Slot Trading Opportunities in Collaborative Ground Delay Programs

Thomas W.M. Vossen
Leeds School of Business,
University of Colorado,
Boulder, CO 80309-0419

Michael O. Ball
Robert H. Smith School of Business
and Institute for Systems Research,
University of Maryland,
College Park, MD 20742-1815

May 13, 2005

Abstract

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.

The CDM paradigm has led to fundamental changes in the implementation of Ground Delay Programs. A key component has been the introduction of the Compression procedure, which allows for the exchange of arrival slots between airlines. In this paper, we consider opportunities for increased airline control by interpreting the Compression procedure as a mediated slot trading mechanism. Based on this interpretation, we propose an extension that allows airlines to submit so-called "at-least, at-most" offers. We develop an efficient integer programming model to solve the mediator's problem, and show that the resulting mechanism can substantially improve the ability of airlines to optimize their internal cost functions.

1 Introduction

The Federal Aviation Administration (FAA) issues Ground Delay Programs (GDPs) to manage temporary reductions in an airport's arrival capacity. In a GDP, flights bound for congested airports are delayed on the ground (prior to their departure), so as to balance the arrivals with the reduced capacity at the airport under consideration. 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. The allocation of ground delays can be interpreted as a resource allocation problem in which available arrival capacity (i.e., a sequence of arrival slots) has to be distributed among the inbound flights. Given the allocation of arrival slots to flights, it is straightforward to determine the resulting ground delays. It is, however, important to note that the resulting allocation problem is complicated by a number of factors that are peculiar to the context in which GDPs are implemented. First, GDPs are intrinsically dynamic in nature, due to the inherent uncertainty related to weather predictions. GDPs, for instance, are implemented anywhere from 6 to 2 hours before the flights are actually scheduled to land, leading to a very short time frame for making decisions. In addition, the implementation of a GDP can have a severe negative impact on an airline's schedule integrity, given that capacity reductions of 50% for extended periods of time are not uncommon. Finally, it is worthwhile to note that GDPs in fact occur frequently. Between September of 1998 and April of 1999, for instance, there was an average of 1.6 GDPs per day (Hall, 1999).

GDPs, as well as other traffic flow management initiatives, have been used by the FAA for nearly 20 years now (see Nolan, 99, for their origins). The advent of the Collaborative Decision Making (CDM) paradigm, however, has had a profound effect on the implementation of traffic flow management initiatives. CDM is based on the recognition that improved data exchange and communication between the FAA and the airlines will lead to better decision making (Wambsganss, 96). 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. While the CDM paradigm applies to a wide range of applications in Air Traffic Flow Management (ATFM), the primary focus so far has been the implementation of Ground Delay Program (GDP) enhancements.

The number of enhancements that have recently been implemented under CDM are numerous: examples include improved data-exchange, better situational awareness tools, and increased flexibility for airlines (see Ball et al., 1998, 2000). However, the most significant improvements have been introduced by fundamental changes in the allocation of arrival capacity. This approach is based on the consensus recognition that airlines have claims on the available arrival capacity, based on their original flight schedules. This realization has led to a fundamental change in the way arrival capacity is allocated, as well as the introduction of a procedure for inter-airline slot exchange. Under CDM, arrival capacity is allocated to airlines by a procedure called *Ration-By-Schedule* (RBS). This procedure has removed disincentives airlines previously had to provide accurate information about delays and cancellations. In addition, CDM has led to the introduction of a new procedure for inter-airline slot exchange, called *Compression*. This procedure seeks to maximize utilization of the available arrival capacity in the presence of delays and cancellations, and attempts to do so in a fair and equitable manner.

This paper is motivated by the inter-airline exchange of slots that currently occurs in the Compression algorithm. The introduction of the Compression algorithm has had a major impact, and is said to have enabled significant reductions in ground delay (Chang et al., 2001). Our main objective is to explore the potential benefits of enhancing the slot exchange capability. First, we interpret the exchange of slots as a form of bartering, in which airlines may submit slot trading offers to the FAA, which acts as the central coordinator (see also Vossen, 2002). Subsequently, we propose an optimization model for the FAAs mediation problem, which must determine the set of offers to accept. This approach generalizes current slot exchange procedures in that it allows airlines to submit so-called at-least, at-most offers; such offers may be viewed as trade-offs between pairs of flights and are motivated by differences in their marginal delay costs. To analyze the potential benefits of this approach, we consider two case studies that use different models of airline decision-making: in the first model, airlines aim to maximize their on-time performance, while in the second model airlines aim to minimize passenger delay costs. We consider experimental results that are based on historical GDP data to show that, on the whole, slot trading may yield significant benefits over the current slot exchange mechanisms.

This paper is organized as follows. Section 2 provides background on GDPs, in particular on the procedures used under CDM, and on the role of inter-airline slot exchange. Section 3 discusses our general approach to increasing slot trading opportunities and describes the optimization models we propose to select a set trades to implement. Section 4 presents the results of our extensive computational experiments. Finally, Section 5 provides conclusions and avenues for further research.

2 Background

The use of ground holding to resolve air traffic congestion was first described systematically by Odoni (Odoni, 1987). The generic flow management problem defined by Odoni is extremely general in that it addresses congestion anywhere in the network. However, he provides a robust rationale for the practical relevance of models where the only capacitated element considered is the airport arrival mechanism. Under this assumption, the problem is commonly known as the Ground Holding Problem (GHP). The basic version of the GHP (see Terrab, 1990) assumes a discrete time horizon, deterministic demand, and deterministic capacity. Given these assumptions, the GHP can be formulated as an assignment problem. Most models that address the GHP, however, concentrate on the trade-off between airborne and ground delays in the presence of stochastic capacity. This version of the GHP was first studied by Odoni (Odoni, 1987) and Andreatta and Romanin-Jacur (Andreatta and Romanin-Jacur, 1987). More efficient models, as well as several extensions have also been proposed (Terrab and Odoni, 1993, Richetta and Odoni, 1993, Richetta, 1995, and Ball et al., 2003). A systematic review of some of these results may be found in (Andreatta et al., 1993). 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 (Vranas et al., 1994, and Andreatta and Brunetta, 1993) and on more general air traffic flow management problems (Bertsimas and Stock-Patterson, 1998, 2000). 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. That is, most ground-holds would be assigned to aircraft with smaller per-unit delay costs (e.g. regional aircraft), while aircraft with higher delay costs (e.g. wide-body aircraft) would be given priority (Odoni, 1987). Consequently, the models described here are perhaps primarily suited for making aggregate decisions (e.g., determining overall flow rates per period (Ball et al., 2000)).

In contrast to the models proposed in the literature, which allocate delays to individual flights, the allocation procedures instituted under CDM primarily address the distribution of delays among airlines. CDM has its origin in early efforts by the FAA to acquire up-to-date airline schedule information (Wambsganss, 1996). Though human-in-the-loop experiments with air traffic flow management (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 (see Vossen, 2002). 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, Ration-By-Schedule (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

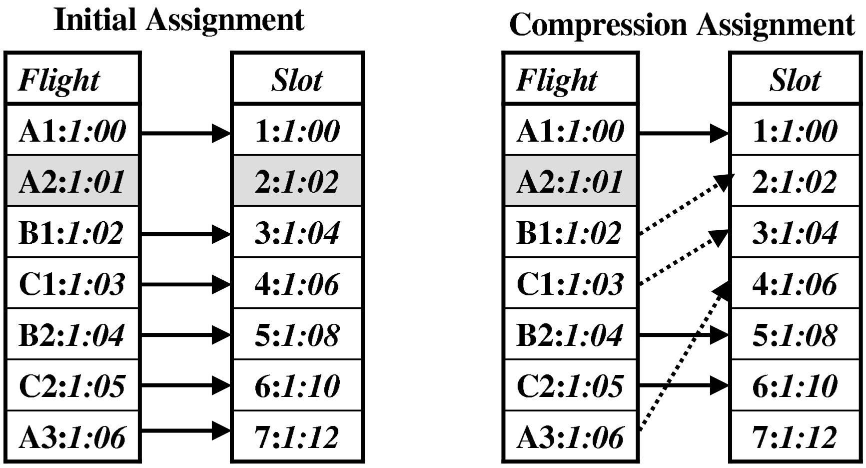

Figure 1: Compression Example.

flight schedules. Under RBS, flights are assigned to slots 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 happened prior to CDM. 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 this initial allocation, airlines are free to reschedule flights according to their private objectives, through flight substitutions and cancellations.

After a round of substitutions and cancellations the utilization of slots can usually be improved. The reason for this is that the resultant schedule, which results after the airlines perform a round of flight substitutions and cancellations, may have "holes" due to the high degree of disruption to airline schedules that can occur on bad weather days. In other words, 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. To illustrate the Compression algorithm, let us consider the example shown in Figure 1. The leftmost figure represents the flight-slot

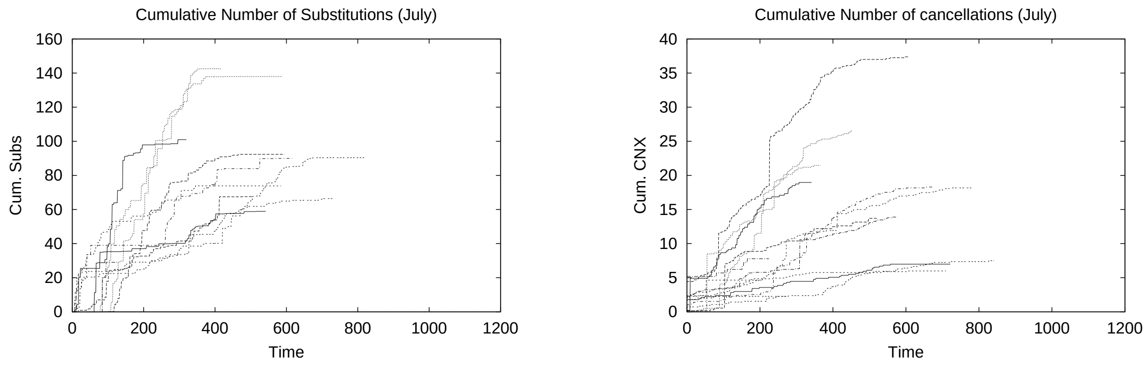

Figure 2: Airline GDP behavior at O'Hare Airport, July 2000

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 cancelled flight, $A2$. The rightmost figure shows the flight schedule after execution of the Compression algorithm: as a first step, the algorithm attempts to fill slot 2 owned by airline A. Since there is no flight from airline A that can use the slot, the slot is allocated to flight $B1$ and airline A takes ownership of slot 3. This process is repeated for slot 3, which results in the assignment of flight $C1$ to slot 3 and transfer of ownership of slot 4 to A. Now, slot 4 can be used by airline A and will be assigned to flight $A3$. 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 (see Vossen, 2002, for a more detailed description).

From an airline standpoint, the ability to substitute and cancel flights is clearly the single most important aspect of a GDP. The increased control 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 2, 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 2 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 as a result of the RBS procedure, 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 2 represents the cumulative percentage of flights that have been cancelled during the course of a GDP. The graphs in Figure 2 show first that airlines perform a large number of flight substitutions and cancellations, and second that airlines perform flight transactions 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 (e.g. 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 envision more general exchange mechanisms. In fact, a more dynamic form of slot exchange functionality, known as *Slot Credit Substitutions* (SCS), has been defined by the CDM working group (Howard, 2001) and is now operating. Under SCS, airlines submit "conditional" cancellation 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 monitors such requests on a continuous basis, and, if possible, immediately implements the associated exchange(s) of slots. SCS can be viewed as a "real-time" version of the Compression procedure, which is a "batch" process run periodically. The conditional nature of SCS requests as well as the real-time response provides increased trading opportunities over compression.

3 Slot Trading

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 airlines would bargain amongst themselves (Adams et al., 1997). It is, however, difficult to envision the use of such highly decentralized mechanisms as near- or medium-term solutions 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. In this paper, we therefore consider more modest generalizations of the Compression and SCS framework. Under this framework, airlines may submit offers to exchange slots. The FAA, on the other hand, would act as a mediator who evaluates and selects possible trades. We note that, not only does this approach allow for more complex trade offers, but, by solving a formal mediation problem, it allows for explicit consideration of equity in determining the overall set of offers accepted.

3.1 Model Concepts

The general idea behind this approach is based on interpreting the Compression procedure as a form of mediated bartering. This interpretation relies on the observation 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. Each flight cancellation may be therefore be viewed as an offer to trade the resulting open slot for a later slot that will reduce the delay on one of the releasing airline's subsequent flights. This type of offer is depicted in Figure 3. In addition, we assume that airlines are 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. This offer type is depicted in Figure 4.

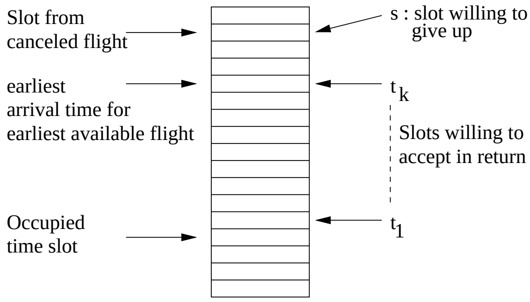

Figure 3: Offer associated with cancelled or delayed flights

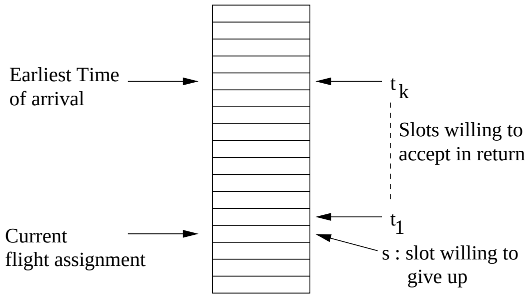

Figure 4: "Default" Offers

Given a resulting set of offers, the FAA (in its role as mediator) will have 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. Several criteria could be used to select the actual trades that are executed: one possibility, for instance, is to use a bilevel programming approach in which offers to move down are given priority. This approach yields solutions that are similar to the Compression Algorithm.

Under the interpretation of Compression as slot trading, only one-for-one trades are allowed. More generally, one might consider k-for-n trades. To illustrate the potential benefits, consider the two cases illustrated in Figure 5. On the left of this Figure, we see an example of a slot state that would motivate one-for-one trading (Compression). Such trading

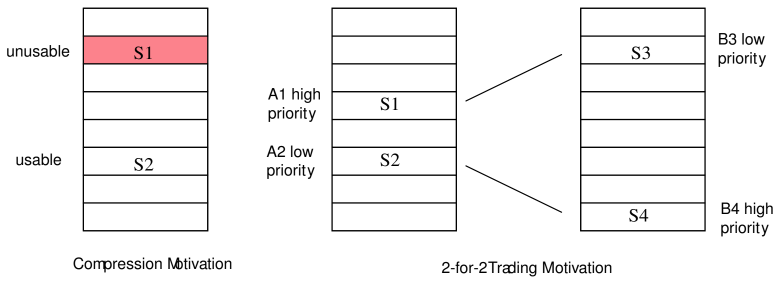

Figure 5: Value Proposition for one-for-one and two-for-two trades

always must start with a case where an airline is willing to give up an earlier slot for a later one. The economic motivation for such a move is that an airline cannot use the earlier slot, i.e. to that airline, it has a zero value. On the other hand, as illustrated in the right side of the Figure, airlines A and B might both have a pair of usable slots, which are occupied by flights with differing priority (delay cost) levels. It could easily be that the value to A of slots $s1, s2, s3$ and $s4$ are \$1500, \$500, \$2000 and \$300 and the value to B of the same four slots are: \$500, \$2500, \$800 and \$1800. Thus, the value proposition to A of trading $s1$ and $s2$ for $s3$ and $s4$ is: 2000 - 1500 + 300 - 500 = \$300 and the value proposition to B of the same trade is: 500 - 800 + 2500 - 1800 = \$400. Thus, both airlines place positive values on all of the slots but because of the manner in which the slots can be mapped to the airlines' flights both airlines benefit from the trade. Such trade opportunities should be much more common than the ones motivated by the Compression algorithm which arise mainly as a result of cancelled flights.

The remainder of this section describes trading mechanisms and models to support trading of the type described above. Section 3.2 introduces a general approach and mediator model that supports general k-for-n trades. Subsequently, in Section 3.3, we discuss a specific model of airline decision-making within GDPs that leads to a restricted but more practical approach. Finally, in Section 3.4, we propose a mediator optimization model that incorporates these restrictions. The experimental results presented in Section 4 show that the trading model we propose is capable of providing significant performance improvements over those achieved by the Compression procedure, indicating that it represents a practical new approach to slot trading that is relevant in the current system.

3.2 Model Overview

The general slot trading framework we envision consists of the follow basic "iteration" between the airlines and the mediator:

- Periodically (say every 15 or 30 minutes), airlines submit a list of trade offers they would desire.
- Subsequently, the mediator (FAA) chooses a set of offers to accept. There are potentially many alternate approaches to formulating and solving this mediator decision problem.

The model we are proposing can be viewed as a "direct" extension of the one-for-one trading model described in the previous section in the sense that each offer will be considered independently of the others. Specifically, the mediator has the ability to accept or reject any individual offer or combination of offers. Thus, the approach does not support offers that might be contingent on the acceptance of other offers or other complex transactions. While it might be tempting to include such complexities, we wish to maintain a certain level of simplicity in order to achieve a practical system. The notation introduced below allows us to specify the precise nature of the trade offers and facilitates the development of a rational framework for the type of offers proposed by the airlines as well as the mediators decision on which offers to accept.

- $F = \{f_1, \ldots, f_n\}$, the flights affected by the GDP; for each flight $f_i \in F$, we let $erta_i$ earliest its earliest runway time of arrival;
- $S = \{s_1, \ldots, s_n\}$, the sequence of arrival slots in the GDP (We note that there are several conditions that can lead to differences in the number flights and slots but for notational convenience we assume they are the same (n)). For each slot $s_j \in S$, we let t_j represents the slot time;

- A, the airlines involved in the GDP;
- $O : F \to A$, a mapping that defines the flight-airline relation. For each airline $a \in A$, F_a represents the flights from that airline, i.e., $F_a = \{f \in F \mid O(f) = a\}$. At the start of a period of trading, all flights have been assigned a slot; we assume that flight f_i is assigned to slot s_i for all $f_i \in F$. This assignment specifies each airline's allotment of slots, that is, $S_a = \{s_i \in S : f_i \in F_a\}$ represents the set of slots owned by airline a.

Given these initial allotments, we can associate with each airline a a set of offers $T_a \subseteq 2^{S_a} \times 2^{S-S_a}$. That is, each offer $\tau = (U_\tau; V_\tau) \in T_a$ specifies that airline a would be willing to offer the slots in U_τ in return for the slots in V_τ. We assume that U_τ and V_τ have the same cardinality. In addition to the offers proactively provided by the airlines, we also assume the availability of default offers, D, which specify that an airline would always be willing to reduce the delay of any of its flights,

$$D = \{(s_i, s_j) : 1 \leq i, j \leq n, erta_i \leq t_j \leq t_i\},$$

i.e. an airline is willing to trade the current slot s_i assigned to a flight for an earlier slot s_j up to the flight's earliest feasible arrival slot. Note that the airline-provided information does not include any information about its relative value for these trades. We assume implicitly that airlines only create offers that "make sense" for their flights, i.e. if any offer is executed then, afterwards, there will be a feasible assignment of an airline's flights to the slots it owns.

Given these offers, the mediator's task is to choose a compatible set of offers to accept. Of course, one could evaluate a set of offers along several dimensions in trying to choose the best set to accept. Here, we use the relatively simple criterion of accepting the maximum number of compatible offers. This problem can be formulated as a set-partitioning problem, which has a variable associated with both the offers submitted by the airlines and the default offers, that is,

- $y_\tau \in \{0, 1\}$ for all $a \in A$, $\tau \in T_a$. $y_\tau = 1$ if and only if offer τ is selected.

- $x_{ij} \in \{0,1\}$ for all $(s_i, s_j) \in D$. $x_{ij} = 1$ 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 objective function equals

$$Max \sum_{a \in A, \tau \in T_a} y_\tau, \tag{1a}$$

although, as stated before, other objective functions could also be used. The constraints can be represented as

$$\sum_{s_j:(s_i,s_j)\in D} x_{ij} + \sum_{a\in A,\tau\in T_a: s_i \in U_\tau} y_\tau = 1 \quad \text{for all } s_i \in S \tag{1b}$$

$$\sum_{s_i:(s_i,s_j)\in D} x_{ij} + \sum_{a\in A,\tau\in T_a: s_j \in V_\tau} y_\tau = 1 \quad \text{for all } s_j \in S \tag{1c}$$

The right hand side of this formulation represents the set of slots twice and the constraints insure that each slot is assigned to both an offer that gives it up and an offer that receives it. Constraint (1b) states that each slot is assigned to some offer (default or airline provided) that proposes to give up that slot. Constraint (1c) states that each slot is assigned to some offer that requires that slot in return. The situation where slot s_i is not traded corresponds to selecting the default offer (s_i, s_i).

3.3 Airline Decision-Making and Offer Restrictions

While the set-partitioning formulation provides a general framework for slot trading, the arbitrary complexity of the offers that can be submitted reduces its practical relevance. To enable a more realistic approach we first consider a basic model of airline decision-making within GDPs, in which airline preferences are represented by an assignment model (see Brennan, 2001, Niznik, 2001, and Yu and Luo, 1999).

In formulation (2) below, x_{ij} and y_i are binary variables. Variable x_{ij} takes value 1 if and only if flight f_i as assigned to slot s_j; y_i is a slack variable that takes value 1 if and only if f_i is cancelled. Expression (2a) defines $u_a(S')$, the optimal value to airline $a \in A$ of any subset of slots $S' \subseteq S$. The objective function coefficients, w_{ij} and c_i, represent the value of assigning flight f_i to slot j and the cost of cancelling flight f_i respectively.

$$u_a(S') = Max \sum_{f_i \in F_a, s_j \in S} w_{ij}x_{ij} - \sum_{f_i \in F_a} c_i y_i \tag{2a}$$

subject to the constraints

$$\sum_{s_j \in S' : t_j \geq erta_i} x_{ij} + y_i = 1 \qquad \text{for all } f_i \in F_a \tag{2b}$$

$$\sum_{f_i \in F_a : t_j \geq erta_i} x_{ij} \leq 1 \qquad \text{for all } s_j \in S' \tag{2c}$$

$$x_{ij}, y_i \geq 0, \tag{2d}$$

where w_{ij} represents the value of assigning flight f_i to slot s_j and c_i represents the cost of cancelling flight f_i. Note the following two important special cases of this model:

Intra-airline optimization – $S' = S_a$**:** Here, formulation (2) can be interpreted as the intra-airline optimization airlines perform in the substitution/cancellation process, i.e. it seeks to find the best assignment of that airline's flights to the slots it "owns".

System-wide optimization – let a **represent all airlines in** A **and** $S_a = S$**:** The formulation can be viewed as a global inter-airline optimization model. The model seeks a solution that achieves system-wide efficiency. Although employing such a model and "forcing" implementation of such a solution is incompatible with the CDM paradigm, the resulting value will provide an upper bound on the benefits that can be obtained by slot trading.

In trying to understand the possible motivation for slot trading, it is instructive to consider the structure of a flight's delay costs. Delay costs are oftentimes reasonably approximated by a staircase structure as shown in Figure 6 (Brennan, 2001). 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

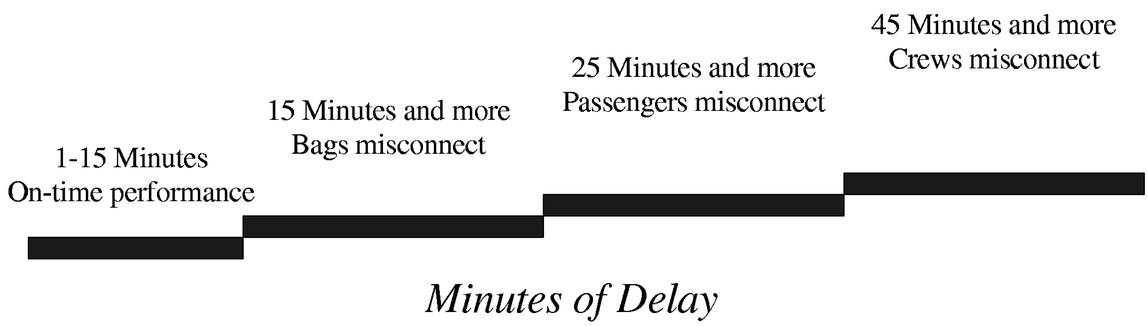

Figure 6: Delay Cost Structure

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.

The intra-airline optimization is actually implemented via the CDM substitution process in which airlines interchange the slot assignment of pairs of flights. For instance, an airline could decide to delay a flight with few passengers while reducing the delay on a more heavily loaded flight that would allow its passengers to make their connections. The current use of pairwise exchanges as well as their potential benefits suggest a possible restriction to the general slot trading framework, 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). 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 (i.e. both flights are moved down), or (3) the offer expresses a trade for one earlier slot and one later slot (i.e. 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. It is worthwhile to note the resulting generalization of SCS requests: whereas as an SCS request states "I am willing to cancel flight f_i in return for a reduction in the delay of flight f_j", a two-for-two trade offer states "I am willing to delay (but not necessarily cancel) flight f_i in return for a reduction in the delay of flight f_j".

While these restrictions in theory limit the potential exchanges during the course of a GDP, our experiments show that two-for-two trades yield substantial benefits over current procedures. Moreover, the restriction to two-for-two trades significantly reduces the 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. This observation is crucial as experience in other CDM initiatives has shown that it is essential to simplify airline information requirements as well as operational processes in order to achieve acceptance and eventual implementation.

3.4 Model Formulation under Offer Restrictions

Even though the set-partitioning formulation of the mediation problem can be used to find compatible trades in the case of "at-least, at-most" offers, the large number of variables makes this approach intractable for all but the smallest cases. When only two-for- two trades are allowed, the resulting set of offers can be defined more succinctly, motivating the alternate formulation of the mediation problem presented in this section. The formulation below models the mediation problem as a network flow problem with side constraints. Defining T as the set of all possible trades, the two-for-two offers to trade are characterized by a tuple $(f_d, s_{d'}; f_u, s_{u'}) \in T$, which states that the airline is willing to move down flight f_d to a slot no later than $s_{d'}$ in return for moving up flight f_u to a slot that is no later than $s_{u'}$. In the remainder of this section we 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. Flow proceeds from a flight source node to a

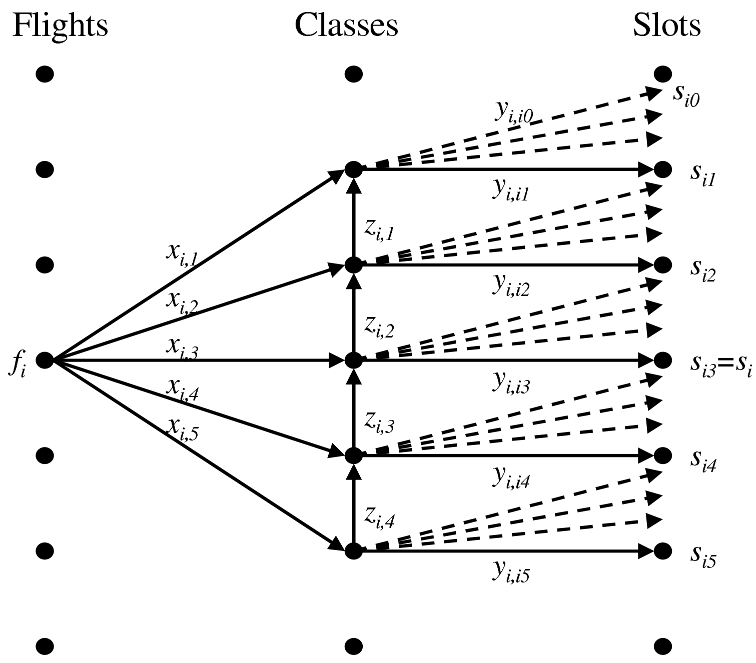

Figure 7: Structure of Network Flow Constraints

class node and finally to a slot node. The slot-defined classes represent either the maximum amount of delay or the minimum delay reduction requested for that flight if the offer is executed. For instance, if the offer $(f_d, s_{d'}; f_u, s_{u'})$ is executed, flight f_d will be assigned to class $s_{d'}$ and f_u will be assigned to class $s_{u'}$. This assignment will ensure that f_u will receive a slot no later than $s_{u'}$ and that f_d will receive a slot no later than $s_{d'}$. The side constraints are needed to ensure that only assignments corresponding to proposed offers are selected. To illustrate this idea, we start by considering a single flight, f_i, and examine all the offers in which it occurs. These offers determine a sequence of classes, which we can represent by the indices of the corresponding slots. That is, the offers specify a sequence $i_1, \ldots, i_{k_i}$, where $t_{i_k} < t_{i_{k'}}$ if $k < k'$. Thus, if $i_k < i$ there exists a trade offer which requests that flight f_i is moved up (earlier in time) to at least slot s_{i_k} (from its current assignment to slot s_i). If, on the other hand, $i_k > i$ there exists a trade offer to move f_i down (later in time) to a slot no later than s_{i_k}. We assume that there is one $k : 1 \leq k \leq k_i$ such that $i_k = i$, corresponding to the default offer. Moreover, we use i_0 to represent the index of the earliest slot that flight f_i can be assigned to, i.e. $i_0 = \min_{j:t_j \geq erta_i} j$. Consequently, if flight f_i is assigned to class i_k it will receive a slot in the range $s_{i_0}, \ldots, s_{i_k}$.

To represent the IP formulation, we partition the set of offers to trade as follows: $D_T = \bigcup_{(f_d, s_{d'}; f_u, s_{u'}) \in T} (f_d, s_{d'})$, which contains the classes that correspond to downward moves for each flight $f_d \in F$; $U_T = \bigcup_{(f_s, s_{d'}; f_u, s_{u'}) \in T} \{(f_u, s_{u'})\}$, which contains the classes that correspond to upward moves for each flight $f_d \in F$; and, $N_T = \bigcup_{f_i \in F} \{(f_i, s_i)\}$, which contains the classes corresponding to default offers for each flight $f_i \in F$. The resulting IP formulation uses the following variables.

- $x_{ik} \in \{0, 1\}$ for $f_i \in F, 1 \leq k \leq k_i$; $x_{ik} = 1$ iff f_i is assigned to class i_k.
- $y_{ij} \in \{0, 1\}$ for $f_i \in F, s_j \in S$; $y_{ij} = 1$ iff flight f_i is assigned to slot s_j.
- $z_{ik} \in \{0, 1\}$ for $f_i \in F, 1 \leq k \leq k_i$; $z_{ik} = 1$ iff f_i has been assigned to a class with index lower than k in the sequence of classes for f_i but receives at least slot s_{i_k}. The use of this variable is explained in more detail below.
- $\hat{x}_{dd'uu'} \in \{0, 1\}$ for $(f_d, s_{d'}; f_u, s_{u'}) \in T$. $\hat{x}_{dd'uu'} = 1$ iff offer $(f_d, s_{d'}; f_u, s_{u'})$ is executed.

The following constraints represent the network flow component of the IP formulation, in which the flights are assigned to classes and classes are assigned to slots.

$$\sum_{k=1}^{k_i} x_{ik} = 1 \quad \text{for all } f_i \in F \tag{3a}$$

$$x_{i1} + z_{i1} - \sum_{j=i_0}^{i_1} y_{ij} = 0 \quad \text{for all } f_i \in F \tag{3b}$$

$$x_{ik} + z_{ik} - z_{i(k-1)} - \sum_{j=i_{k-1}+1}^{i_k} y_{ij} = 0 \quad \text{for all } f_i \in F, 1 < k \leq k_i \tag{3c}$$

$$\sum_{f_i \in F: i_0 \leq j \leq i_k} y_{ij} = 1 \quad \text{for all } s_j \in S \tag{3d}$$

Constraint (3a) represents the assignment of flights to classes, while constraints (3b) and (3c) represent the subsequent assignment of classes to slots. Finally, constraint (3d) represents the restriction that each slot is assigned exactly once. The resulting network flow structure is shown in Figure 7, for a single flight f_i with $k_i = 5$. The leftmost nodes in Figure 7 correspond to flights, which are first assigned to classes. After a flight has been assigned to class i_k, it can either be assigned to a slot with an index in the range $i_{k-1}+1, \ldots, i_k$ or it can be assigned to higher class. The latter is achieved with the z_{ik} variables, which ensure that when flight f_i is assigned to class i_k, it can can also receive a slot in the range $i_0, \ldots, i_{k-1}$.

In addition to the network flow constraints, the IP formulation also contains the following side constraints

$$x_{dd'} = \sum_{(f_d, s_{d'}, f_u, s_{u'}) \in T} \hat{x}_{dd'uu'} \quad \text{for all } (f_d, s_{d'}) \in D_T \tag{3e}$$

$$x_{uu'} = \sum_{(f_d, s_{d'}, f_u, s_{u'}) \in T} \hat{x}_{dd'uu'} \quad \text{for all } (f_u, s_{u'}) \in U_T \tag{3f}$$

Constraints (3e) and (3f) ensure that the resulting trades only include offers proposed by the airlines. In words, the constraint states that a flight is moved down if and only if another flight is moved up, in accordance with one of the proposed trades.

The constraints in formulation (3) define the set of feasible trades. An optimal set of trades is achieved by selecting an appropriate objective function. The objective function may be as simple as maximizing the number of trades that are executed. However, one may also consider objective functions that incorporate system-wide performance measures such as maximizing on-time performance or minimizing passenger delays, which we examine in the experiment described in Section 4. It is important to note that, even though it is beyond the scope of our work in this paper, our model is also capable of incorporating equity considerations via the objective function and/or additional constraints.

4 Experimental Results

In this section, we analyze the slot trading approach presented in sections 3.3 and 3.4 using two case studies. The objective of these case studies is threefold: (1) to analyze potential benefits of increased coordination, (2) to consider the impact of airline behavior in the trading process, and (3) to analyze the efficiency of the underlying optimization models. The IP formulations in our experiments were solved using Cplex 7.1 on a Sun Ultra 10 workstation.

In the first case study the airlines' objective is to maximize on-time performance; in the second, the airlines objective is to minimize passenger delays. These objectives represent two of the most important factors that airlines consider in their decision-making process. The importance of on-time performance is discussed in Luo and Yu (1998). The minimization of passenger delays, an equally important objective, has been used by American Airlines in their GDP decision support tools (Vasques-Marquez, 1991, Niznik, 2001).

For each case study, we assess the potential improvement to the airlines' objective under consideration that could be obtained by trading slots among airlines. Our analysis is based on historical data from a set of GDPs at Boston's Logan airport between January and April of 2001. For each of these GDPs, we collected the program data after RBS was first executed. The resulting data provided us with the sets of flights, airlines, and slots (based on the actual airport acceptance rates), parameter values such as the earliest runway arrival times and slot times, as well as the initial allocation of flights to slots. Given these parameters, we fixed the assignment of all the flights that were exempted from the program, and the resulting flight-slot assignment of the remaining (non-exempt) flights was used to determine each airline's initial allotment of slots, S_a.

In each case study, we analyze the results of solving the following optimization problems:

i) We solved formulation (2) for each airline using the objective function under consideration in the case study. The resulting slot allocation represents an optimal allocation without coordination. This is the equivalent of the substitution process in the current system before compression takes place.

ii) We also solved formulation (2) without considering slot ownership. In this case, the resulting slot allocation represents an optimal allocation with coordination, i.e., the solution that could be obtained if a centralized optimization approach were possible, which of course is incompatible with the decentralized planning approach introduced by CDM. However, obtaining this solution is important as it provides an upper bound to the benefits that could be achieved by inter-airline trading of slots in the current system.

iii) In order to analyze the potential benefits of the two-for-two slot trading mechanism we started with the flight-to-slot allocation that maximized internal airline performance, i.e. the allocation described in i) above using formulation (3). Then, for each airline, we generated all 2-for-2 trades that improved the airline objective of interest and solved optimization model (3) to find a mutually compatible set of trades to implement.

iv) For the on-time performance case, we also assessed the benefit provided by the Compression Algorithm in place in the current system. We made the assumption that airlines would cancel flights with excessive delays, i.e., delays of two hours or more, prior to the execution of Compression. We did not perform this step in the passenger delay cost case study since there was no natural way to represent airline cancellation behavior or to assign a cost penalty to passengers on cancelled flights. We should also note that the analysis of this case for the on-time-performance case study demonstrated significantly better performance for 2-for-2 trading.

We conclude each study with a high level assessment on running times that suggests the trading models we propose are suitable for decision-making in real time.

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 arrival time as scheduled in the official airline guide, oag_i (we note that $erta_i$, defined earlier, and oag_i can differ; in general, $erta_i \geq oag_i$ but $erta_i$ can be strictly greater than oag_i, most typically due to "upstream" delays that induce unavoidable delay on flight f_i). For the on-time performance case, an airline's objective is to maximize the number of flights that are delayed at most 15 minutes. Restricting the airlines' 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 formulation (2) with coefficients

$$w_{ij} = \begin{cases} M & \text{if } (t_j - oag_i) < 15, \\ 0 & \text{otherwise.} \end{cases}$$

with $M >> 0$. A slightly different approach will be discussed in Section 4.1.1.

Let us now consider applying the trading model introduced in the previous section, under a scenario where airlines submit any offers that would improve on-time performance. An offer of this type would state that an airline is willing to move down any flight not arriving on-time in the current allocation in return for a delay reduction that makes another flight arrive on time. Note that such offers would never involve a flight that is already arriving on time in the current allocation. In this case, we can distinguish three possible 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 (i.e. a default offer), and (3) a flight will be assigned at most the last slot in the GDP (i.e. the flight is delayed).

As before, the resulting classes can be represented by a sequence of slot indices. If a flight f_i is already arriving on time only the default offer applies, and therefore we have $k_i = 1$ and $i_1 = i$. For all other flights, we have $k_i = 3$ and the slot indices for the three classes outlined above can be defined as $i_1 = \max_{s_j \in S: t_j - oag_i < 15} j$, corresponding to on-time arrivals, $i_2 = i$, corresponding to the default offers, and $i_3 = n$ corresponding to the flight delays (recall that n is the index of the latest slot in the GDP). These class definitions suggest replacing constraints (3e) and (3f) with:

$$\sum_{f_i \in F_a} x_{i3} \leq \sum_{f_i \in F_a} x_{i1} \quad \text{for all } a \in A, \tag{3g}$$

which ensures that for each airline only offers which improve on-time performance are executed. Thus, since airlines are willing to arbitrarily trade off additional delays for an improvement in on-time performance it is possible to represent the acceptable offers with a single constraint for each airline.

Finally, we note that the choice of an objective function comprises several possibilities. The obvious choice is to maximize the number of flights that have been moved on time (i.e., Maximize $\sum_{f_i \in F} x_{i1}$). However, since each delayed flight will induce another flight to arrive on time, one can also consider maximizing the number of flights that have been delayed (i.e., Maximize $\sum_{f_i \in F} x_{i3}$). Finally, we can consider minimizing the number of default offers executed (i.e., Minimize $\sum_{f_i \in F} x_{i2}$) as this maximizes the execution of offers in which flights are moved from their current slot. The impact of these first two possibilities is discussed in Section 4.1.1 below.

4.1.1 Slot Trading Benefits under the On-Time Performance Objective

Figures 8 and 9 give the results of our experiments under the on-time performance objective. Figure 8 shows the improvement in on-time performance relative to the best 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, i.e. as a result of solving the optimization model described in ii) above. The dashed line represents the relative improvements that would be obtained by 2-for-2 slot trading, i.e. as a result of solving the optimization problem described in iii) above, using an objective function of maximizing the number of flights moved up and constraint (3g). The dotted line represents the improvements obtained by the Compression Algorithm, i.e. as a result of solving the optimization model described in iv) above. On average, the

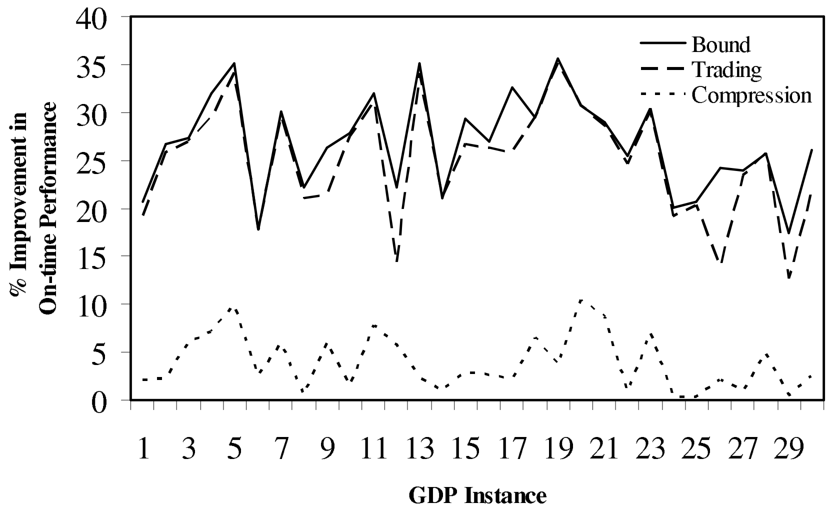

Figure 8: On-time Performance Improvements from Slot Trading

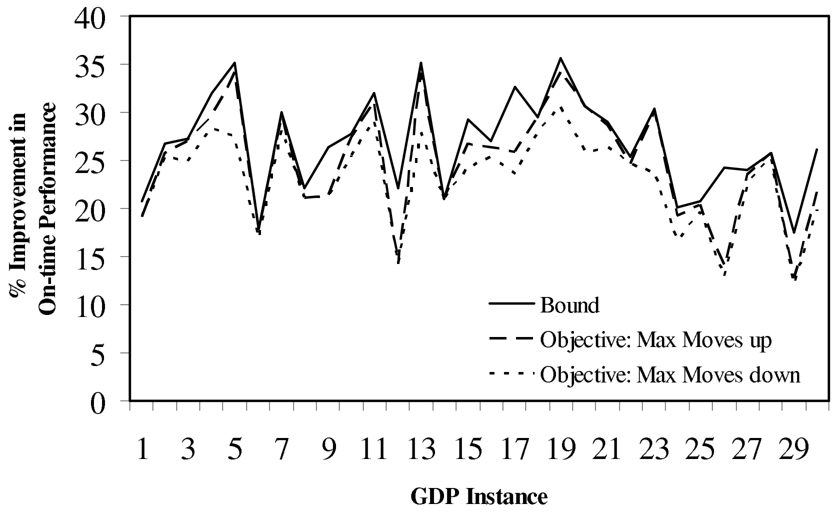

Figure 9: Impact of Objective Function Choice

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.

We note, however, that the objective function choice impacts the performance of the resulting IP model. Figure 9 shows the relative improvements in on-time performance under two objective function alternatives, 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 yield substantial improvements, however the use of the first objective leads to slightly better results and yields significantly faster running time; indicating that, 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. The average solution time when the number of flights moved up is maximized was 8.3 seconds. However, the number of flights moved down is maximized, the average solution time increased to 163 seconds, a most significant difference. It is important to note that we limited the maximum number of nodes visited in the branch and bound tree to 1000; while the node limit was reached occasionally when the number of flights moved down was maximized, the removal of this restriction did not impact the solution values shown in Figure 9. That is, the optimal solution was found within the node limit in all cases.

4.1.2 Slot Trading Dynamics

In this case study, we have so far 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 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_{ij} of assigning flight f_i to slot s_j in formulation (2). More specifically, we define the coefficients w_{ij}, whose units are minutes, as

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

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 penalty cost (reduction in value). We assume that the value decreases linearly in the additional delay up to a maximum delay of 120 minutes, i.e. there is no benefit associated with reducing the delay of flights if the delay remains over 120 minutes. Note that M should be set to a value much greater than 120.

Given this value function for each flight, the number of offers submitted by an airline $a \in A$ can be limited by the specification of aspiration levels $v_a \geq 0$. An aspiration level v_a signifies that an airline will only agree to delay a flight if the increase in delay is no more than the aspiration level v_a. Of course, an airline would only accept this delay if another of its flights were to arrive on time. As in Section 4.1, therefore, we still recognize three classes for flights that are not arriving on time. In the presence of aspiration levels, however, we define the class i_3 corresponding to flight delays for a flight $f_i \in F_a$ as $i_3 = \max_{s_j \in S: w_{ii} - w_{ij} \leq v_a} j$. Note that aspiration levels consider increases in delay cost with respect to a flight's current arrival time, that is, the arrival time after RBS and an airline's internal schedule adjustments have been executed.

With aspiration levels, an airline will no longer accept an arbitrary increase in delay in return for an increase in on-time performance. As such, aspiration levels provides a simple

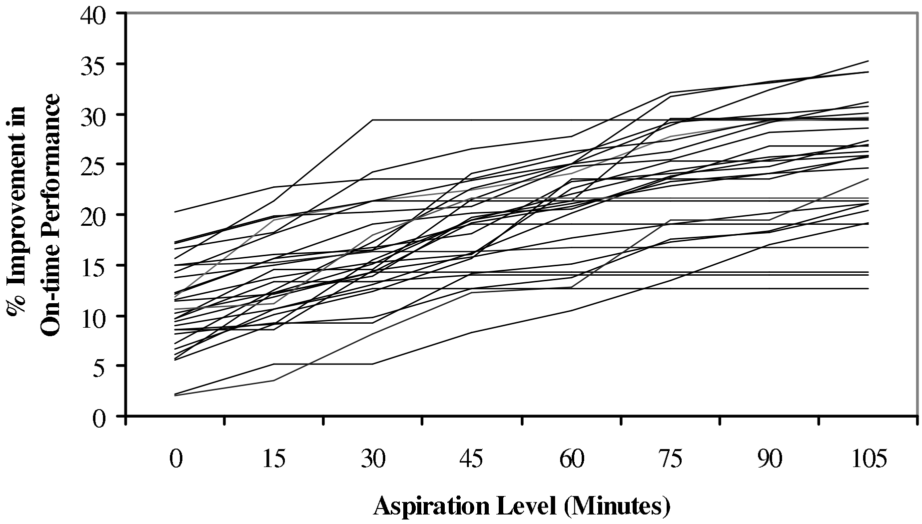

Figure 10: On-time Performance Improvements by Aspiration Levels

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 10 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 10 corresponds to a single GDP. These results show that even for small aspiration levels (e.g. 0 to 30 minutes), considerable improvements in on-time performance can be obtained. It is interesting to note 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 when the number of offers is limited.

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. A flight's passenger delay is defined as the total number of passengers on that flight, multiplied by the delay assigned to that flight. As such, the trade-offs that occur in this case will typically consider the overall benefits of delaying flights with few passengers while reducing the delays of more heavily loaded flights. There are several possibilities to incorporate passenger delay minimization into the airlines' performance function shown in formulation (2). A straightforward approach is to define the value w_{ij} of assigning flight f_i to slot s_j in formulation (2) as

$$w_{ij} = -p_i(t_j - oag_i),$$

where p_i represents the number of passengers on flight f_i (note that the coefficient is negative so as to represent a cost). We will refer to an objective function based on these coefficients as a standard passenger delay objective. Another, somewhat more involved approach, is based on the staircase cost structure shown in Figure 6. This applies in situations where the exact slot assignment is less important than the specific interval in which the flight is assigned, based on the occurrence of operationally significant events at certain time epochs (see Brennan, 2000). In this approach, we define for each flight f_i a set of *critical times*, $\bar{t}_{i1} \ldots, \bar{t}_{ik_i}$, which define slot classes, $1, ..., k_i$, corresponding to the interval boundaries of the cost structure shown in Figure 6. Given these critical times, the value w_{ij} of assigning flight f_i to slot s_j in formulation (2) can be defined as

$$w_{ij} = \begin{cases} 0 & \text{if } (t_j - oag_i) \leq \bar{t}_{i,1}, \\ -p_i\bar{t}_{ik} & \text{if } \bar{t}_{ik} < (t_j - oag_i) \leq \bar{t}_{i(k+1)},\ 1 < k < k_i. \end{cases}$$

We will refer to an objective function based on these coefficients as a critical times passenger delay objective.

Recall that the on-time performance objective allowed us to represent the offers proposed using a single constraint. On the other hand, due to the increase in the number of flight classes that can be considered in the airlines' offers when minimizing passenger delay costs, it is no longer feasible to represent the set of offers that may be proposed by means of a single constraint. If airlines use passenger delay minimization as their objectives, the potential number of offers can be substantial: in principle, an airline may submit any "at least, at most" offer that will reduce its passenger delay cost. That is, for any two flights $f_d, f_u \in F_a$ and any two slots $s_{d'}, s_{u'}$ such that $t_d < t_{d'}$, $t_{u'} < t_u$, and $w_{d,d} - w_{d,d'} < w_{u,u'} - w_{u,u}$, the offer

$(f_d, s_{d'}; f_u, s_{u'})$ might be considered desirable. However, many of these offers are redundant and thus there is no need to consider them.

Suppose, for instance, that an airline submits two offers $(f_d, s_{d_1}; f_u, s_{u'})$ and $(f_d, s_{d_2}; f_u, s_{u'})$ with $t_{d_1} < t_{d_2}$. That is, an airline would be willing accept at most slot s_{d_1} or at most slot s_{d_2} for flight f_1, in return for giving flight f_u at least slot $s_{u'}$. Clearly, the first offer, $(f_d, s_{d_1}; f_u, s_{u'})$, will be redundant. A similar situation occurs with two offers $(f_d, s_{d'}; f_u, s_{u_1})$ and $(f_d, s_{d'}; f_u, s_{u_2})$ with $t_{u_1} < t_{u_2}$. That is, an airline would be willing accept at most slot $s_{d'}$ for flight f_d, in return for giving flight f_u at least slot s_{u_1} or at least slot s_{u_2}. Any of these redundant trades can be removed from consideration. In addition, the critical arrival times further reduce the number of potential offers. It is safe to assume that trades will only involve slots corresponding to critical times (i.e. we could define an at-least class corresponding to all slots with a delay less than 15 minutes, etc.). In order to reduce the number of offers considered, we also applied the definition of classes just presented when considering the standard passenger delay objective (of course, the relative value of these classes was different under the two objectives). Thus, the potential offers evaluated in formulation (3) was the same for both the standard and critical times passenger delay objectives. When evaluating both the passenger delay objectives, we employed the same objective function in formulation (3), i.e. to maximize the number of flights moved up.

4.2.1 Slot Trading Benefits under the Passenger Delay Cost Objective

To analyze the benefits of our slot trading approach when the objective is to minimize passenger delay costs, we again 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, and collected the program data after RBS was first executed. In addition, we estimated the number of passengers per flight by considering capacities of the various aircraft types (Bhogadi, 2002), assuming a load factor of 75%. The initial flight-slot assignment of the remaining (non-exempt) flights was used to determine each airline's initial allotment of slots, S_a.

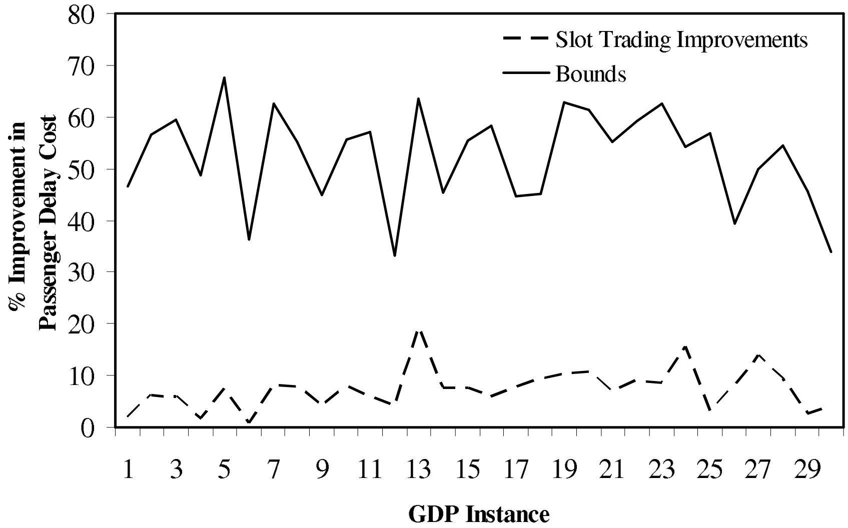

Figure 11: Slot Trading Benefits with Standard Passenger Delay Objective

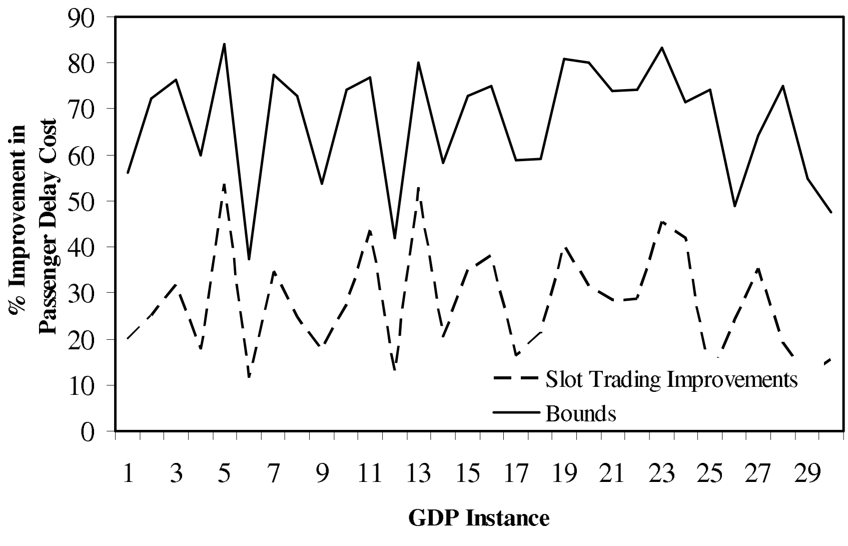

Figure 12: Slot Trading Benefits with Critical Times Passenger Delay Objective

Subsequently, we first considered the slot trading benefits that could be achieved when airlines use the standard passenger delay objective. As before we started by minimizing the passenger delay cost for each individual airline using formulation (2), corresponding to the airlines' internal substitution process, i.e. model i) described in the introduction to Section 4. After that, we generated trade offers and applied the mediation problem to the allocation that would have resulted after each airline minimized its individual passenger delay costs, i.e. model iii) from the introduction to Section 4. In addition, we obtained a bound on the overall benefits by minimizing passenger delay cost without considering slot ownership corresponding to the centralized optimization approach, i.e. model ii) from the introduction to Section 4. The results are shown in Figure 11. Here, the dashed line shows the relative improvement achieved by the slot trading model, while the solid line represents the bounds on potential improvements that are possible. The potential reductions in passenger delay costs are substantial, ranging from 33% to 67% relative to the passenger delay costs produced after each airline minimized its individual delay costs (the average reduction equals 52%). The average reduction in passenger delay obtained by slot trading is approximately 7.7%. This represents approximately 15% of the average potential reduction.

The same analysis was also performed for the critical times passenger delay objective, and the results are shown in Figure 12. The critical times we used in the objective function for formulation (2) are $15, 30, 45, 75,$ and 120 minutes. Again, the dashed line shows the relative improvements achieved by the slot trading model, while the solid line represents the bounds on potential improvements that are possible. In this case, the bounds on the reduction in passenger delay costs range from 37% to 84%, and the average reduction equals 67%. The average reduction in passenger delay costs produced by the 2-for-2 slot trading model, i.e. the dashed line in Figure 12, equals 36%. This represents approximately 54% of the average improvement achieved by the bounds solution.

We see that for the standard passenger delay objective the trading model was only able to achieve a very small percentage of the benefits obtained by the centralized optimization model, whereas for the critical times passenger delay objective the trading model achieved over 50% of the centralized benefits. In addition to the improvement in passenger delay costs, we also determined the number of flights that were moved up (i.e., assigned to an earlier class) for both alternatives. For the standard passenger delay case, an average of 83 flights were moved up, compared to an average of 93.5 flights moved up for the critical times case. Since the number of flights moved up is a measure of trading activity we can see that, under both airline objectives, there was substantial activity. Finally, it is important to note that the bounds produced by the centralized optimization approach are infeasible in a system that operates under the CDM paradigm. For both passenger delay objectives, globally optimal allocations are usually achieved by assigning the majority of delay to flights employing smaller aircraft. Thus, the centralized approach penalized airlines that operate smaller aircraft in favor of airlines operating larger aircraft. This violates the FAA's equal access policy, a crucial component of CDM and the FAA's air traffic management philosophy.

With respect to the efficiency of solutions under the passenger delay cost alternatives considered, we note that using the critical times version of the objective function of formulation (2) had little impact on the performance of formulation (3) in the on-time performance study experiments. Specifically, the average solution time for the critical times passenger delay case was 12.5 seconds. This compares well to the average solution time of 8.3 seconds achieved for the on-time performance case when the objective was to maximize the number flights moved up. However, the average running time under the standard passenger delay costs objective increased significantly to 120 seconds. We have been unable to explain this increase in running time but note that even 120 seconds is a relatively modest time, which would be acceptable within an application setting.

5 Conclusions

This paper explores the potential benefits of increased coordination during GDPs. We introduced a general bartering framework in which airlines may engage in slot trading by submitting offers to the FAA, which acts as the central coordinator. We proposed an optimization model for the FAA's mediation problem, which generalizes current slot exchange procedures in that it allows airlines to submit so-called "at-least, at-most" offers. These offers to trade slots may be viewed as tradeoffs between pairs of flights, and 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 benefits that approached the theoretical bound. Moreover, the level of performance achieved remained even when the number of proposed offers was limited. Significant benefits from trading were also obtained for the case where airlines use passenger delay minimization as their objectives, in particular if passenger delay costs increased according to a step-wise functional form.

We feel that the results of this paper provide strong evidence of the practical usefulness of the slot trading models defined. We are now working on simplifying data input requirements so that these capabilities can be incorporated into Flight Schedule Monitor (FSM), the CDM decision support tool. Further research in this area is warranted to support this move toward implementation and also the development of more advanced models.

While the IP formulation for the mediation problem is quite efficient, initial results show that alternative formulations may substantially improve IP performance (Vossen, 2002), suggesting that further research into more efficient models seems merited. In any implementation of the trading models discussed, airlines would have substantial flexibility in defining offers. Thus, it would be worthwhile to study more complex models of decision-making at the airline level with the goal of developing airline trading strategies and evaluating how these strategies affect overall system performance. Finally, the results obtained with the passenger delay cost objectives suggest using exchanges that involve side payments as a possible way of compensating carriers operating smaller aircraft for absorbing unilateral increases in delay.

Even though adding side payments increases the complexity of trading models significantly and also represents a major change in the underlying business paradigm, such an approach is consistent with recent moves toward market mechanisms on the part of the air traffic management community (see Ball, Donohue and Hoffman, 2004).

Acknowledgments

This work was supported by NEXTOR, the National Center of Excellence for Aviation Operations Research, under Federal Aviation Administration Research Grant Number 96-C-001 and contract number DFTA03-97-D00004. Any opinions expressed herein do not necessarily reflect those of the FAA or the U.S. Department of Transportation. We thank the editor and two referees for detailed and constructive comments that have significantly improved this paper.

References

Adams, M., S. Kolitz , J. Milner, A. Odoni. 1997. Evolutionary Concepts for Decentralized Air Traffic Management. *Air Traffic Control Quart.* **4** 281-306.

Andreatta, G., L. Brunetta. 1993. Multiairport Ground Holding Problem: A Computational Evaluation of Exact Algorithms. *Oper. Res.* **46** 57-64.

Andreatta, G., G. Romanin-Jacur. 1987. Aircraft Flow Management Under Congestion. *Transportation Sci.* **21** 249-253.

Andreatta, G., A.R. Odoni, O. Richetta. 1993. Models for the Ground Holding Problem. L. Bianco, A.R. Odoni, eds. *Large-Scale Computation and Information Processing in Air Traffic Control*, Springer-Verlag, Berlin, Germany, 125-168.

Ball, M.O., G. Donohue and K. Hoffman. 2004. Auctions for the Safe, Efficient and Equitable Allocation of Airspace System Resources, to be published in Cramton, P., Y. Shoham and R. Steinberg, eds., *Combinatorial Auctions.*

Ball, M.O., R. Hoffman, W. Hall, A. Muharremoglu. 1998. Collaborative Decision Making in Air Traffic Management: A Preliminary Assessment. *NEXTOR tech. rep. RR-99-3.*

Ball, M.O., R. Hoffman, D. Knorr, J. Wetherly, M. Wambsganss. 2000. Assessing The Benefits Of Collaborative Decision Making In Air Traffic Management. *Proc. 3rd USA/Europe air Traffic Management R&D Sem.*, http://atm2001.eurocontrol.fr.

Ball, M.O., R. Hoffman, A. Odoni, R. Rifkin. 2003. Efficient Solution of a Stochastic Ground Holding Problem. *Oper. Res.* **51** 167-171.

Bertsimas, D., S. Stock-Patterson. 1998. The Air Traffic Flow Management Problem with Enroute Capacities. *Oper. Res.* **46** 406-422.

Bertsimas, D., S. Stock-Patterson. 2000. The Traffic Flow Management Rerouting Problem in Air Traffic Control: A Dynamic Network Approach. *Transportation Sci.* **34** 239-255.

Bhogadi, N. 2002. Personal Communication.

Brennan, M. 2001. Simplified Substitutions - Enhancements to Substitution Rules and Procedures During Ground Delay Programs. *Proc. AGIFORS Airline Operations Meeting.*

Chang, K., K. Howard, L. Shisler, M.C. Wambsganss, M. Tanino, R. Oiesen. 2001. Enhancements to the FAA Ground-Delay Program Under Collaborative Decision Making. *Interfaces* **31** 57-76.

Hall, W.D. 1999. Efficient Capacity Allocation in a Collaborative Air Transportation System. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Howard, K. 2001. Slot Credit Substitutions Issues. Internal CDM Memorandum.

Luo, S., G. Yu. 1998. Airline Schedule Perturbation Problem: Landing and Takeoff with Non-Splittable Resource for the Ground Delay Program. G. Yu, ed. *Operations Research in the Airline Industry*, Kluwer Academic Publishers, 404-432.

Niznik, T. 2001. Optimizing the Airline Response to Ground Delay Programs. *Proc. AGIFORS Airline Operations Meeting.*

Nolan, M.S. 1999. *Fundamentals of Air Traffic Control.* Brooks/Cole Publishing Company, Pacific Grove, CA.

Odoni, A. 1987. The Flow Management Problem in Air Traffic Control. A. Odoni, L. Bianco, G. Szego, eds. *Flow Control of Congested Networks.* Springer-Verlag, Berlin, Germany, 269-288.

Richetta, O., A. Odoni. 1993. Solving Optimally the Static Ground Holding Policy Problem in Air Traffic Control. *Transportation Sci.* **24** 228-238.

Richetta, O. 1995. Optimal Algorithms and a Remarkably Efficient Heuristic for the Ground-Holding Problem in Air Traffic Control. *Oper. Res.* **43** 758-770.

Terrab, M. 1990. Ground Holding Strategies for Air Traffic Control. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Terrab, M., Odoni, A. 1993. Strategic Flow Management for Air Traffic Control. *Oper. Res.* **41** 138-152.

Vasquez-Marquez, A. 1991. American Airlines Arrival Slot Allocation System (ASAS). *Interfaces* **21** 42-61.

Vossen, T. 2002. Fair Allocation Methods in Air Traffic Management. Ph.D. thesis, University of Maryland, College Park, MD.

Vossen, T., Ball, M. 2003. Optimization and Mediated Bartering Models for Ground Delay Programs. Working paper, University of Maryland, College Park, MD, http://bmgt1-notes.umd.edu/faculty/km/papers.nsf.

Vranas, P., Bertsimas, D., Odoni A. 1994. The Multi-Airport Ground Holding Problem in Air Traffic Control. *Oper. Res.* **42** 249-261.

Vranas, P., Bertsimas, D., Odoni A. 1994. Dynamic Ground-Holding Policies for a Network of Airports. *Transportation Sci.* **28** 275-291.

Wambsganns, M. 1996. Collaborative Decision Making Through Dynamic Information Transfer. *Air Traffic Control Quart.* **4** 107-123.