either. Similarly, the rightmost graph in Figure 6.1 represents the cumulative percentage of flights that have been cancelled during the course of a GDP. The graphs in Figure 6.1 show first that airlines perform a large number of flight substitutions, and second that airlines perform flight substitutions throughout the course of GDP.

Given the sheer volume of flight substitutions, it is not difficult to imagine that potential benefits could be obtained by allowing the exchange of slots between *different* airlines. That is, by coordinating their flight schedule adjustments airlines might be able to achieve mutual benefits that they would not be able to achieve by themselves. This, of course, is already inherent in the Compression procedure: slots that an airline cannot use (i.e. due to flight cancellations) are exchanged in such a way that all parties involved will receive a reduction in their flight delays. Using the Compression procedure and its reported benefits as a starting point, one could also envision more general exchange mechanisms. In fact, a basic form of such a slot exchange functionality, known as "Slot Credit Substitutions", is currently under consideration in the CDM working group ([30]). Under this proposal, airlines would be able submit what amounts to conditional cancellations: airlines would be able to submit requests of the form "I am willing to cancel flight f_1 (and release its currently assigned slot s_1) if I can move flight f_2 up into (a later) slot $s_{1'}$". The FAA would monitor such requests on a continuous basis, and if possible implement the exchange(s) of slots required to satisfy the request.

The introduction of slot trading during the course of a GDP introduces a wide range of possibilities, in that a number of schemes could potentially be used to coordinate the exchange of slots. One approach, for instance, could be a market-based mechanism in which airlines would be able to buy and sell slots. Another approach could be a system where airline would bargain amongst themselves (see [2] for an classification of potential approaches). As discussed in Chapter 3, however, it is difficult to envision the use of such highly decentralized mechanisms within the context of GDPs. Among others, the high level of uncertainty, the very dynamic environment, and the potential impact on other ATFM initiatives all present significant barriers[1]. In this chapter, we therefore consider more modest generalizations of the slot-credit substitution framework. Under this framework, airlines may submit offers to exchange slots (which could be more general than those allowed under the slot-credit substitution proposal). The FAA, on the other hand, would act as a mediator who evaluates and selects possible trades. To illustrate this general concept, we first discuss how the Compression procedure might be interpreted as a form of mediated bartering. Subsequently, we give a general model representation of the resulting framework.

6.1.1 Compression as Mediated Bartering

In Chapter 5 we discussed how the Compression procedure may be viewed as a form of (re)rationing, instigated by flight cancellations and delays. An alternate interpretation, however, is to view the inter-airline exchange of slots as a form

[1]In addition, it is significant to note that antitrust regulations prohibit direct negotiations between airlines.

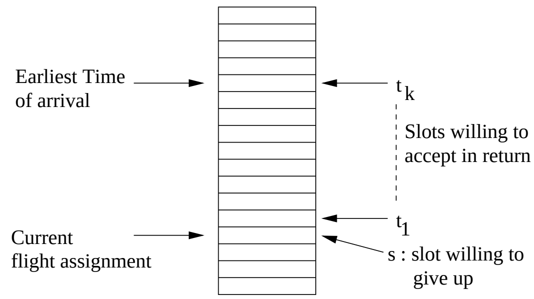


Figure 6.2: "Default" Offers

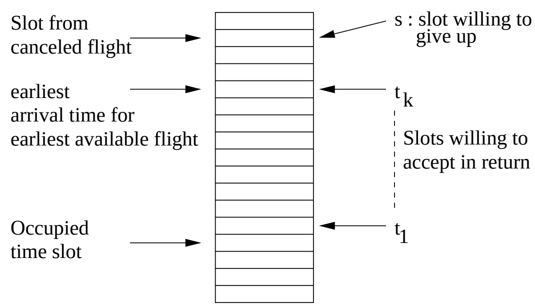


Figure 6.3: Offer associated with cancelled or delayed flights

of bartering, in which the FAA acts as a broker matching offers proposed by the airlines. To illustrate this interpretation, we first observe that all slot exchanges are instigated by a slot that is made available through a cancelled or a delayed flight. Such a slot leads to a series of slot exchanges, in which flights are repeatedly moved up in a way that maximizes the return for the releasing airline. To formalize the bartering interpretation we define an exchange process that is driven by a set of offers made by each airline. There are two generic types of offers, which are depicted in Figures 6.2 and 6.3.

The default offers depicted in Figure 6.2 simply state that an airline would be willing to offer a slot currently occupied by one of its flights in return for an earlier slot, as long as the new slot is not earlier than the earliest time of arrival for the flight. The offers shown in Figure 6.3, on the other hand, apply when a flight is cancelled or delayed. Here, the releasing airline is willing to give up the slot in return for a reduction in the delay of a subsequent designated flight. A single cancellation can lead to multiple offers of this type to effect a set of progressive moves for a single airline's flights.

Given the resulting set of offers, the FAA (in its role as mediator) has to determine which offers to select and execute. In the case of Compression, all exchanges are one-for-one (i.e., a single slot owned by one airline is exchanged for a single slot owned by another airline). As a result, the problem of finding a feasible set of exchange sequences is equivalent to a finding a set of non-intersecting trade cycles, which correspond to the solutions of an assignment problem (see [81] for a detailed discussion). Several criteria could be used to select the actual trades that are executed: one possibility is to use a bilevel programming approach in which offers to move down are given priority (see [81]). This approach yields solutions that are similar to the Compression Algorithm.

6.1.2 Model Description

Under the interpretation of Compression as slot trading, only one-for-one trades are allowed. Here, we describe a more general slot trading model. As in the previous chapters, we let $\mathcal{F} = \{f_0, \ldots, f_{n-1}\}$ represent the flights in the GDP, and $\mathcal{S} = \{s_0, \ldots, s_{n-1}\}$ the slots available during the GDP. The airlines are represented by a set $\mathcal{A}$, and for each airline $a \in \mathcal{A}$, $\mathcal{F}_a \subseteq \mathcal{F}$ represents the flights operated by airline a. At the start of a period of trading, all flights

$$
\begin{array}{lll}
u_a(p(B)) = \text{Max} & \sum_{f\in\mathcal{F}_a, s\in\mathcal{S}} w_{fs}x_{fs} + \sum_{f\in\mathcal{F}_a} c_f y_f & \\
\quad\text{subject to:} & & \\
& \sum_{s\in\mathcal{S}} x_{fs} + y_f = 1 & \text{for all } f \in \mathcal{F}_a \\
& \sum_{f\in\mathcal{F}_a} x_{fs} \leq p(B)_s & \text{for all } s \in \mathcal{S} \\
& x_{fs}, y_f \geq 0 &
\end{array}
$$

Figure 6.4: Airline preferences

have been assigned a slot; we assume that flight f_i is assigned to slot s_i for all $i \in 0, \ldots, n-1$. This assignment specifies each airline's allotment of slots, that is, $\mathcal{S}_a = \{s_i \in \mathcal{S} : f_i \in \mathcal{F}_a\}$ represents the set of slots owned by airline a.

Given these initial allotments, we can associate with each airline a a set of offers $\mathcal{T}_a \subseteq 2^{\mathcal{S}_a} \times 2^{\mathcal{S}-\mathcal{S}_a}$. That is, each offer $t_a = (O_{a,t}, R_{a,t}) \in \mathcal{T}_a$ specifies that airline a would be willing to offer slots in $O_{a,t}$ in return for the slots in $R_{a,t}$. Airline preferences over these offers are implied by a value $w_{a,t}$ for each offer $t_a \in \mathcal{T}_a$.

In the remainder of this chapter, we assume that an airline's preferences can be expressed by an assignment model as shown in Figure 6.4. Here, $p(B) \in \mathbb{R}^n_+$ with $p(B)_j = 1$ if $j \in B$ and 0 otherwise. Thus, an airline's value for the bundle of slots S is obtained by solving as assignment model, where w_{fs} represents the value of assigning flight f to slot s, and c_f represents the cost of cancelling flight f. As such, an airline's value $w_{a,t}$ for an offer $t_a = (O_{a,t}, R_{a,t})$ can be defined as $w_{a,t} = u_a(p(\mathcal{S}_a - O_{a,t} + R_{a,t})) - u_a(\mathcal{S}_a)$.

6.2 Background

The implementation of this general framework poses a number of issues. First, we have to specify which offers to allow and (potentially) how airlines may submit their preferences over different offers. Given this information, the framework requires a criterion or mechanism for determining which offers to accept, which may involve a number of criteria. One common criterion could be the (pareto) efficiency of the resulting allocation. Another desirable aspect could be the stability of the resulting allocation, which may involve equity considerations. An additional concern is introduced by the incentives the mechanism may generate, that is, the airlines may strategically misrepresent their preferences. In this section, we outline two potential allocation criteria, and illustrate their limitations within the context of slot trading during GDPs.

Cooperative Games without Side Payments

Let us assume, for now, that the mediator has complete knowledge of each airline's preferences. Thus, our only concern is the criterion for determining trades. One possibility is to represent the model as a cooperative game without side payments. A cooperative game without side payments can be defined as follows (see [58]).

Definition 6.2.1. *A* Cooperative Game without Side Payments *is defined as a tuple $\langle N, X, V, (\succeq_i)_{i \in N} \rangle$, where N represents the set of players, X represents the set of possible outcomes, $V : 2^N \rightarrow 2^X$ is a function that associates with each coalition G a set of outcomes $V(G)$, and $\succeq_i$ is a preference relation over X for all $i \in N$.* □

The core of a cooperative game without side payments is defined as follows.

Definition 6.2.2. *The core of the cooperative game $\langle N, X, V, (\succeq_i)_{i \in N} \rangle$ is the set of all $x \in V(N)$ for which there is no coalition G and $y \in V(G)$ such that $y \succ_i x$ for all $i \in G$.* □

It is relatively straightforward to represent the slot trading model as a cooperative game without side payments in which the players correspond to the airlines. X can be defined as the set of all possible allocations of slots to airlines. For any coalition G, we can define $V(G)$ as the subset of those allocations in which the airlines in G have been assigned slots in $\bigcup_{a \in G} \mathcal{S}_a$. Thus, $V(G)$ represent the allocations that can be achieved if the airlines in G trade amongst themselves. Finally, the preference relationships $\succeq_a$ are defined by the utility functions u_a. Intuitively, therefore, the core represents the set of allocations such that no group of airlines could each improve by trading amongst themselves.

In the special case that each airline owns exactly one flight, the resulting cooperative game corresponds to the well-known "housing" market proposed by Shapley and Scarf ([52], [67]). It is well-known that in this case the core is non-empty, and that under certain restrictions on the preference relations the core consists of a single allocation. Moreover, an intuitive procedure (the so-called "top-trading cycles" algorithm) can be used to determine core allocations. Unfortunately, however, these nice results do not extend to the more general case in which airlines may have more than one flight. In this case the core may be empty; that is, no stable allocations may exist. This is illustrated in the counterexample shown in Figure 6.5, which is a slight adaptation from the example given by Konishi et al. ([37]). Figure 6.5 contains a simple GDP instance, in which there are six flights and slots owned by four airlines. Among all the possible allocations of slots to the airlines, there are only four that are

- $\mathcal{A} = \{a, b, c, d\}$, $\mathcal{F} = \{f_1, \dots, f_6\}$, $\mathcal{S} = \{s_1, \dots, s_6\}$;
- $\mathcal{F}_a = \{f_1, f_2\}$, $\mathcal{S}_a = \{s_1, s_2\}$, $\mathcal{F}_b = \{f_3\}$ and $\mathcal{S}_b = \{s_3\}$, $\mathcal{F}_c = \{f_4\}$, $\mathcal{S}_c = \{s_4\}$, $\mathcal{F}_d = \{f_5, f_6\}$, $\mathcal{S}_d = \{s_5, s_6\}$;
- w_{fs} equals

w_{fs}	s_1	s_2	s_3	s_4	s_5	s_6
f_1	0.01	10	6	0.1	0.01	5
f_2	0.01	10	6	0.1	0.01	5
f_3	0.01	0.1	1	3	2	0.01
f_4	0.01	0.1	2	1	0.01	3
f_5	0.01	20	0.1	6	10	5
f_6	0.01	20	0.1	6	10	5

Figure 6.5: Counterexample Data

individually rational (that is, allocations in which no airline is worse off than it was in the initial allocation):

- $X_1 : X_{1,a} = \{s_1, s_2\}, X_{1,b} = \{s_3\}, X_{1,c} = \{s_4\}, X_{1,d} = \{s_5, s_6\}$,
- $X_2 : X_{2,a} = \{s_1, s_2\}, X_{2,b} = \{s_4\}, X_{2,c} = \{s_3\}, X_{2,d} = \{s_5, s_6\}$,
- $X_3 : X_{3,a} = \{s_1, s_2\}, X_{3,b} = \{s_3\}, X_{3,c} = \{s_6\}, X_{3,d} = \{s_4, s_5\}$,
- $X_4 : X_{4,a} = \{s_3, s_6\}, X_{4,b} = \{s_5\}, X_{4,c} = \{s_4\}, X_{4,d} = \{s_1, s_2\}$.

However, allocation X_1 is blocked by allocation X_2 through $\{b, c\}$ (that is, airlines b and c would be better off by trading amongst themselves). Similarly, allocation X_2 is blocked by allocation X_3 through $\{c, d\}$. Allocation X_3 is blocked by allocation S_4 through $\{a, b, d\}$. Finally, allocation X_4 is blocked by allocation X_2 through $\{b, c\}$. Consequently, the core of the corresponding cooperative game without side payments must be empty.

Randomized Slot Trading

Since the core may be empty, we cannot expect to use it as a criterion for trading slots. As such, we cannot expect to apply a procedure like the top-trading cycle algorithm to implement our slot trading framework. A potential remedy to this problem would be to allow, as was done in Chapter 4, slots to be divisible. This would induce a form of randomized trading, similar to the procedure proposed by Hylland and Zeckhauser ([31]) for the allocation of students to dorms. To formalize this idea, we represent the slot trading framework as an exchange economy. An exchange economy can be defined as follows.

Definition 6.2.3. *An* Exchange Economy *is defined as a tuple* $\langle N, n, (\omega_i)_{i \in N}, (\succeq_i)_{i \in N} \rangle$*, where* N *represents the set of players,* n *represents the number of commodities,* ω_i *represents the initial endowment of player* i*, and* $\succeq_i$ *is a preference relation over the bundles in* $\mathbb{R}^n_+$ *for all* $i \in N$*.* □

Again, it is relatively straightforward to represent the slot trading model as an exchange economy: the players corresponds to the airlines, the commodities to the slots, the initial endowments to the slots owned by each airline (i.e. $\omega_{a,j} = 1$ if $j \in \mathcal{S}_a$ and 0 otherwise), and airline preferences are represented by the utility function u_a. Feasible allocations to the resulting model are given by the set

$$X = \{x \in \mathbb{R}^{\mathcal{A} \times n}_+ : \sum_{a \in \mathcal{A}} x_{a,j} = \sum_{a \in \mathcal{A}} \omega_{a,j}\},$$

which generalizes the previous model in that fractional allocations of slots to airlines are allowed. A fractional assignment x can be interpreted as a "lottery" over integral allocations, which follows by representing x as a convex combination of the extreme points of X. It is important to note, however, that this approach requires the assumption that airlines have so-called von Neumann-Morgenstern utility functions. In other words, an airline's utility for a fractional assignment equals its expected utility in the lottery over integral allocations (and airlines compare lotteries by comparing their expected utilities).

An important concept in exchange economies is the notion of a competitive equilibrium, which is defined as follows.

Definition 6.2.4. *A* Competitive Equilibrium *is an allocation-price pair* (x, p) *where* $x \in X$, $p \in \mathbb{R}^n_+$ *and for all* $a \in \mathcal{A}$, $y \in \mathbb{R}^n_+$,

$$u_a(y) > u_a(x) \quad \Rightarrow \quad \sum_{j=0}^{n-1} p_j y_j > \sum_{j=0}^{n-1} p_j \omega_{a,j}.$$

□

It is well-known (see [59]) that if (x, p) is a competitive equilibrium, the allocation x will be in the core of the corresponding cooperative game without side payments (in which fractional assignments are allowed). Moreover, it can be shown that if the utility functions u_a are continuous, concave and monotone non-decreasing in each variable, a competitive equilibrium exists (see [59]). Thus, since the utility functions shown in Figure 6.4 are piecewise linear and concave (see [53], p.42), the core will be non-empty under this representation of the slot trading model[2].

As such, the interpretation as an exchange economy presents a possible approach to the design of slot trading mechanism: given each airline's preferences, the mediator could determine a competitive equilibrium, resolve the resulting lottery (such a procedure is discussed in [31]), and implement the final allocation. Of course, a critical issue still to be resolved would be a scheme that would

[2]It should be noted that this result is only valid if the airlines value for an assignment can be expressed through an assignment model as in Figure 6.4.

induce airlines to reveal their preferences. Aside even from this issue, however, it is unlikely that such a form of randomized trading would yield a satisfactory approach. In particular, the critical assumption that airlines have von Neumann-Morgenstern utility functions is unlikely to hold within the context of GDPs. For example, it is difficult to envision that an airline would cancel or even delay one of its flights in return for a *probability* that another flight's delay is reduced.

6.3 Approach

In light of the results discussed in the previous section, the remainder of this chapter sets out a somewhat different approach. Instead of using the core or competitive equilibria as the allocation criterion, we consider a system in which the mediator simply aims to maximize the number of possible trades, or optimizes some objective function that embodies certain system-wide performance goals. The resulting slot trading framework can be summarized as follows:

- Periodically (say every 15 or 30 minutes), airlines submit a list of trade offers they would desire.

- Subsequently, the mediator (FAA) will either maximize the number of trades that can be executed, or sequentially execute as many feasible trades as possible.

Note that the airline-provided information does not include any information about its relative value for these trades. We assume, implicitly, that each offer specifies how each flight would be assigned to the slots traded for; this would be necessary to maintain a feasible allocation of slots to flights. In addition to the offers, $\mathcal{T}_a$, proactively provided by the airlines, we also assume the availability

$$
\begin{aligned}
&\text{Max} \quad \textstyle\sum_{a\in\mathcal{A},t_a\in\mathcal{T}_a} y_{a,t}\\
&\text{subject to:}\\
&\qquad \textstyle\sum_{(s_i,s_j)\in\mathcal{D}} x_{i,j} + \sum_{a\in\mathcal{A},t_a\in\mathcal{T}_a:s_i\in O_{a,t}} y_{a,t} = 1 \quad \text{for all } s_i \in \mathcal{S}\\
&\qquad \textstyle\sum_{(s_i,s_j)\in\mathcal{D}} x_{i,j} + \sum_{a\in\mathcal{A},t_a\in\mathcal{T}_a:s_j\in R_{a,t}} y_{a,t} = 1 \quad \text{for all } s_j \in \mathcal{S}\\
&\qquad x_{i,j}, y_{a,t} \in \{0,1\}
\end{aligned}
$$

Figure 6.6: IP formulation for Mediation Problem

of default offers, which specify that an airline would always be willing to reduce the delay of any of its flights, i.e.

$$\mathcal{D} = \{(s_i, s_j) : 0 \le j, i \le n-1, e_i \le j \le i\}.$$

Given these offers, the mediator's task is to find the maximum number of desirable offers that are compatible. This problem can be formulated as a set-partitioning problem, as shown in Figure 6.6. In this formulation, the variables $y_{a,t}$ are associated with the trade offers $t_a \in \mathcal{T}_a$, that is, $y_{a,t} = 1$ if and only if offer t_a is selected. The variables $x_{i,j}$ correspond to the default offers $(s_i, s_j) \in \mathcal{D}$, and $x_{i,j} = 1$ if and only if slot s_i is exchanged for slot s_j (or equivalently, flight f_i is assigned to slot s_j). The first constraint states that each slot is assigned to an offer (default or airline provided) that proposes to give up the slot. The second constraint states that each slot is assigned to an offer that requires the slot in return (note that the situation where slot s_i is not traded corresponds to selecting the default offer (s_i, s_i)).

Offer Structure

In principle, the set-partitioning formulation can be used to accommodate any offer an airline might find desirable. Here, however, we consider a more restricted approach, in which airlines are allowed to propose only "two-for-two" trade offers (i.e., an offer consists of an exchange of two slots for two other slots). The motivation for introducing this restriction is the reduction in complexity of the resulting framework. This reduction in complexity not only applies to the mediator's problem, but also to the evaluation and generation of potential offers by each individual airline. While this may be a complex problem in general, it is relatively straightforward to evaluate these pairwise offers.

Even though these restrictions limit the potential exchanges during the course of a GDP, the oftentimes specific nature of airline objectives indicates that two-for-two trades may still be of substantial use. Before discussing this further, however, it is worthwhile to first look at the structure of two-for-two trades. Any two-for-two trade offer involves two flights, whose assigned slots are offered for two other slots. As such, these offers can be separated into three classes: (1) the offer expresses a trade for two earlier slots (i.e. both flights are moved up), (2) the offer expresses a trade for two later slots (e.g. both flights are moved down), or (3) the offer expresses a trade for one earlier slot and one later slot (e.g. one flight is moved up while the other is moved down). It is safe to dismiss the first two classes: the first class is subsumed by the default offers while it is hard to imagine why an airline would submit an offer in the second class. As such, we can safely interpret (a class of) two-for-two trades as an "at-least, at-most" offer which indicates that an airline demands a certain minimum delay reduction on one flight in return for a maximum amount of additional delay imposed on another flight.

Thus, two-for-two trades could allow airlines to make local adjustments to

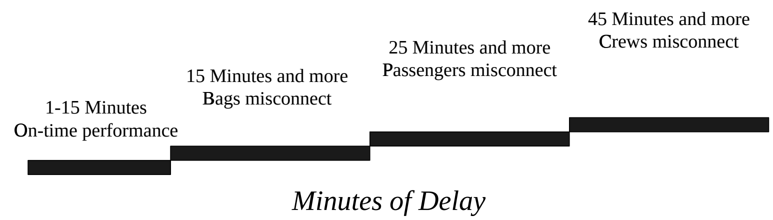


Figure 6.7: Delay Cost Structure

their flight schedule by trading off the marginal delay costs between pairs of flights. For instance, an airline could offer to delay a flight with few passengers in return for delay reduction on a more heavily loaded flight that would allow its passengers to make their connections. More generally, a flight's delay costs are oftentimes reasonably approximated by a staircase structure as shown in Figure 6.7 ([18],[26]). This structure is motivated by operationally significant delay levels within each carrier. For instance, the industry standard for an on-time arrival is 0 to 15 minutes delay beyond scheduled arrival time. Thus, the difference between 4 and 9 minutes of delay is not nearly as significant as the difference between 14 and 19 minutes of delay. Similarly, between 15 and 25 minutes, the rate of missed baggage connections begins to increase, and between 25 and 45 minutes of delay, passengers begin to miss connections. With delays over 45 minutes, crews begin to miss connections. Of course, the exact times and significance of these classes may differ on a flight to flight basis. Yet these examples illustrate that, in general, there may be substantial differences in the marginal delay costs for different flights, which motivates the potential use of pairwise trade-offs.

Model Formulation

While the formulation as a set-partitioning problem could also be used to find

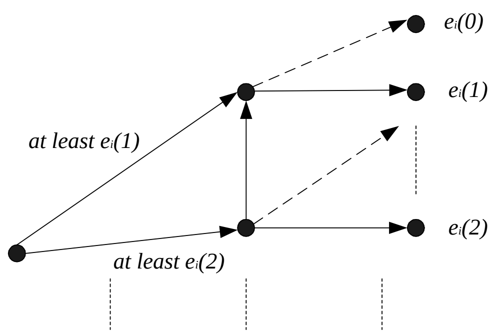


Figure 6.8: Flight Assignment Structure

compatible trades in the case of "at-least,at-most" offers, the large number of variables would likely make this approach intractable for all but the smallest cases. When only two-for-two trades are allowed, however, the resulting set of offers can be defined more succinctly.

Specifically, each trade $t \in \mathcal{T}$ can be characterized by a tuple (d^t, m^t, u^t, l^t), which states that the airline is willing to move down flight f_{d^t} to a slot no later than m^t in return for moving up flight f_{u^t} to a slot that is no later than l^t. In the remainder of this section we therefore discuss an alternate formulation of the mediation problem, which takes into account the underlying offer structure.

This formulation may be viewed as a network flow problem with side constraints. The general idea is that each flight is assigned to a class, which represents the amount of delay or delay reduction that each flight receives (i.e. at least d units reduction in delay, at most d units additional delay). The side constraints are needed to ensure that only assignments corresponding to proposed offers are selected. To illustrate this idea, we first consider a single flight f_i and examine all the offers it occurs in. These offers determine a sequence of classes $e_i(0) < e_i(1) < \cdots < e_i(k_i)$, where $e_i(0) = oag_i$ (that is, the earliest time of arrival for flight f_i). Thus, if $e_i(k) < i$ there is an offer which contains a demand

$$
\begin{array}{ll}
\text{Max} \ \ \sum_{(i,k)\in D_{\mathcal{T}}} x_{i,k} & \\
\text{subject to:} & \\
\quad \sum_{k=1}^{k_i} x_{i,k} = 1 & \text{for all } i \in \mathcal{S} \\
\quad x_{i,1} + z_{i,1} = \sum_{s=e_i(0)}^{e_i(1)} y_{i,s} & \text{for all } i \in \mathcal{S} \\
\quad x_{i,k} + z_{i,k} - z_{i,k-1} = \sum_{s=e_i(k-1)+1}^{e_i(k)} y_{i,s} & \text{for all } i \in \mathcal{S}, k \in 2, \ldots, k_i \\
\quad \sum_{s=e_i(0)}^{e_i(k_i)} y_{i,s} = 1 & \text{for all } s \in \mathcal{S} \\
\quad x_{i,k} \leq \sum_{(i,k,u,l)\in\mathcal{T}} \hat{x}_{i,k,u,l} & \text{for all } (i,k) \in D_{\mathcal{T}} \\
\quad x_{u,l} = \sum_{(i,k,u,l)\in\mathcal{T}} \hat{x}_{i,k,u,l} & \text{for all } (u,l) \in U_{\mathcal{T}} \\
\quad x_{i,k}, \hat{x}_{i,k,u,l}, y_{i,k}, z_{i,s} \in \{0,1\} &
\end{array}
$$

Figure 6.9: IP formulation for Restricted Mediation Problem

of at least $e_i(k)$ for slot i. Similarly, if $e_i(k) > i$ there is an offer to move down f_i to at most position $e_i(k)$. In addition, we assume that there is one $k : 0 \leq k \leq k_i$ such that $e_i(k) = i$. Intuitively, each of the elements in this sequence represent classes that flight f_i can be assigned to, as shown in Figure 6.8. Figure 6.8 also shows that once a flight is assigned to a class, it will subsequently be assigned a slot according to the bounds implied by the class. To represent the IP formulation, we define $D_{\mathcal{T}} = \bigcup_{(i,k,u,l)\in\mathcal{T}}\{(i,k)\}$, $U_{\mathcal{T}} = \bigcup_{(i,k,u,l)\in\mathcal{T}}\{(u,l)\}$, and $N_{\mathcal{T}} = \bigcup_{i\in\mathcal{S}}\{(i,i)\}$. Thus, $D_{\mathcal{T}}$ contains the classes that correspond to downward moves, $U_{\mathcal{T}}$ contains the classes that correspond to upward moves, while $N_{\mathcal{T}}$ contains the classes corresponding to default offers. The resulting IP formulation is shown in Figure 6.9. The variables in this formulation can be interpreted as follows.

- The variables $x_{i,k}$ represent the assignment of a flight to a class, that is,

$x_{i,k} = 1$ iff f_i is assigned at least slot $e_i(k)$ for $i \in \mathcal{F}, 1 \leq k \leq k_i$.

- The variables $\hat{x}_{i,k,u,l}$ represent the execution of an offer, that is, $x_{i,k,u,l} = 1$ if offer (i, k, u, l) is executed for $(i, k, u, l) \in \mathcal{T}$.
- The variables $y_{i,s}$ represent the actual assignment of a flight to a slot, i.e. $y_{i,s} = 1$ iff f_i is assigned to slot s for $i \in \mathcal{F}, s \in \mathcal{S}$.
- The variables $z_{i,k}$ are used to complete the assignment of flights to classes, i.e. $z_{i,k} = 1$ iff f_i has been assigned to a class lower than k but receives at least slot $e_i(k)$.

The first constraint in the IP formulation represents the assignment of flights to classes. The second and third constraints represent the subsequent assignment of classes to slots, and the fourth constraint represents the restriction that each slot is assigned exactly once. The fourth constraint ensures that the resulting trades only include offers proposed by the airlines. Specifically, the constraint states that a flight is moved down only if another flight is moved up, in accordance with one of the proposed trades. The final constraint states that at most one flight will be moved down for any flight that is moved up.

6.4 Case Studies

The remainder of this chapter further pursues the approach to slot trading outlined in the previous section, using two case studies. The case studies we consider rely on two different models of airline decision-making: the first model assumes that the airlines' objective is to maximize on-time performance, while in the second model the airline objective is to minimize passenger delays. While these basic models cannot accurately represent the full complexity of the decisions and trade-offs airlines are faced with during the course of a GDP, we believe these models do incorporate important factors in the airlines' decision-making process (see [41], [40], [54]). The objective of these case studies is threefold: (1) to analyze potential benefits of increased coordination, (2) to analyze the efficiency of the underlying optimization models, and (3) to consider the impact of airline behavior in the trading process.

6.4.1 On-Time Performance

In the first case we consider, we assume that each airline's objective is to maximize on-time performance. A flight is said to be on-time if it arrives within 15 minutes of its scheduled arrival time; thus, an airline's objective is to maximize the number of flights that are delayed at most 15 minutes. While this may be the simplest model of airline decision-making imaginable, the importance of on-time performance should not be underestimated. The primary reason for this is that the FAA tracks and publishes the aggregate on-time performance for each airline; the publication of these statistics is said to have a significant impact on consumer preference (cf. [41]).

Restricting the airline objectives to maximizing on-time performance offers a substantial simplification of the trading model. To illustrate this, we first observe that in this case we can represent the airline's performance function using the IP formulation shown in Figure 6.4 with coefficients

$$w_{fs} = \begin{cases} M & \text{if } (t_s - e_f) \leq 15, \\ 0 & \text{otherwise.} \end{cases}$$

Max OV

subject to:

$$x_i^U + x_i^N + x_i^D = 1 \qquad \text{for all } f_i \in \mathcal{F}$$

$$x_i^U + z_i^U = \sum_{j \in U_i} y_{i,j} \qquad \text{for all } f_i \in \mathcal{F}$$

$$x_i^N + z_i^N = z_i^U + \sum_{j \in N_i} y_{i,j} \qquad \text{for all } f_i \in \mathcal{F}$$

$$x_i^D = z_i^N + \sum_{j \in D_i} y_{i,j} \qquad \text{for all } f_i \in \mathcal{F}$$

$$\sum_{f_i \in \mathcal{F}: e_i \leq j} y_{i,j} = 1 \qquad \text{for all } s_j \in \mathcal{S}$$

$$\sum_{f_i \in \mathcal{F}_a} x_i^D \leq \sum_{f_i \in \mathcal{F}_a} x_i^U \qquad \text{for all } a \in \mathcal{A}$$

$$x_f^U, x_f^N, x_f^D, z_f^U, z_f^N, y_{i,j} \in \{0,1\}$$

Figure 6.10: IP formulation for Mediation Problem with On-time Performance Objective

with $M >> 0$ (a slightly different approach will be discussed later in this section). This indicates an application of the trading model introduced in the previous section, in which an airline would submit any offer that would improve on-time performance. Such an offer would state that an airline is willing to move down any flight (that is not already arriving on-time in the current allocation) in return for a delay reduction that makes another flight arrive on time. Note that such an offer would never involve a flight that is already arriving on time in the current allocation. In this case, we can distinguish three possible "at-least,at-most" classes for each flight: (1) a flight will be assigned at least a slot corresponding to an on-time arrival, (2) a flight will be assigned at least the slot it currently occupies (e.g. a default offer), and (3) a flight will be assigned at most the last slot in the GDP (e.g. the flight is delayed). In this case, the general formulation of the mediation problem discussed before can be simplified to the formulation shown in Figure 6.10.

In this formulation, the three sets U_i, N_i and D_i correspond to the three different classes for each flight f_i. As indicated before, the definition of these sets depends on flight f_i's current delay. If flight f_i is currently arriving on time only the default offer applies, and therefore we have $U_i = D_i = \emptyset$ and $D_f = \{e_i, i\}$. Otherwise, the flight can either moved up or down and we have $U_i = \{j \in \mathcal{S} : t_j - e_i \leq 15\}, N_i = \{j \in \mathcal{S} : j \leq i\}/U_i$ and $D_i = \{j \in \mathcal{S} : j > i\}$. As before, the variables x_i^U, x_i^D, x_i^N represent the assignment of a flight to a class, the variables $y_{i,j}$ represent the assignment of a flight to a slot, and the variables z_i^U, z_i^N complete the assignment of flights to classes. The constraints in Figure 6.10 are analogous to the constraints in the general formulation with the exception of the final constraint, which ensures that for each airline only offers which improve on-time performance are executed. The fact that an airline is willing to arbitrarily trade off additional delays for an improvement in on-time performance allows us to represent the acceptable offers with a single constraint for each airline. Finally, we note that a number of different objective functions are possible, since any solution for which $\sum_{f_i \in \mathcal{F}} x_i^N < n$ results in the execution of at least one desired trade. Thus, some possibilities would be to maximize the number of flights that have been moved on time (e.g. Maximize $\sum_{f_i \in \mathcal{F}} x_i^U$), to maximize the number of flights that have been delayed (Maximize $\sum_{f_i \in \mathcal{F}} x_i^D$), or to minimize the number of default offers executed (Minimize $\sum_{f_i \in \mathcal{F}} x_i^N$). The impact of these different possiblities is discussed later in this section.

Initial Empirical Results

As a first step in our analysis, we considered the *potential* benefits that could be obtained by trading slots among airlines. For our analysis, we used historical data from a set of GDPs at Boston's Logan airport between January and April of 2001 (the same data was also used in Chapter 5). For each of these GDPs, we first executed RBS and fixed the assignment of all the flights that were exempted from the program. The resulting flight-slot assignment of the remaining (non-exempt) flights was used to determine each airline's initial allotment of slots. Subsequently, we maximized on-time performance for each individual airline using the assignment problem formulation shown in Figure 6.4. In practice, this allocation would result using the substitution process currently in place. The total number of flights arriving on time after this stage represents the optimum that could be obtained *without* coordination. In addition, we also maximized the overall on-time performance (using a single assignment problem formulation without considering slot ownership). This provides a bound on the benefits that could be obtained *with* coordination.

To analyze the potential benefits of our 2-for-2 slot trading mechanism, we applied the mediation problem to the allocation that would have been obtained after each airline individually maximized its on-time performance. The objective function we used was to maximize the number of flights that were moved up. We compared these benefits with the improvement in on-time performance that could be achieved with the Compression Algorithm. To obtain an estimate of the Compression benefits, we assumed that airlines would cancel all flights with excessive delays (i.e. delays of 2 hours or more). After cancelling these flights, we applied the Compression Algorithm. The results are shown in Fig-

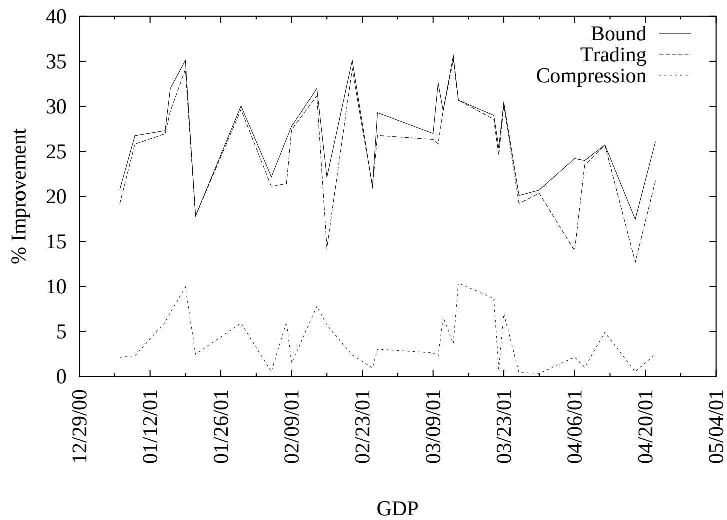


Figure 6.11: On-time Performance Improvements from Slot Trading

ure 6.11, which shows the improvements in on-time performance relative to the optimum that could be obtained without coordination (as a percentage of the total number of non-exempted flights in the GDP). Here, the solid line represents the upper bound on the increase that could be obtained by coordination. The dashed line represents the relative improvements that would be obtained by 2-for-2 slot trading, while the dotted line represents the improvements obtained by the Compression Algorithm. On average, the potential increase in on-time performance would be 26.8% (the average number of flights in a GDP was 216.1, while the average number of flights arriving on-time without coordination was 100.1). The average relative improvement obtained by slot trading was 24.9%, while the average improvement with the Compression Algorithm was 3.9%. These results clearly indicate that slot trading could yield substantial benefits: while the Compression Algorithm would only lead to modest improvements in on-time performance, the use of slot trading nearly always yielded improvements close to the theoretical maximum.

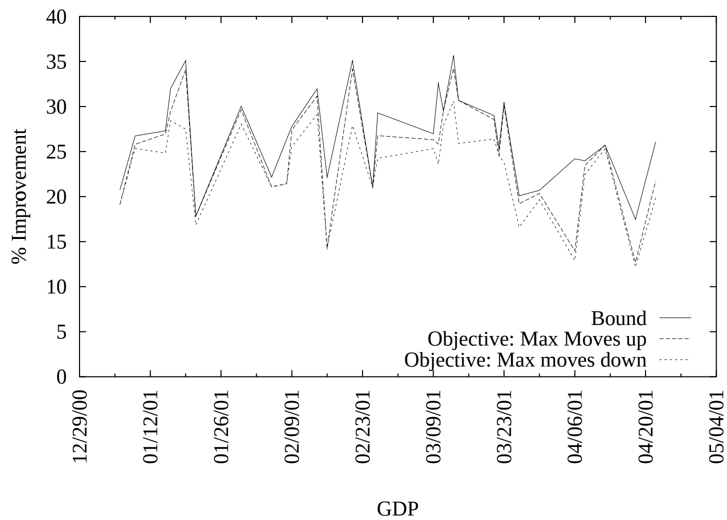


Figure 6.12: Impact of Objective Function Choice

Alternate Objectives and Formulations

We now turn to the performance and efficiency of the slot trading model itself, and consider the impact of using alternate objectives and formulations in the mediation problem. To illustrate theses issues, we first compared the improvements in on time performance that would be obtained with different objective functions. Figure 6.12 shows the relative improvements in on-time performance under two objective functions, maximizing the number of flights moved up and maximizing the number of flights moved down. The solid line again represents the theoretical maximum, the dashed line represents the relative improvements when the number of flights moved up is maximized, while the dotted line represents the improvements when the number of flights moved down is maximized. Both objectives yields substantial improvements, though the use of the first objective leads to slightly better results. In addition to the differences in on-time performance, the objective function choice can also lead to substantial differences in the efficiency of the resulting IP formulations. If we maximized the number of flights moved up, for instance, the average solution time equals 8.3 seconds[3]. However, if we maximized the number of flights moved down the average solution time was 163 seconds[4]. Given the the impact of objective function choice on the solution times, we also considered the use of alternate IP formulations for the mediation problem. We analyzed two alternatives to the IP formulation shown in Figure 6.10: the first formulation aims to improve efficiency by strengthening the formulation, while the second aims to reduce the size of the resulting IP formulations.

To strengthen the formulation, we can reformulate the constraints that express the assignment of flights to classes for each airline $a \in \mathcal{A}$, i.e.

$$x_i^U + x_i^N + x_i^D = 1 \quad \text{for all } f_i \in \mathcal{F}_a, \text{ and} \tag{6.1}$$

$$\sum_{f_i \in \mathcal{F}_a} x_i^D \leq \sum_{f_i \in \mathcal{F}_a} x_i^U. \tag{6.2}$$

The key to reformulating these constraints is the observation that feasible solutions to these constraints can be represented as shortest paths in appropriately defined graphs (akin to the formulation of knapsack problems as dynamic programming problems). To illustrate this, we define for each airline $a \in \mathcal{A}$ with $\mathcal{F}_a = \{f_1^a, \dots, f_{n_a}^a\}$, the set of nodes $V_a = \{(i,j) : 1 \leq i \leq n_a, -\lfloor \frac{n_a}{2} \rfloor \leq j \leq \lfloor \frac{n_a}{2} \rfloor\}$. Intuitively, each node represents a state; that is node (i,j) would indicate that after assigning flights $f_1^a, \dots, f_{i-1}^a$ the difference between the flights moved up and the flights moved down equals j. Observe that the maximum difference equals $\lfloor \frac{n_a}{2} \rfloor$, since there has to be a flight moved up for every flight moved down. Given this set of nodes, the assignment of flights to classes could be

[3] the IP formulations were solved using Cplex 7.1, using a Sun Ultra 10 work station.

[4] We limited the maximum number of nodes visited in the branch and bound tree to 1000.

represented by a set arcs $E_a = E_a^U \cup E_a^N \cup E_a^D$, with $E_a^U = \{((i,j),(i+1,j+1)) : (i,j),(i+1,j+1) \in V_a\}$, $E_a^N = \{((i,j),(i+1,j)) : (i,j),(i+1,j) \in V_a\}$, and $E_a^D = \{((i,j),(i+1,j-1)) : (i,j),(i+1,j-1) \in V_a\}$. Intuitively, each arc corresponds to the assignment of a flight to a class; for instance the arc $((i,j),(i+1,j+1)) \in E_a^U$ would correspond to moving flight f_i^a up, e.g. $x_{a_i}^U = 1$. As such, any path from node $(1,0)$ to a node (n_a, j) with $j \geq 0$ would correspond to a solution that satisfies constraints 6.1 and 6.2.

The resulting formulation can be obtained by introducing the following variables (corresponding to the arcs in E_a).

- $\hat{x}_{a_i,j}^U \in \{0,1\}$, $a \in \mathcal{A}$, $(a_i, j) \in V_a$;
- $\hat{x}_{a_i,j}^N \in \{0,1\}$, $a \in \mathcal{A}$, $(a_i, j) \in V_a$;
- $\hat{x}_{a_i,j}^D \in \{0,1\}$, $a \in \mathcal{A}$, $(a_i, j) \in V_a$.

Using these variables, the flow balance constraints can be represented as

$$\hat{x}_{a_1,0}^U + \hat{x}_{a_1,0}^N + \hat{x}_{a_1,0}^D = 1 \quad \text{for all } a \in \mathcal{A}, \tag{6.3}$$

$$\hat{x}_{a_{i-1},j-1}^U + \hat{x}_{a_{i-1},j}^N + \hat{x}_{a_{i-1},j}^D = \hat{x}_{a_i,j}^U + \hat{x}_{a_i,j}^N + \hat{x}_{a_i,j}^D \quad \text{for all } a \in \mathcal{A},\ (i,j) \in V_a,\ i > 1, \tag{6.4}$$

$$\sum_{j \geq -1} \hat{x}_{a_n,j}^U + \sum_{j \geq 0} \hat{x}_{a_n,j}^N + \sum_{j \geq 1} \hat{x}_{a_n,j}^D = 1 \quad \text{for all } a \in \mathcal{A}. \tag{6.5}$$

Replacing constraints 6.1 and 6.2 by constraints 6.3, 6.4, and 6.5 in the original formulation, and substituting

$$x_{a_i}^U = \sum_{(a_i,j) \in V_a} \hat{x}_{a_i,j}^U, \quad x_{a_i}^N = \sum_{(a_i,j) \in V_a} \hat{x}_{a_i,j}^U, \quad x_{a_i}^D = \sum_{(a_i,j) \in V_a} \hat{x}_{a_i,j}^U$$

would yield an alternative formulation.

A different reformulation of the original IP is motivated by the fact that the sometimes large size of the formulations is primarily due to the number of assignment variables $y_{i,j}$. To reduce the size of the resulting formulations, we therefore consider an approach which projects out these assignment variables. The idea behind this approach is that each class can be viewed as a potential "job" to be scheduled in the interval $1, \ldots, n$; each job has both a release time and a deadline. For instance $x_i^N = 1$ would imply that we have to schedule a job with release time e_i and deadline i. With this information, we could replace the requirement that each flight is assigned a slot in its assigned class with the requirement that each potential set of classes ("jobs") can be scheduled.

This less stringent requirement can be formalized by stating that for each interval $j, \ldots, k$, the number of jobs that both arrive and are due in that interval should be no more than the number of slots in that interval. Formally, this can be expressed as follows.

$$\sum_{f_i \in \mathcal{F}: U_i \subseteq I} x_i^U + \sum_{f_i \in \mathcal{F}: N_i \subseteq I} x_i^N + \sum_{f_i \in \mathcal{F}: N_i \subseteq I} x_i^D \leq k - j + 1, \tag{6.6}$$

for all $I = \{j, \ldots, k\}$, $1 \leq j \leq k \leq n$. The validity of this reformulation is easily shown by induction on the number of slots. Consequently, the second formulation can be obtained by eliminating the assignment variables (and all the constraints they appear in) in the original formulation, and adding constraint 6.6. We note that feasible solutions to the resulting formulation do not represent an assignment of flights to slots; rather the constraints ensure that a feasible assignment of flights to slots exists. To obtain an assignment of flights to slots, we use a separate assignment problem.

We evaluated the performance of the resulting formulations, using the two objective functions described above. The results are shown in Appendix B. In

Appendix B, formulation 1 represents the standard formulation of the mediation problem shown in Figure 6.10. Formulation 2 represents the formulation that is obtained by using constraints 6.3, 6.4, and 6.4, while formulation 3 represents the formulation that would result from the elimination of the assignment variables. The results shown in Appendix B clearly show the impact of using alternate formulation. Formulation 2 is stronger as indicated by the reduced value of the LP relaxation value (in particular if the objective function used is to maximize the number of flights moved down). However, the increased size of the IPs can lead to an increase in solution time. The results for formulation 3 clearly show the reduction in the size of the formulations. While the elimination of assignment variables does not strengthen the formulation, the reduced size may lead to significantly lower computation times. To conclude, we note that both reformulations are complementary, that is, they can be applied simultaneously. We have not pursued this possibility.

Slot Trading Dynamics

So far, the slot trading problem we considered assumed that an airline would agree to any amount of additional delay for a flight in return for a reduction in delay that would make another flight arrive on time. In other words, an airline would submit all “at-least, at-most” offers that would improve its on-time performance. As a final step in our analysis, we now consider the benefits that may be obtained if airlines only submit smaller sets of offers.

To analyze this situation, we consider cases in which an airlines may no longer accept an arbitrary increase in delay in return for an increase in on-time performance. The approach we follow is based on the use of a slightly more complex value w_{fs} of assigning flight f to slot s (in the formulation shown in Figure 6.4). More specifically, we define the coefficients w_{fs} as

$$w_{fs} = \begin{cases} M & \text{if } (t_s - e_f) \leq 15, \\ \max(120 - t_s, 0) & \text{otherwise.} \end{cases}$$

As before, a large value is associated with on-time performance. However, in this case additional delays to a flight that is not arriving on time will incur a cost (reduction in value). We assume that the value decreases linearly in the additional delay; there is no additional decrease in value for flights with excessive delays (i.e. 2 hours or more).

Given this value function for each flight, the number of offers submitted by an airline can be limited by the specification of aspiration levels $v_a (a \in \mathcal{A})$. Intuitively, an aspiration level v_a signifies that an airline will only agree to delay a flight (in return for an increase in on-time performance) if the increase in cost is no more than the aspiration level v_a. As such, the aspiration levels define the classes D_i in the formulation shown in Figure 6.10, that is,

$$D_i(v_a) = \{j \in \mathcal{S} : j > i \wedge w_{ii} - w_{ij} \leq v_a\} \quad \text{for all } a \in \mathcal{A},\ f_i \in \mathcal{F}_a.$$

The use of aspiration levels provides a simple yet intuitive way to analyze the impact of limiting the number of offers proposed; by varying the aspiration levels, the number of proposed offers can be adjusted.

Figure 6.13 shows the relative improvements in on-time performance as a function of the aspiration level, for the same set of GDPs in Boston that was used before. Each curve in Figure 6.13 corresponds to a single GDP. These results show that even for smaller aspiration levels (e.g. 0 to 30 minutes), considerable improvements in on-time performance can be obtained. It is interesting to note

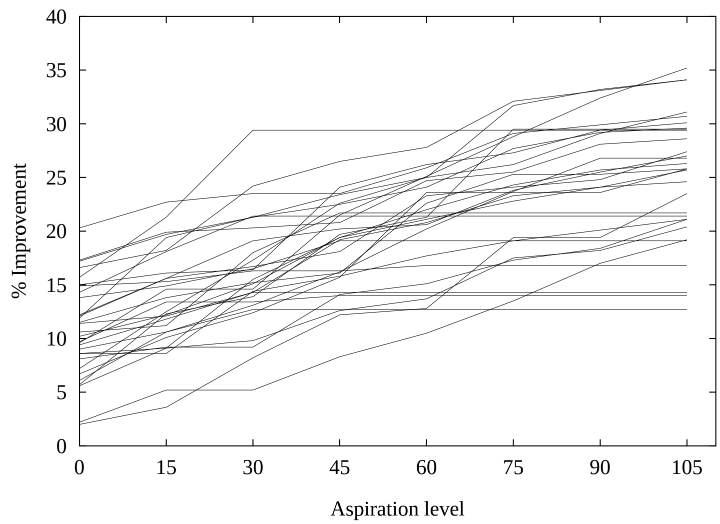


Figure 6.13: On-time Performance Improvements by Aspiration Levels

that even with aspiration levels of 0 (in this case, an airline will only allow flights with excessive delays to be moved down), the improvements can be substantially higher than those obtained by the Compression Algorithm. Overall, these results indicate that slot trading may be beneficial even if with limited numbers of offers.

6.4.2 Passenger Delay Costs

The second case study considers the situation in which each airline's objective is to minimize (total or average) passenger delays. Total passenger delay is defined as the total number of passengers in a flight multiplied by the delay of that flight summed over all flights. As such, the trade-offs that occur in this case will typically consider the overall benefits of delaying flights with few passengers in return for delay reductions for more heavily loaded flights. As in the case of on-time performance objectives, minimizing passenger delays plays an important role in the airlines' decision-making process during GDPs ([41], [79], [54]).

The use of passenger delay minimization as airline objectives again allows us to represent the airline's performance function as an assignment problem. Using the IP formulation shown in Figure 6.4, passenger delays can be represented by using coefficients

$$w_{fs} = -p(f)(t_s - e_f),$$

where $p(f)$ represents the number of passengers on flight f (note the coefficient is negative so as to represents costs). Another, somewhat more course grained approach, could be to define for each flight a set of *critical times* $k_1(f), \ldots, k_{max}(f)$ and use coefficients

$$w_{fs} = \begin{cases} 0 & \text{if } (t_s - e_f) \leq k_1(f), \\ -k_i(f)p(f) & \text{if } k_i(f) < (t_s - e_f) \leq k_{i+1}(f),\ 1 < i < max. \end{cases}$$

The use of critical times leads to the staircase cost structure shown in Figure 6.7. This applies in situations where the exact slot assignment is less important than the specific interval in which the flight is assigned (see also [41], where critical arrival times are based on the impact on downstream delays).

Whereas the on-time performance objective allowed us to represent the offers proposed using a single constraint, the increased number of possible trade-offs (i.e. flight classes) that may occur in the minimization of passenger delay costs make this approach less applicable. To analyze the benefits that could be achieved by slot trading, we therefore used the slot trading model shown in Figure 6.9. If airlines use passenger delay minimization as their objectives, the potential number of offers could substantial: in principle, an airline might submit any "at least, at most" offer that would reduce its passenger delay cost. That is, for any two flights $f_i, f_j \in F_a$ and any two slots d, u such that $i < d$, $u < j$ and

$$w_{f_i,d} - w_{f_i,i} < w_{f_j,j} - w_{f_j,u},$$

the offer (i, d, j, u) might be considered desirable. However, many of these offers may be considered redundant. Suppose, for instance, that an airline submits two offers (i, d_1, j, u) and (i, d_2, j, u) with $d_1 < d_2$. That is, an airline would be willing accept at most slot d_1 or at most slot d_2 for flight f_1, in return for giving flight f_j at least slot u. Clearly, the first offer, (i, d_1, j, u), will be redundant. A similar situation occurs with two offers (i, d, j, u_1) and (i, d, j, u_2) with $u_1 < u_2$. That is, an airline would be willing accept at most slot d for flight f_1, in return for giving flight f_j at least slot u_1 or at least slot u_2. Thus, any of these redundant trades can be removed from consideration. In addition, we note that the use of critical arrival times further allows us to reduce the number of potential offers. Specifically, we can define the classes in the IP formulation to correspond to the latest slot with a given value in the staircase cost structure (i.e. we could define an at-least class corresponding to all slots with a delay less than 15 minutes, etc.). To reduce the number of offers when using the standard passenger delay objective, we applied the same definition of classes in this case. As such, the potential offers evaluated will be the same for both objectives; however, the desirability of these offers may be different depending upon the objective function used.

Empirical Results

As a first step in our analysis, we considered the *potential* benefits that could be obtained by coordination. For our analysis, we used the same set of GDPs at Boston's Logan airport. Estimates of the number of passengers per flight were obtained by considering capacities of the various aircraft types ([16]), and assuming a load factor of 75%. For each of these GDPs, we first executed RBS

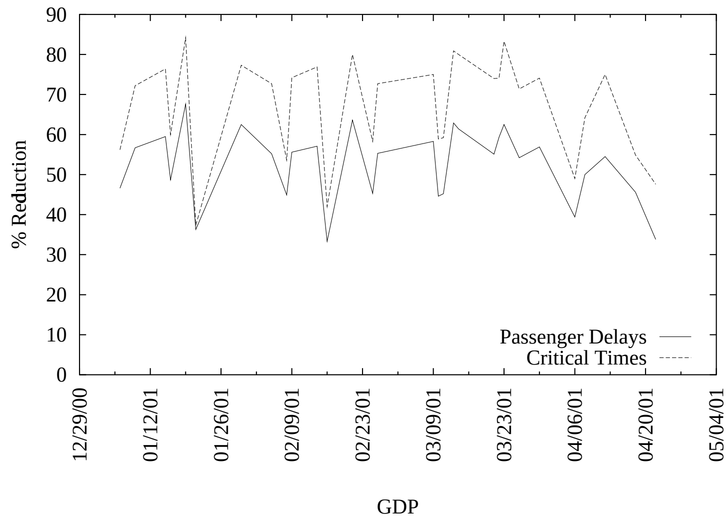


Figure 6.14: Potential Reduction in Passenger Delays

and fixed the assignment of all the flights that were exempted from the program. The resulting flight-slot assignment of the remaining (non-exempt) flights was used to determine each airline's initial allotment of slots.

Subsequently we minimized the total passenger delay cost for each individual airline, using both objective functions discussed before. The critical times we used in the second objective are $15, 30, 45, 75$, and 120 minutes. In addition, we also minimized the overall passenger delay costs, to determine a bound on the benefits that could be obtained with coordination. The potential improvements are shown in Figure 6.14, which shows the relative reduction in passenger delay costs for both objective functions. The solid line represents the improvements using the first objective, while the dashed line represents the improvements obtained using the objective which uses critical times.

Figure 6.14 shows that the potential reductions in passenger delay are substantial, ranging from 33% to 67% (the average reduction equals 52%). The results are even more pronounced when the second objective function is used; in

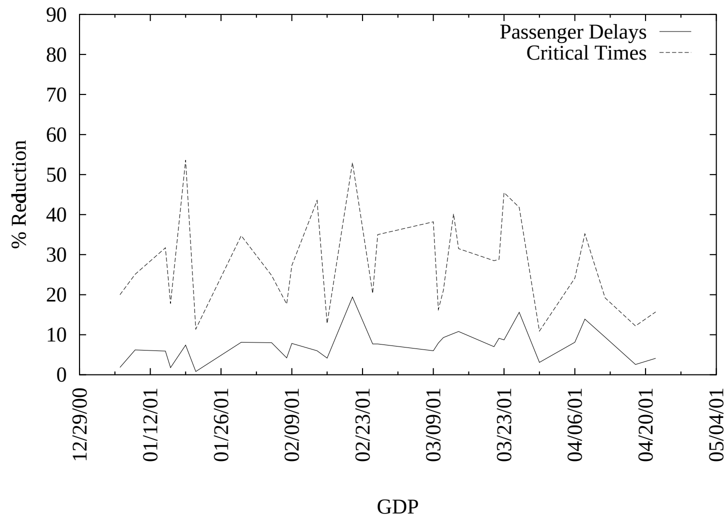


Figure 6.15: Reduction in Passenger Delay Costs using Slot Trading Potential Reduction in Passenger Delays

this case the cost reduction ranges from 37% to 84%, and the average reduction equals 67%. It should be noted, however, that the magnitude of these improvements may be somewhat misleading; typically, globally optimal allocations are usually achieved by assigning the majority of delay to General Aviation flights and commuter carriers (which use smaller aircraft). As such, these one-sided increases in delay would imply that GA flights and commuter carriers would unilaterally offer to increase their delays.

To analyze the benefits from our slot trading model, we applied the IP formulation with both objective functions, using the offers discussed above. The improvements obtained by our slot trading model are shown in Figure 6.15, which shows the relative reduction in passenger delay costs for both objective functions. The results show that the potential benefits are significantly less than those obtained using the on-time performance objective. Using the first objective, the average reduction in passenger delay is approximately 7.7% . Under the second objective, the average reduction in passenger delay cost equals 36%, which is substantial but significantly less than the global maximum. While the use of critical times in the second objective function has little impact on IP performance (the average solution time equals 12.5 seconds), the use of the passenger delay objective causes a significant increase in solution times (120 seconds on average[5]). In addition, we determined the number of flights that were moved up (i.e. assigned to an earlier class): using the first objective, an average of 83 flights were moved up, while using the second objective an average of 93.5 flights were moved up. Overall, it appears therefore that the use of "at least, at most" trades will have a much larger impact if the airlines' objective functions can be represented by a staircase objective with the use of critical times. This may actually be more realistic, as it has been shown that in many cases airline flight cost functions do have jumps associated with events such as passengers missing connections and crews timing out.

6.5 Discussion

The purpose of this chapter was to analyze the potential benefits that could be obtained by the introduction of slot trading during GDPs. Motivated by practical concerns, we considered a mediated bartering framework in which the FAA acts as a broker matching offers proposed by the airlines. Given that economic approaches do not appear to be applicable, we developed an optimization model for the mediation problem faced by the FAA for the case where airlines can specify "at least, at most" offers.

[5] the number of branch-and-bound nodes was limited to 250.

Experimental results in two case studies that use different models of airline decision-making showed considerable promise. Under a basic model of airline decision-making, in which each airline's objective is to maximize its on-time performance, slot trading yielded significant benefits. Moreover, the experiments showed that the IP formulation is highly efficient and that significant benefits can be obtained even when the number of proposed offers is limited. In the case where airlines use passenger delay minimization as their objectives, the relative benefits appeared to be substantially less. Nevertheless, in the case where passenger delay costs increase in a limited number of stages, the slot trading mechanism still resulted in substantial benefits.

Chapter 7

Conclusions

This dissertation has been motivated by the fairness considerations that arise in a collaborative air traffic management environment. With the advent of CDM, the equitable allocation of airspace capacity has become increasingly important, and a key concern in procedural modifications and enhancements. The objectives of this dissertation have been threefold: (1) to formalize and to analyze potential fairness concepts that may apply during GDPs, (2) to study the impact of program dynamics and propose methods to manage them, and (3) to consider the potential benefits of increased airline control by the introduction of slot trading mechanisms. The remainder of this chapter summarizes the conclusions for each of these topics, and outlines potential areas for future research.

Fair Slot Allocation Concepts

Chapter 4 discussed potential approaches to the fair allocation of arrival slots during a GDP, based on the CDM-introduced notion that airlines are entitled to shares of the capacity based on their original flight schedules. First, we considered approaches that rely on the equitable distribution of delays, using both multi-objective optimization methods and cost-sharing methods (e.g. the Shapley value). While equity is commonly measured in terms of the resulting delays, we saw that methods that are explicitly based on the assignment of delays appeared to have several drawbacks. In particular, such allocation schemes are not invariant if the allocation is decomposed into stages, which often occurs during GDPs because of weather uncertainty. In light of these drawbacks, we studied an axiomatic approach to the allocation of slots, in which we posed certain desirable properties as axioms that an allocation rule would have to satisfy. We showed that, under certain intuitive axioms, any such rule can be characterized by an underlying priority standard over the scheduled arrival times. While this provides a strong basis for the RBS procedure, the result also indicates other possibilities. In particular, we identified the so-called proportional random assignment mechanism as a potential alternative. We argued that RBS and proportional random assignment are based on fundamentally different interpretations of the entitlement airlines derive from their flight schedules, and that proportional random assignment might be more applicable in situations where significant numbers of flights are bound to be cancelled. Surprisingly enough, however, empirical results do not appear to indicate major differences between RBS and proportional random assignment. While significant differences in the delay may occur at any given day, on the aggregate there appear to be no systematic biases. Moreover, our empirical analysis showed that probabilistic allocation schemes introduce substantial variance in the delay, which would likely be unacceptable for the airlines.

The overall objective in Chapter 4 was to provide a theoretical basis for potential concepts of fairness in the allocation of slots during GDPs. An attractive area of further research could therefore be to find an axiomatic characterization of the proportional random assignment method. An axiomatic characterization of this method could formally clarify the difference in entitlement with the Shapley value. Another area of further research could be the investigation of allocation rules that are not necessarily collusion-proof. Of particular interest would be methods similar to the (round-robin like) uniform gains method (see [86]); such methods may prove to be useful in situations where airlines could strategically submit their demands.

Another possibility would be to incorporate broader policy objectives into the slot allocation process. In the axiomatic approach discussed in Chapter 4, we distinguished flights by a single attribute, the scheduled time of arrival. However, it would also be possible to prioritize flights according to a richer set of attributes. Examples could be aircraft size, number of passengers, or even historical data quality. It should be noted, however, that such methods would likely require restrictions to the airlines' subsequent substitution process to maintain the validity of the original priorities. For this reason, it is unlikely that airlines would agree with such an approach.

Fair Slot Allocation: Equity as near may be

Chapter 5 considered methods that can be used to approximate fair shares in situations where the ideal may not be attainable. While these methods could also be used when the allocation schemes proposed in Chapter 4 lead to an unacceptably high level of variance, the primary focus of Chapter 5 was the management of program dynamics, that is, the flight cancellations and delays that occur during the course of a GDP and the timing of GDPs (which leads to flight exemptions). Based on similarities with apportionment and balanced JIT scheduling problems, we first discussed optimization models that can be used to reallocate slots when flight cancellations and delays cause the current schedules to be infeasible and/or suboptimal. One approach in particular, which minimizes the deviations from predefined ideal positions in the schedule, provided a potentially attractive alternative to the Compression Algorithm. The resulting procedure can be interpreted as a form of rerationing according to given sets of airline priorities and, as such, unifies both RBS and Compression.

Subsequently, we showed that the time at which a GDP is implemented can have a significant impact on the distribution of delays. Adjustments of the previously described approaches introduced an alternative method to manage flight exemptions. Empirical results showed that the resulting optimization models significantly reduce the systematic biases that exist under current procedures. Further analysis indicated that the use of these methods does not significantly change the airlines' distribution of delays. The use of these methods did introduce a systematic reduction in the delays of smaller aircraft. However, due to the relative small number of smaller aircraft, the overall impact of this reduction appeared to be limited. Given these results, we believe that the optimization-based approach presents a potentially attractive alternative to the methods that are currently used to manage exemptions.

Finally, we considered the application of these optimization schemes with alternative definitions of the ideal shares. We analyzed two possible alternatives: the ideal shares that follow from the proportional random assignment scheme and the "standard" proportional shares used in balanced jit problems. The second definition is similar to the notion of entitlement that underlies the Shapley value. While the use of proportional random assignment shares yielded little difference from RBS, the use of proportional shares had a significant impact on the resulting allocations. We conjectured that this impact might be due to strategic behavior by GA operators. While this conjecture requires additional analysis, it indicates an attractive feature of using proportional shares: since these shares only depend on the numbers of flights for each airline, they reduce current incentives to misrepresent arrival times. Additional research in this area could consider the impact of situations where scheduled carriers artificially increase their demand.

Overall, the methods proposed in Chapter 5 provided a general framework for the (re)allocation of slots that allows the incorporation of many practical "constraints". The notion that schedules are balanced according to appropriately defined quota could also be applied to address other issues, such as the management of "pop-up" flights and of en-route resource allocation problems.

Slot Trading during GDPs

Finally, Chapter 6 explored the potential benefits of increased coordination during GDPs. To do this, we introduced a general bartering framework in which airlines may submit offers to trade slots to the FAA, which acts as the central coordinator. Given the apparent limitations of economic approaches that were discussed, we further proposed an optimization model for the FAA's mediation problem. This model generalizes current (and proposed) slot exchange procedures in that it allows airlines to submit so-called "at-least, at-most" offers, which may be viewed as tradeoffs between pairs of flights and that are motivated by operationally significant delay levels.

To analyze the potential benefits of this approach, we considered two case studies that use different models of airline decision-making. Empirical results using historical GDP data showed considerable promise. Under a basic model of airline decision-making, in which airlines aim to maximize their on-time performance, slot trading yielded significant benefits. The IP formulation of the mediation problem is highly efficient, and various alternative formulation further improve IP performance. Moreover, significant benefits could be obtained even when the number of proposed offers was limited. In the case where airlines use passenger delay minimization as their objectives, the relative benefits appeared to be less. Nevertheless, in the case where passenger delay costs increase in a limited number of stages, the slot trading mechanism still resulted in substantial benefits.

The use of slot trading during GDPs offers numerous areas for further research. First, the performance differences under alternate passenger delay cost objectives merit further analysis. This can include the possibility of changing the sets of offers evaluated (and consequently) proposed, by increasing the number of classes in our formulation of the mediation problem. In this case, the increased size might degrade IP performance, which might make the use of alternative formulations more attractive. More generally, it may be desirable to study more complex models of airline decision-making. Such a study should not only involve the performance of the resulting mediation problem, but also include the evaluation of offers by airlines.

Appendix A

Proofs

A.1 Proof of Proposition 4.3.3

A probabilistic allocation rule X associates with every possible combination of capacities $c \in \{0,1\}$ and set of flights $F \subseteq \mathcal{F}$ a random allocation $X(\tau_f, P(F,c))$ that is both feasible and efficient. In principle, X could be specified by enumeration, associating an allocation with each possible combination (F,c). If we impose impartiality and consistency, however, the allocation rules can be characterized succinctly. To prove this, we first analyze the case where the capacity equals a unit vector e_j and consider for any allocation rule X the allocation $X(\tau_{\mathcal{F}}, P(\mathcal{F}, e_j))$. By the feasibility requirement, we have

$$X(\tau_{\mathcal{F}}, P(\mathcal{F}, e_j))_{f,j'} = 0 \quad \text{for } j \neq j' \text{ and } \quad X(\tau_{\mathcal{F}}, P(\mathcal{F}, e_j))_{f,j} = 0 \quad \text{if } \tau_f > j,$$

and impartiality implies that

$$X(\tau_{\mathcal{F}}, P(\mathcal{F}, e_j))_{f,j} = X(\tau_{\mathcal{F}}, P(\mathcal{F}, e_j))_{f',j} \quad \text{if } \tau_f = \tau_{f'}.$$

Suppose now that each eligible flight receives a positive share of slot j, that is, $X(\tau_{\mathcal{F}}, P(\mathcal{F}, e_j))_{f,j} > 0$ if $\tau_f \leq j$. Then, consistency implies that $X(\tau_{\mathcal{F}}, P(\mathcal{F}, e_j))$

completely specifies the allocation for all $F \subseteq \mathcal{F}$.

Lemma A.1.1. *Let X be any impartial, consistent allocation rule, and let e_j represent a unit capacity vector. Let $X(\tau_{\mathcal{F}}, P(\mathcal{F}, e_j))_{f,j} = \lambda_i^j$ when $\tau_f = i$, such that $\lambda_i^j > 0$ if $i \leq j$ and $\lambda_i^j = 0$ if $i > j$. Then, for every $F \subseteq \mathcal{F}$ we have*

$$X(\tau_F, P(F, e_j))_{f,j} = \frac{\lambda^j_{\tau_f}}{\sum_{f \in F} \lambda^j_{\tau_f}}. \tag{A.1}$$

Proof. The proof follows by induction on the number of flights. Clearly, equality A.1 holds if $F = \mathcal{F}$. Suppose now A.1 holds for a given set of flights F and consider the set $F - \{f\}$ that results from removing flight f. Then, by consistency we have

$$X(\tau_F, P(F, e_j))_{f',j} = (1 - \frac{\lambda^j_{\tau_f}}{\sum_{f \in F} \lambda^j_{\tau_f}})X(tau_{F-\{f\}}, P(F - \{f\}, e_j))_{f',j}$$

for all $f' \in F - \{f\}$, and therefore

$$X(\tau_{F-\{f\}}, P(F - \{f\}, e_j))_{f',j} = \frac{\sum_{f \in F} \lambda^j_{\tau_f}}{\sum_{f \in F-\{f\}} \lambda^j_{\tau_f}} X(\tau_F, P(F, e_j))_{f',j} = \frac{\lambda^j_{\tau_{f'}}}{\sum_{f \in F-\{f\}} \lambda^j_{\tau_f}}$$

for all $f' \in F - \{f\}$. $\square$

Thus, if (1) we restrict the possible capacity profiles to unit vectors and (2) the possible allocation rules to those that assign each eligible flight a positive share, a consistent and impartial allocation rule can can be characterized by a set of weights $\lambda_i^j (0, \leq i, j < n)$. Intuitively, these weights assign a relative priority to each of the flights: for instance, if $\lambda^j_i / \lambda^j_{i'} = k$ then a flight whose oag equals i will always get a share of slot j that is k times the share of a flight whose oag is i'.

Proposition 4.3.3 considers the possibility that not every eligible flight is assigned a positive share. In this case, an impartial and consistent allocation rule can be characterized by the combination of a set of weights and a preordering $\succeq_j$ over the OAG times .

Proposition A.1.2. *Let X be a consistent, impartial allocation rule, let e_j represent a unit capacity vector whose capacity at slot j equals 1, and let τ_F be any demand profile. Then, there exists a set of weights $\lambda_i^j (0 \leq i \leq j)$ and a weak ordering*[1] *$\succeq_j$ over the OAG times $0 \leq i \leq j$ such that*

$$X(\tau_F, P(F, e_j))_{f,j} = \frac{\lambda_{\tau_f}^j}{\sum_{g \in F} \lambda_{\tau_g}^j} \quad \text{if } \tau_f \succeq_j \tau_{f'} \text{ for all } f' \in F, \text{ and}$$

$$X(\tau_F, P(F, e_j))_{f,j} = 0 \quad \text{otherwise.}$$

Proof. Consider again the set of flights $\mathcal{F}$, and suppose as before that $X(\tau_{\mathcal{F}}, P(\mathcal{F}, e_j))_{f,j} = \lambda_{\tau_f}^j$. Now we can partition the eligible oag times into two classes: $S^+ = \{i : 0 \leq i \leq j, \lambda_i^j > 0\}$ and $S^0 = \{i : 0 \leq i \leq j, \lambda_i^j = 0\}$. Thus, flights whose oag is in S^+ receive a positive share of slot j while flights whose oag is in S^0 do not.

Observe that for any set of flights $F \subseteq \mathcal{F}$ which contains at least one flight whose oag is in S^+, we can still apply Lemma A.1.1. Therefore, $X(\tau_F, P(F, e_j))_{f,j} = 0$ if $\tau_f \in S^0$ and

$$X(\tau_F, P(F, e_j))_{f,j} = \frac{\lambda_{\tau_f}^j}{\sum_{f \in F : \tau \in S^+} \lambda_{\tau_f}^j}$$

if $\tau_f \in S^+$. Consequently, flights whose oag is in S^+ will always have absolute priority over flights whose oag is in S^0. Consequently, we can partially specify the set of weights λ_i^j for all $i \in S^+$ and the preordering $\succeq_j$ as $i \succeq_j i'$ for all $i \in S^+, i' \in S^0$ and $i \succeq_j i', i' \succeq_j i$ for all $i, i' \in S^+$.

To fully characterize the allocation rule, we still have to consider the sets of flight F in which no flights whose oag is in S^+ are present. But this can be done

[1] a weak ordering or preordering is an ordering relation $\succeq_P$ that is connected (i.e. $j \succeq_P j'$ or $j' \succeq_P j$ or both) and transitive.

by considering the set in which all flights are present expect those whose oag is in S^+. Thus, repeating this process until all flights have been assigned a weight will eventually give the desired result. $\square$

Thus, if capacity profiles are restricted to unit vectors the allocation rules that are impartial and consistent can be characterized by a preordering which partitions the oag times into priority classes, and a set of weights which specifies the relative priorities within each class. Furthermore, since the allocations under more general capacity profiles can be determined by reducing the recursion in the consistency axiom to the unit capacity case, the following corollary follows.

Corollary A.1.3. *A probabilistic allocation rule that is consistent and satisfies equal treatment of equals is characterized by the n sets of weight λ_i^j and preorderings $\succeq_j$, which define the allocation when the capacity profile is a unit vector.* $\square$

It is an open question whether the reverse also holds.

A.2 Proof of Theorem 4.3.8

Proposition 4.3.3 shows that impartial, consistent allocation rules can be characterized by a set of weights and preorderings, which specify how the rule allocates each individual slot. As such, a flight f_1 could have priority over flight f_2 at slot j, but f_2 could have priority over f_1 at slot $j+1$. The consequence of time independence is that the weights and preorderings are identical at each slot. Thus, an impartial, consistent, and time independent allocation rule is characterized by a single set of weights $\lambda_i \in \mathbb{R}_+^n$ and a preordering $\succeq_Q$.

The following lemma shows that the addition of the composition axiom forces the weights of any two flights that are in the same priority class to be equal.

Lemma A.2.1. *Let X be any impartial, consistent, and time independent allocation rule, which is characterized by the set of weights $\lambda_i \in \mathbb{R}^n_+$ and the preordering $\succeq_Q$. If X satisfies composition, then*

$$\lambda_i = \lambda_j \quad \textit{for all } i, j \textit{ such that } i \succeq_Q j \textit{ and } j \succeq_Q i.$$

Proof. Assume without loss of generality that $i < j$. The proof follows by considering a set of three flights $F = \{f_1, f_2, f_3\}$ with $\tau_{f_1} = \tau_{f_2} = i$ and $\tau_{f_3} = j$, and a capacity profile $c = e_j + e_{j+1}$. For this situation, we can calculate the resulting allocations both by applying the consistency axiom and by applying the composition axiom.

First, we calculate the slot shares $X(\tau_F, P(F, c))$ by applying the consistency axiom. Applying the consistency axiom by first assigning flight f_1 to a slot yields

$$\begin{aligned}
&X(\tau_F, P(F,c))_{f',j'} \\
&= X(\tau_F, P(F,c))_{f_1,j} X(\tau_{\{f_2,f_3\}}, P(\{f_2, f_3\}, e_{j+1}))_{f',j'} \\
&+ X(\tau_F, P(F,c))_{f_1,j+1} X(\tau_{\{f_2,f_3\}}, P(\{f_2, f_3\}, e_j))_{f',j'} \\
&+ (1 - X(\tau_F, P(F,c))_{f_1,j} - X(\tau_F, P(F,c))_{f_1,j}) X(\tau_{\{f_2,f_3\}}, P(\{f_2, f_3\}, c))_{f',j'},
\end{aligned}$$

for $f' = f_1, f_2$, $j' = j, j+1$. Solving these equations (using the previous results) leads to the following slot share values

$$X(\tau_F, P(F,c))_{f_1,j}, X(\tau_F, P(F,c))_{f_2,j} = \frac{\lambda_i}{2\lambda_i + \lambda_j},$$

$$X(\tau_F, P(F,c))_{f_3,j} = \frac{\lambda_j}{2\lambda_i + \lambda_j},$$

$$X(\tau_F, P(F,c))_{f_1,j+1}, X(\tau_F, P(F,c))_{f_2,j+1} = \frac{\lambda_i + \lambda_j}{\lambda_i + 2\lambda_j} \frac{\lambda_i}{2\lambda_i + \lambda_j} + \frac{\lambda_j}{\lambda_i + 2\lambda_j}$$

$$X(\tau_F, P(F,c))_{f_3,j+1} = \frac{\lambda_i + \lambda_j}{\lambda_i + 2\lambda_j} \frac{\lambda_j}{2\lambda_i + \lambda_j} + \frac{\lambda_j}{\lambda_i + 2\lambda_j}.$$

Alternatively, the allocation $X(\tau_F, P(F,c))$ can also be obtained by applying the composition axiom,e.g.

$$\begin{aligned}
X(\tau_F, P(F,c)) = X(\tau_F, P(F,e_j)) & \\
& + X(\tau_F, P(F,e_j))_{f_1,j} X(\tau_{\{f_2,f_3\}}, P(\{f_2,f_3\}, e_{j+1})) \\
& + X(\tau_F, P(F,e_j))_{f_2,j} X(\tau_{\{f_1,f_3\}}, P(\{f_1,f_3\}, e_{j+1})) \\
& + X(\tau_F, P(F,e_j))_{f_3,j} X(\tau_{\{f_1,f_2\}}, P(\{f_1,f_2\}, e_{j+1})).
\end{aligned}$$

Solving these equations gives the following slot shares

$$\begin{aligned}
X(\tau_F, P(F,c))_{f_1,j+1} &= \frac{\lambda_i}{\lambda_i + \lambda_j} X(\tau_F, P(F,c))_{f_2,j} + \frac{1}{2} X(\tau_F, P(F,c))_{f_3,j}, \\
X(\tau_F, P(F,c))_{f_2,j+1} &= \frac{\lambda_i}{\lambda_i + \lambda_j} X(\tau_F, P(F,c))_{f_1,j} + \frac{1}{2} X(\tau_F, P(F,c))_{f_3,j}, \\
X(\tau_F, P(F,c))_{f_3,j+1} &= \frac{\lambda_j}{\lambda_i + \lambda_j} (X(\tau_F, P(F,c))_{f_1,j} + X(\tau_F, P(F,c))_{f_2,j}).
\end{aligned}$$

Clearly, a probabilistic allocation rule that satisfies both consistency and composition should yield identical shares under both derivations. Thus, the following equality should hold for methods that satisfy both axioms.

$$\begin{aligned}
& \frac{\lambda_i + \lambda_j}{\lambda_i + 2\lambda_j} \frac{\lambda_i}{2\lambda_i + \lambda_j} + \frac{\lambda_j}{\lambda_i + 2\lambda_j} \\
= & \\
& \frac{\lambda_i}{\lambda_i + \lambda_j} \frac{\lambda_i}{2\lambda_i + \lambda_j} + \frac{1}{2} \frac{\lambda_j}{2\lambda_i + \lambda_j}.
\end{aligned}$$

This equation can be rewritten as

$$\frac{\lambda_i}{\lambda_i + \lambda_j} \left(\frac{\lambda_i}{\lambda_i + \lambda_j} - 1 \right) \left(\frac{\lambda_i}{\lambda_i + \lambda_j} - \frac{1}{2} \right) = 0.$$

Since $\lambda_i, \lambda_j > 0$ (because i, j are in the same priority class), it follows that $\frac{\lambda_i}{\lambda_i + \lambda_j} = \frac{1}{2}$ and therefore that $\lambda_i = \lambda_j$. $\square$

Theorem 4.3.8 follows almost immediately from Proposition 4.3.3 and Lemma A.2.1.

Theorem A.2.2. *Let c be any capacity profile and F be any set of flights. Then, for any probabilistic allocation rule X that is impartial, consistent, time independent, and satisfies composition, there is a priority standard Q such that*

$$X(\tau_F, P(F,c)) = \sum_{x \in Q(F,c)} \frac{1}{|Q(F,c)|} x.$$

Proof. By Proposition 4.3.3, an impartial, consistent allocation rule can be characterized by a set of weights and preorderings, which specify how the rule allocates each individual slot. Time independence further implies that an allocation rule can be specified by a single set of weights $\lambda_i \in \mathbb{R}^n_+$ and a preordering $\succeq_Q$. By Lemma A.2.1, we furthermore have that within each priority class the weights λ_i are equal.

Consequently, a probabilistic allocation rule that satisfies all axioms assigns each each slot according to the preordering $\succeq_Q$. Since the allocation can be decomposed into stages, it follows that the possible allocations correspond to the priority method based on Q. Finally, since flights within a priority class are selected with equal probability, we have that each of these allocations is equiprobable. $\square$

A.3 Proof of Theorem 5.2.1

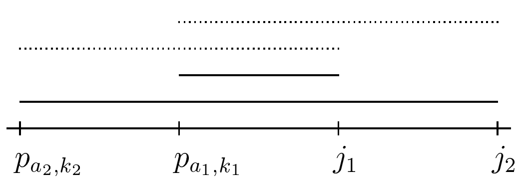
Theorem A.3.1. *A solution x obtained by the greedy algorithm shown in Figure 5.7 is an optimal solution for the IP formulation shown in Figure 5.6.*

Proof. The proof follows using an interchange argument. Suppose x is not an optimal solution to the IP formulation. Then, there exist another solution y

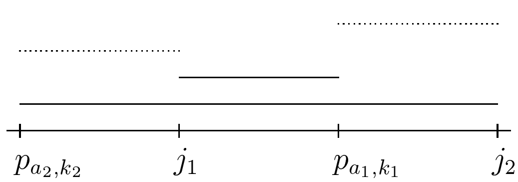
which is optimal and differs from x in at least one position. Let j_1 be the first position at which the allocations differ, and let a_1, a_2, k_1, k_2 be such that $x_{a_2,k_2,j_1} = 1$ and $y_{a_1,k_1,j_1} = 1$. By construction of the greedy algorithm, we know that $p_{a_2,k_2} < p_{a_1,k_1}$ (note that the strict inequality is due to the fact that all ideal positions are different). Moreover, we also know that $y_{a_2,k_2,j_2} = 1$ for some $j_2 > j_1$.

As a consequence, we can separate the following six cases, which are depicted graphically below. In each of these cases, the solid lines represent the differences between the ideal position and the actual assignment in the optimal solution y, and the dotted lines represent the differences that would result from an interchange of the assignments (that is, if we let $y_{a_1,k_1,j_2} = 1$ and $y_{a_2,k_2,j_1} = 1$).

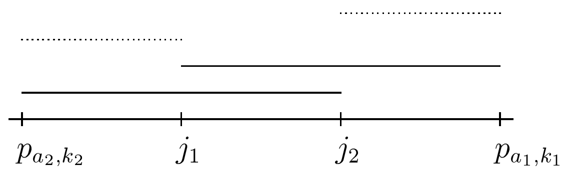
- $p_{a_2,k_2} < p_{a_1,k_1} \leq j_1 < j_2$,

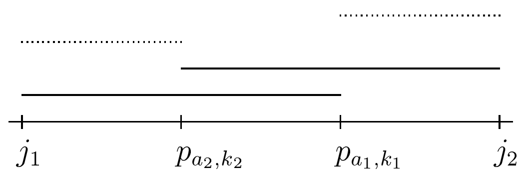


- $p_{a_2,k_2} \leq j_1 < p_{a_1,k_1} \leq j_2$,



- $p_{a_2,k_2} \leq j_1 < j_2 < p_{a_1,k_1}$,
- $j_1 \leq p_{a_2,k_2} < p_{a_1,k_1} \leq j_2$,
- $j_1 \leq p_{a_2,k_2} \leq j_2 < p_{a_1,k_1}$
- $j_1 < j_2 \leq p_{a_2,k_2} < p_{a_1,k_1}$.





It follows by inspection that in each of these cases,

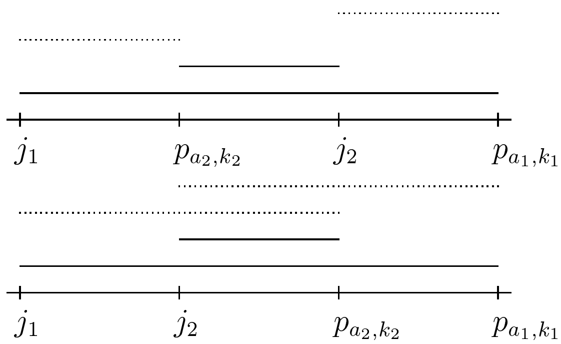
$$(p_{a_1,k_1} - j_2)^2 + (p_{a_2,k_2} - j_1)^2 < (p_{a_2,k_2} - j_1)^2 + (p_{a_1,k_1} - j_2)^2.$$

Thus, interchanging the positions of the two flights in y will yield an allocation with lower cost and, by construction, any such allocation is feasible. This, however, contradicts the assumption that y is an optimal solution, which completes the proof. □

A.4 Proof of Theorem 5.3.1

Theorem A.4.1. *A solution x obtained by the greedy algorithm shown in Figure 5.7 is an optimal solution for the IP formulation shown in Figure 5.6 with constraints 5.3 added.*

Proof. The proof again follows using an interchange argument. Suppose x is not an optimal solution to the IP formulation. Then, there exist another solution y which is optimal and differs from x in at least one position. Let j_1 be the first position at which the allocations differ, and let a_1, a_2, k_1, k_2 be such that $x_{a_2,k_2,j_1} = 1$ and $y_{a_1,k_1,j_1} = 1$. It follows that $L_{a_1,j_1} < k_1$ and $L_{a_2,j_1} < k_2$, that is,



neither of the flights are due at time j. By construction, we therefore know that $p_{a_2,k_2} < p_{a_1,k_1}$. Moreover, we also know that $y_{a_2,k_2,j_2} = 1$ for some $j_2 > j_1$ and therefore that $L_{a_2,j_2} \leq k_2$.

Suppose now that $L_{a_1,j_2} \leq k_1$. In that case, we can interchange the flights using the argument given in Theorem 5.2.1 (e.g. the interchange will not violate the lower bounds). Now consider the case where $L_{a_1,j_2} > k_1$. Thus, flight k_1 of airline a_1 is due before time j_2, but after time j_1. Now let us look at all the flights $f_{a',k'}$ occupying the positions $j_1+1,\ldots,j_2-1$, and suppose all these flights were due before time j_2, e.g., $L_{a',j_2} > k'$. This implies that $|\{f \in F^e : eta_f = j'\}| > 1$ for at least one $j' \in j_1+1,\ldots,j_2-1$, since we know that $L_{a_1,j_2} > k_1$. This, however, would contradict our assumption and therefore there is at least one $j' \in j_1+1,\ldots,j_2-1$ such that $y_{a',k',j'} = 1$ and $L_{a',j_2} \leq k'$. Using the arguments from theorem 5.2.1, we can therefore interchange the assignments of f_{a_2,k_2} and $f_{a',k'}$ in y without increasing the cost. This yields an new assignment y where the distance between f_{a_1,k_1} and f_{a_2,k_2} has decreased. Thus, by repeating this argument we would eventually be able to interchange the flights such that f_{a_2,k_2} would be assigned to j_1, which shows that x is an optimal solution to the IP. $\square$

Appendix B

Slot Trading Model Results

This appendix contains empirical results for the formulation discussed in Section 6.4.1. As stated before, formulation 1 represents the standard formulation of the mediation problem shown in Figure 6.10. Formulation 2 represents the formulation that is obtained by using constraints 6.3, 6.4, and 6.4, while formulation 3 represents the formulation that would result from the elimination of the assignment variables. The tables show the results by solving the various formulations for a number of GDPs using Cplex 7.1 on a Sun Ultra 10 workstation. “Rows” and “Cols” represent the number of constraints resp. variables. “Nodes” represent the number of branch-and-bound nodes, “Time” denotes the solution times in seconds, “LP” represents the optimal value of the LP relaxation, and “Opt” the value of the best solution value found. In all experiments, we limited the number of branch-and-bound nodes to 1000.

Table B.1: Formulation 1(OV: Maximize number of flights moved up)

GDP	Rows	Cols	Nodes	Time(s.)	LP	Opt
bos01-06-01	639	17046	0	6.2	36	36
bos01-09-01	740	19090	10	21	56	56
bos01-15-01	825	20207	0	8	76	76
bos01-16-01	668	19024	10	19	58	58
bos01-19-01	990	25871	0	4.4	103	103
bos01-21-01	661	23085	0	8.2	34	34
bos01-30-01	819	20298	0	2.5	75	75
bos02-05-01	609	12973	50	18	39	39
bos02-08-01	579	15184	7	8.6	39	39
bos02-09-01	914	25519	14	20	74	74
bos02-14-01	830	20709	1	5.2	76	76
bos02-16-01	386	9390	0	1.8	20	20
bos02-21-01	691	14472	0	1.3	71	71
bos02-25-01	729	17846	35	30	44	44
bos02-26-01	614	12711	4	3.5	53	53
bos03-09-01	946	25346	20	22	80	80
bos03-10-01	282	3581	5	1.3	26	23
bos03-11-01	398	6417	3	1.8	36	36
bos03-13-01	756	16123	1	2.1	77	77
bos03-14-01	807	20278	4	10	77	77
bos03-21-01	764	19433	1	3.9	66	66
bos03-22-01	745	17911	10	17	57	57
bos03-23-01	832	19793	0	7.4	86	86
bos03-26-01	488	8211	0	2.2	44	44
bos03-30-01	860	23254	20	22	58	58
bos04-06-01	585	18401	1	12	28	26
bos04-08-01	631	11931	11	9.8	46	46
bos04-12-01	819	20535	10	16	63	63
bos04-18-01	643	18339	10	13	24	24
bos04-22-01	528	13326	7	7.4	36	35

Table B.2: Formulation 1(OV: Maximize number of flights moved down)

GDP	Rows	Cols	Nodes	Time(s.)	LP	Opt
bos01-06-01	639	17046	0	9.8	36	36
bos01-09-01	740	19090	480	160	56	51
bos01-15-01	825	20207	1000	250	59.5	54
bos01-16-01	668	19024	1000	280	58	54
bos01-19-01	990	25871	1000	250	73	67
bos01-21-01	661	23085	11	31	34	34
bos01-30-01	819	20298	1000	230	62	53
bos02-05-01	609	12973	672	110	39	36
bos02-08-01	579	15184	0	6.4	39	38
bos02-09-01	914	25519	1000	880	72.5	67
bos02-14-01	830	20709	1000	250	74	60
bos02-16-01	386	9390	0	2	20	20
bos02-21-01	691	14472	1000	120	52	47
bos02-25-01	729	17846	30	48	44	40
bos02-26-01	614	12711	1000	100	46.5	41
bos03-09-01	946	25346	1000	440	75	70
bos03-10-01	282	3581	16	2.5	25	19
bos03-11-01	398	6417	1000	37	34	27
bos03-13-01	756	16123	1000	95	58.5	52
bos03-14-01	807	20278	1000	110	60.5	54
bos03-21-01	764	19433	1000	250	60	53
bos03-22-01	745	17911	896	230	57	54
bos03-23-01	832	19793	1000	230	62	56
bos03-26-01	488	8211	1000	42	37.5	35
bos03-30-01	860	23254	600	370	58	54
bos04-06-01	585	18401	2	11	27	25
bos04-08-01	631	11931	173	50	46	41
bos04-12-01	819	20535	1000	280	61	57
bos04-18-01	643	18339	0	8.1	24	23
bos04-22-01	528	13326	27	14	36	31

Table B.3: Formulation 2(OV: Maximize number of flights moved up)

GDP	Rows	Cols	Nodes	Time(s.)	LP	Opt
bos01-06-01	2927	23597	0	73	36	36
bos01-09-01	2557	24063	0	73	51	51
bos01-15-01	2865	25379	0	28	54	54
bos01-16-01	2146	22372	0	15	54	54
bos01-19-01	3553	32572	0	44	67	67
bos01-21-01	3615	33091	0	170	34	34
bos01-30-01	2639	24538	0	37	53	53
bos02-05-01	1952	15964	1	44	36	36
bos02-08-01	2003	19254	0	23	38	38
bos02-09-01	4451	35065	3	530	67	67
bos02-14-01	3300	27085	2	140	60	60
bos02-16-01	1159	12084	0	6.9	20	20
bos02-21-01	2143	18314	1	27	47	47
bos02-25-01	3630	25929	0	200	40	40
bos02-26-01	1764	15985	5	61	41	41
bos03-09-01	4684	37297	3	500	70	70
bos03-10-01	657	4769	3	5.7	19	19
bos03-11-01	981	7850	2	6.7	27	27
bos03-13-01	2684	21069	2	100	52	52
bos03-14-01	2803	24726	0	42	54	54
bos03-21-01	2934	24475	2	110	53	53
bos03-22-01	3052	24666	0	61	54	54
bos03-23-01	3267	26842	0	92	56	56
bos03-26-01	1208	9922	6	13	35	35
bos03-30-01	3230	29833	0	140	54	54
bos04-06-01	2301	22986	0	37	25	25
bos04-08-01	2202	17149	0	46	41	41
bos04-12-01	2632	25393	1	130	57	57
bos04-18-01	3056	24922	2	200	23	23
bos04-22-01	1948	17334	5	84	31	31

Table B.4: Formulation 2(OV: Maximize number of flights moved down)

GDP	Rows	Cols	Nodes	Time(s.)	LP	Opt
bos01-06-01	2927	23597	0	140	36	36
bos01-09-01	2557	24063	0	87	56	56
bos01-15-01	2865	25379	0	58	76	76
bos01-16-01	2146	22372	4	29	58	58
bos01-19-01	3553	32572	2	260	103	103
bos01-21-01	3615	33091	0	330	34	34
bos01-30-01	2639	24538	0	32	74	74
bos02-05-01	1952	15964	4	50	39	39
bos02-08-01	2003	19254	1	23	39	39
bos02-09-01	4451	35065	2	450	74	74
bos02-14-01	3300	27085	3	240	76	76
bos02-16-01	1159	12084	0	7.6	20	20
bos02-21-01	2143	18314	1	45	71	71
bos02-25-01	3630	25929	0	340	44	44
bos02-26-01	1764	15985	0	21	53	53
bos03-09-01	4684	37297	4	1000	80	80
bos03-10-01	657	4769	0	1.7	23	23
bos03-11-01	981	7850	0	3.9	36	36
bos03-13-01	2684	21069	2	230	77	77
bos03-14-01	2803	24726	0	43	77	77
bos03-21-01	2934	24475	0	46	66	66
bos03-22-01	3052	24666	4	300	57	57
bos03-23-01	3267	26842	0	110	86	86
bos03-26-01	1208	9922	0	6.6	44	44
bos03-30-01	3230	29833	3	270	58	58
bos04-06-01	2301	22986	0	48	26	26
bos04-08-01	2202	17149	6	160	46	46
bos04-12-01	2632	25393	5	290	63	63
bos04-18-01	3056	24922	2	370	24	24
bos04-22-01	1948	17334	0	16	35	35

Table B.5: Formulation 3(OV: Maximize number of flights moved up)

GDP	Rows	Cols	Nodes	Time(s.)	LP	Opt
bos01-06-01	691	299	0	0.44	36	36
bos01-09-01	1015	349	5	1.7	56	56
bos01-15-01	1861	353	0	2	76	76
bos01-16-01	435	277	7	0.63	58	58
bos01-19-01	2405	431	0	3.3	103	103
bos01-21-01	975	335	0	1.2	34	34
bos01-30-01	1471	358	0	1.6	75	75
bos02-05-01	822	268	1	0.76	39	39
bos02-08-01	150	178	0	0.12	39	39
bos02-09-01	1719	425	0	2	74	74
bos02-14-01	1382	383	0	1.2	76	76
bos02-16-01	52	82	0	0.01	20	20
bos02-21-01	854	296	0	0.58	71	71
bos02-25-01	1649	373	0	1.7	44	44
bos02-26-01	649	259	0	0.28	53	53
bos03-09-01	2182	438	0	2.6	80	80
bos03-10-01	160	121	0	0.07	26	23
bos03-11-01	160	150	0	0.08	36	36
bos03-13-01	1036	343	0	0.69	77	77
bos03-14-01	1091	343	0	0.83	77	77
bos03-21-01	915	330	0	0.8	66	66
bos03-22-01	1440	352	0	1.4	57	57
bos03-23-01	1716	354	0	1.9	86	86
bos03-26-01	540	205	0	0.31	44	44
bos03-30-01	1345	383	0	1.7	58	58
bos04-06-01	207	211	0	0.21	28	26
bos04-08-01	1209	314	3	1.5	46	46
bos04-12-01	1390	350	0	1.2	63	63
bos04-18-01	499	261	0	0.34	24	24
bos04-22-01	272	215	0	0.16	36	35

Table B.6: Formulation 3(OV: Maximize number of flights moved down)

GDP	Rows	Cols	Nodes	Time(s.)	LP	Opt
bos01-06-01	691	303	9	1	36	36
bos01-09-01	1015	351	32	4.4	56	51
bos01-15-01	1861	353	1000	54	59.5	54
bos01-16-01	435	279	554	4.5	58	54
bos01-19-01	2405	431	1000	89	73	67
bos01-21-01	976	356	0	1.1	34	34
bos01-30-01	1471	358	1000	40	62	53
bos02-05-01	822	269	71	4.2	39	36
bos02-08-01	181	211	0	0.13	39	38
bos02-09-01	1719	425	1000	59	72.5	67
bos02-14-01	1382	383	0	2.3	64.5	60
bos02-16-01	76	120	0	0.04	20	20
bos02-21-01	854	297	1000	13	52	47
bos02-25-01	1649	373	0	2.3	44	40
bos02-26-01	650	260	1000	8.2	46.5	41
bos03-09-01	2182	438	1000	77	75	50
bos03-10-01	170	134	0	0.09	25	19
bos03-11-01	170	159	0	0.11	34	27
bos03-13-01	1036	343	1000	21	58.5	52
bos03-14-01	1091	343	1000	24	60.5	54
bos03-21-01	915	330	1000	19	60	53
bos03-22-01	1440	353	39	8.2	57	54
bos03-23-01	1716	354	1000	60	62	56
bos03-26-01	540	205	3	0.6	37.5	35
bos03-30-01	1345	383	0	1.7	58	54
bos04-06-01	207	211	0	0.21	27	25
bos04-08-01	1209	314	6	2.4	46	41
bos04-12-01	1390	351	1000	34	61	57
bos04-18-01	503	279	0	0.32	24	23
bos04-22-01	275	227	0	0.2	36	31

BIBLIOGRAPHY

[1] Airline operational control overview. prepared by Airline Dispatchers Federation and Seagull Technology, Inc., Cooperative Research and Development Agreement 93-CRDA-0034, September 1995.

[2] M. Adams, S. Kolitz, J. Milner, and A. Odoni. Evolutionary concept for decentralized air traffic management. *Air Traffic Control Quarterly*, 4:281–306, 1997.

[3] G. Andreatta and L. Brunetta. Multiairport ground holding problem: A computational evaluation of exact algorithms. *Operations Research*, 46:57–64, 1998.

[4] G. Andreatta, A. Odoni, and O. Richetta. Models for the ground holding problem. In L. Bianco and A.R. Odoni, editors, *Large-Scale Computation and Information Processing in Air Traffic Control*, pages 125–168. Springer-Verlag, Berlin, 1993.

[5] G. Andreatta and G. Romanin-Jacur. Aircraft flow management under congestion. *Transportation Science*, 21:249–253, 1987.

[6] M.L. Balinski and N. Shahidi. A simple approach to the product rate variation problem *via* axiomatics. *Operations Research Letters*, 22:129–135, 1998.

[7] M.L. Balinski and H.P. Young. *Fair Representation:Meeting the Ideal of One Man, One Vote*. Yale University Press, 1982.

[8] M. Ball, C-Y. Chen, R. Hoffman, and T. Vossen. Collaborative decision making in air traffic management: Current and future research directions. In *Proceedings of ATM'99, Workshop on Advanced Technologies and Their Impact on Air Traffic Management in the 21st Century*. 1999.

[9] M. Ball, R. Hoffman, W. Hall, and A. Muharremoglu. Collaborative decision making in air traffic management: A preliminary assessment. Technical report, NEXTOR, The National Center of Excellence for Aviation Operations Research, 1998.

[10] M. Ball, R. Hoffman, D. Knorr, J. Wetherly, and M. Wambsganss. Assessing the benefits of collaborative decision making in air traffic management. In *Proceedings of 3rd USA/Europe air Traffic Management R&D Seminar*. 2000.

[11] M. Ball, R. Hoffman, A. Odoni, and R. Rifkin. Efficient solution of a stochastic ground holding problem. To be published in Operations Research, 2002.

[12] J. Bard, G. Yu, and M.F. Arguello. A time-band optimization model for the irregular operations problem. *IIE Transactions on Operations Engineering*, 33:931–947, 2001.

[13] J. Bautista, R. Companys, and A. Corominas. A note on the relation between the product rate variation (prv) problem and the apportionment problem. *Journal of the Operational Research Society*, 47:1410–1410, 1996.

[14] D. Bertsimas and S. Stock Patterson. The air traffic flow management problem with enroute capacities. *Operations Research*, 46:406–422, 1998.

[15] D. Bertsimas and S. Stock Patterson. The traffic flow management rerouting problem in air traffic control: A dynamic network approach. *Transportation Science*, 34:239–255, 2000.

[16] N. Bhogadi. Private Communication, 2002.

[17] L. Billera and D. Heath. Internal telephone billing rates - a novel application of non-atomic game theory. *Operations Research*, 26:956–965, 1978.

[18] M. Brennan. Simplified substitutions - enhancements to substitution rules and procedures during ground delay programs. Proceedings of AGIFORS Airline Operations Meeting, 2001.

[19] W. Brough, E. Clarke, and N. Tideman. Airport congestion and noise: Interplay of allocation and distribution. *Transportation Research Record*, 1450:3–7, 1995.

[20] J.-M. Cao and A. Kanafani. Real-time decision support for integration of airline flight cancellations and delays, part i: Mathematical formulation. *Transportation Planning and Technology*, 20:183–199, 1997.

[21] J.-M. Cao and A. Kanafani. Real-time decision support for integration of airline flight cancellations and delays, part ii: Algorithm and computational experiments. *Transportation Planning and Technology*, 20:201–217, 1997.

[22] P. Carlson. Exploiting the opportunities of collaborative decision making: A model and efficient algorithm for airline use. *Transportation Science*, 34:381–393, 2000.

[23] W. DeCota. Matching capacity and demand at LaGuardia airport. Presented at the NEXTOR Workshop on Airline and National Strategies for Dealing with Airport and Airspace Congestion, 2000.

[24] S. Grandeau, M. Clarke, and D. Mathaisel. The processes of airline system operations control. In G. Yu, editor, *Operations Research in the Airline Industry*, pages 312–369. Kluwer Academic Publishers, 1998.

[25] W.D. Hall. *Efficient Capacity Allocation in a Collaborative Air Transportation System*. PhD thesis, Massachusetts Institute of Technology, 1999.

[26] R. Hoffman. Refined benefits assessment of the collaborative arrival planner aatt decision support tool. Internal Document, Metron Inc., 2000.

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

[28] R. Hoffman, W. Hall, M. Ball, A. Odoni, and M. Wambsganns. Collaborative decision making in air traffic flow management. Working Paper, 2000.

[29] D. Hoitomt, D. Kraay, and B. Tang. Flight sequencing under the faa's simplified substitution rules. Presentation. AGIFORS Operations, Joint Study Group Meeting, Control Study and Ground Resources., April 1999.

[30] K. Howard. Slot credit substitutions issues. Internal CDM Memorandum, 2001.

[31] A. Hylland and R. Zeckhauser. The efficient allocation of individuals to positions. *Journal of Political Economy*, 87:293–314, 1979.

[32] Iata. Worldwide scheduling guidelines. International Air Transport Association, 3rd Edition, December 2000.

[33] J. Iginizo. *Linear Programming with Multiple Objectives*. John Wiley and Sons, New York, 1978.

[34] R. Inman and R. Bulfin. Sequencing jit mixed-model assembly lines. *Management Science*, 37:901–904, 1991.

[35] A. Jarrah, G. Yu, N. Krishnamurthy, and A. Rakshit. A decision support framework for airline flight cancellations and delays. *Transportation Science*, 27:266–280, 1993.

[36] V. Kolpin. Multi-product serial cost sharing: An incompatibility with the additivity axiom. *Journal of Economic Theory*, 69:227–233, 1995.

[37] H. Konishi, T. Quint, and J. Wako. On the shapley-scarf economy: the case of multiple types of indivisible goods. *Journal of Mathematical Economics*, 35:1–15, 2001.

[38] W. Kubiak. Minimizing variation of production rates in just-in-time systems: A survey. *European Journal of Operations Research*, 66:259–271, 1993.

[39] W. Kubiak and S. Sethi. A note on “level schedules for mixed-model assembly lines in just-in-time production systems”. *Management Science*, 37:121–122, 1991.

[40] S. Luo and G. Yu. Airline schedule perturbation problem: Ground delay program with splittable resources. In G. Yu, editor, *Operations Research in the Airline Industry*, pages 433–460. Kluwer Academic Publishers, 1998.

[41] S. Luo and G. Yu. Airline schedule perturbation problem: Landing and takeoff with non-splittable resource for the ground delay program. In G. Yu, editor, *Operations Research in the Airline Industry*, pages 404–432. Kluwer Academic Publishers, 1998.

[42] H. Luss. On equitable resource allocation problems: A lexicographic minimax approach. *Operations Research*, 47:361–378, 1999.

[43] J. Milner. *Dynamic Slot Allocation with Airline Participation*. PhD thesis, Massachusetts Institute of Technology, 1995.

[44] J. Miltenburg. Level schedules for mixed-model assembly lines in just-in-time production systems. *Management Science*, 35:192–207, 1989.

[45] H. Moulin. On additive methods to share joint costs. *The Japanese Economic Review*, 46:303–332, 1995.

[46] H. Moulin. Incremental cost sharing: Characterization by coalition strategy-proofness. *Social Choice and Welfare*, 16:279–320, 1999.

[47] H. Moulin. Priority rules and other asymmetric rationing methods. *Econometrica*, 68:643–684, 2000.

[48] H. Moulin. The proportional random assignment of indivisble units. Mimeo, Rice University, 2000.

[49] H. Moulin. Axiomatic cost and surplus-sharing. Mimeo, Rice University, 2001.

[50] H. Moulin and S. Shenker. Serial cost sharing. *Econometrica*, 60:1009–1038, 1992.

[51] H. Moulin and R. Stong. Fair queueing and other probabilistic allocation methods. Mimeo, Rice University, 2001.

[52] Hervé Moulin. *Cooperative Microeconomics: A Game-Theoretic Introduction*. Princeton University Press, 1995.

[53] G.L. Nemhauser and L.A. Wolsey. *Integer and Combinatorial Optimization*. John Wiley and Sons, New York, 1988.

[54] T. Niznik. Optimizing the airline response to ground delay programs. Proceedings of AGIFORS Airline Operations Meeting, 2001.

[55] Michael S. Nolan. *Fundamentals of Air Traffic Control*. Brooks/Cole Publishing Company, 3 edition, 1999.

[56] A. Odoni. The flow management problem in air traffic control. In A.R. Odoni, L. Bianco, and G. Szego, editors, *Flow Control of Congested Networks*, pages 269–288. Springer-Verlag, Berlin, 1987.

[57] A. Odoni and T.P. Fan. The potential of demand management as a short-term means of relieving airport congestion. In *ATM-2001 EUROCONTROL-FAA Seminar, Santa Fe, NM*. 2001.

[58] M. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

[59] Guillermo Owen. *Game Theory*. Academic Press, 3rd edition, 1995.

[60] A. Rakshit, N. Krishnamurthy, and G. Yu. System operations advisor: A real-time decision support system for managing airline operations at united airlines. *Interfaces*, 26:50–58, 1996.

[61] S. Rassenti, V. Smith, and R. Buffin. A combinatorial auction mechanism for airport time slot allocation. *Bell Journal of Economics*, 13:402–417, 1982.

[62] J. Rawls. *A Theory of Justice*. Harvard University Press, 1971.

[63] O. Richetta. Optimal algorithms and a remarkably efficient heuristic for the ground-holding problem in air traffic control. *Operations Research*, 43:758–770, 1995.

[64] O. Richetta and A. Odoni. Solving optimally the static ground holding policy problem in air traffic control. *Transportation Science*, 24:228–238, 1993.

[65] W.H. Riker and I. Sened. A political theory on the origin of property rights: Airport slots. *American Journal of Political Science*, 35:951–969, 1991.

[66] A. Sen. *Collective Choice and Social Welfare*. Holden-Day, 1970.

[67] L. Shapley and H. Scarf. On cores and indivisibility. *Journal of Mathematical Economics*, 1:23–37, 1974.

[68] M. Shubik. Incentives, decentralized control, and the assignment of joint costs and internal pricing. *Management Science*, 8:325–343, 1962.

[69] G. Steiner and S. Yeomans. Level schedules for mixed-model just-in-time processes. *Management Science*, 39:728–735, 1993.

[70] D. Teodorovic and S. Guberinic. Optimal dispatching strategy on an airline network afer a schedule perturbation. *European Journal of Operations Research*, 15:178–182, 1984.

[71] D. Teodorovic and S. Guberinic. Model for operational daily airline scheduling. *Transportation Planning and Technology*, 14:273–285, 1990.

[72] M. Terrab. *Ground Holding Strategies for Air Traffic Control*. PhD thesis, Massachusetts Institute of Technology, Department of Civil Engineering, 1990.

[73] M. Terrab and A. Odoni. Strategic flow management for air traffic control. *Operations Research*, 41:138–152, 1993.

[74] B. Thengvall, J. Bard, and G. Yu. Multiple fleet aircraft schedule recovery following hub closures. *Transportation Research, Part A*, 35:289–308, 2001.

[75] B. Thengvall, J.F. Bard, and G. Yu. Balancing user preferences for aircraft recovery during airline irregular operations. *IIE Transactions on Operations Engineering*, 32:181–193, 2000.

[76] W. Thomson. Monotonicity, stability, and egalitarianism. *Mathematical Social Sciences*, 8:15–28, 1984.

[77] W. Thomson. The consistency principle. In T. Ichiishi, A. Neyman, and Y. Tauman, editors, *Game Theory and Applications*, pages 187–215. Academic Press, New York, 1990.

[78] U.S. House of Representatives. H.R. 1407. http://www.house.gov/transportation, 2001.

[79] A. Vasquez-Marquez. American airlines arrival slot allocation system. *Interfaces*, 21:42–61, 1991.

[80] W. Vickrey. Counterspeculation, auctions, and competitive seales tenders. *Journal of Finance*, 16:8–37, 1961.

[81] T. Vossen and M. Ball. Optimization and mediated bartering models for ground delay programs. Working Paper, 2000.

[82] P. Vranas, D. Bertsimas, and A. Odoni. Dynamic ground-holding policies for a network of airports. *Transportation Science*, 28:275–291, 1994.

[83] M. Wambsganns. Collaborative decision making through dynamic information transfer. *Air Traffic Control Quarterly*, 4:107–123, 1996.

[84] CDM working group. Cdm rules of engagement. Internal Memorandum to CDM working group, 1996.

[85] H.P. Young. Distributive justice in taxation. *Journal of Economic Theory*, 44:321–335, 1988.

[86] H.P. Young. *Equity in Theory and Practice*. Princeton University Press, 1994.