



Transportation Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Fairness and Collaboration in Network Air Traffic Flow Management: An Optimization Approach

Dimitris Bertsimas, Shubham Gupta

To cite this article:

Dimitris Bertsimas, Shubham Gupta (2015) Fairness and Collaboration in Network Air Traffic Flow Management: An Optimization Approach. Transportation Science

Published online in Articles in Advance 01 Apr 2015

<http://dx.doi.org/10.1287/trsc.2014.0567>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2015, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Fairness and Collaboration in Network Air Traffic Flow Management: An Optimization Approach

Dimitris Bertsimas

Sloan School of Management and Operations Research Center, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, dbertsim@mit.edu

Shubham Gupta

Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, shubhamg@mit.edu

Air traffic flow management (ATFM) attempts to maintain a safe and efficient flow of aircraft given demand-capacity mismatches while ensuring an equitable distribution of delays among stakeholders. There has been extensive research addressing network effects (such as the presence of multiple airports, sectors, and connectivity requirements) in ATFM, but it has not explicitly incorporated the equitable distribution of delays, as well as work on the equitable distribution of delays in a single-airport setting, such as ration-by-schedule (RBS) as introduced under the collaborative decision-making paradigm. In this paper, we develop a two stage approach for network ATFM that incorporates fairness and airline collaboration. In Stage 1, we propose a discrete optimization model that attempts to incorporate an equitable distribution of delays among airlines by introducing a notion of fairness in network ATFM models—controlling the number of reversals and total amount of overtaking, which is a natural generalization of RBS. For two flights f and f' , a reversal occurs when flight f' arrives before f , when f was scheduled to arrive before f' . In the event a reversal occurs, the number of time periods between the arrival times constitutes overtaking. In Stage 2, we allow for airline collaboration by proposing a network model for slot reallocation. We provide extensive empirical results of the proposed optimization models on national-scale, real-world data sets spanning six days that show interesting trade-offs between fairness and efficiency. We report computational times of less than 30 minutes for up to 25 airports and provide theoretical evidence that illuminates the strength of our approach.

Keywords: air traffic flow management; collaborative decision-making; discrete optimization; fairness in allocation

History: Received: May 2012; revisions received: April 2013, March 2014; accepted: May 2014. Published online in *Articles in Advance*.

1. Introduction

The sustained growth of the aviation industry has put a significant strain on the available resources of the air transportation system. This is evidenced by significant flight delays and severe congestion at airports. In 2011, approximately 20% of the flights in the United States were delayed by more than 15 minutes, while another 2% were cancelled (U.S. Department of Transportation 2012). Moreover, for the calendar year 2010, 99 million minutes of system delay led to an estimated \$6.5 billion in direct operating costs for U.S. airlines (Airlines for America 2012).

Air traffic flow management (ATFM) refers to the set of strategic processes that try to reduce congestion costs and support the goal of safe, efficient, and expeditious aircraft movement. ATFM procedures try to resolve local demand-capacity mismatches by adjusting the aggregate traffic flows to match scarce capacity resources. Ground delay programs (GDPs) are one of the most sophisticated ATFM initiatives currently in use that attempt to address airport arrival capacity

reductions. Under this mechanism, delays are applied to flights at their origin airports that are bound for a common destination airport that is suffering from reduced capacity or excessive demand. The premise for this tool is that it is better to absorb delays for a flight while it is grounded at its origin airport rather than incurring airborne delay near the affected destination airport, which is more costly in terms of fuel costs. An airspace flow program (AFP) is another tool that was introduced in 2006 and is similar to GDPs in terms of operational details. The U.S. Federal Aviation Administration (FAA) uses this tool to control arrival rate into a flow constrained area (FCA), such as a weather affected segment of the airspace. Some of the other ATFM tools include assigning airborne delays, dynamic rerouting, and speed control. We briefly review the literature on the existing ATFM tools below.

Odoni (1987) first conceptualized the problem of scheduling flights in real time to minimize congestion costs. Thereafter, several models have been proposed

to handle different versions of the problem. The problem of assigning ground delays in the context of a single airport (the *single-airport ground-holding problem*) has been studied in Terrab and Odoni (1993), Richetta and Odoni (1993), and Richetta and Odoni (1994), and in the multiple airport setting, the *multiairport ground-holding problem* has been studied in Terrab and Paulose (1993) and Vranas, Bertsimas, and Odoni (1994). The problem of controlling release times and speed adjustments of aircraft while airborne for a network of airports taking into account the capacitated airspace (*ATFM problem*) has been studied in Bertsimas and Stock-Patterson (1998); Helme (1992); and Lindsay, Boyd, and Burlingame (1993). The problem with the added complication of dynamically rerouting aircrafts (*ATFM rerouting problem*) was first studied by Bertsimas and Stock-Patterson (2000). Recently, Bertsimas, Lulli, and Odoni (2011) have presented a new mathematical model for the ATFM problem with dynamic rerouting that has superior computational performance. For a detailed survey of the various contributions and a taxonomy of all of the problems, see Bertsimas and Odoni (1997) and Hoffman, Mukherjee, and Vossen (2012). It is important to highlight that network formulations present significant challenges in computational tractability, and so far, no network models have been able to incorporate equity considerations effectively. In fact, some other papers have focused exclusively on addressing the computational challenges of the network problem. For example, Rios and Ross (2010) study the application of a massively parallel Dantzig-Wolfe decomposition applied to a linear integer program and report three-hour planning horizons for nationwide, real-time, optimal traffic flow scheduling. Similarly, Churchill, Lovell, and Ball (2009) build on the model developed in Bertsimas, Lulli, and Odoni (2011) to evaluate a new formulation for large-scale ATFM for more aggregate-level modeling. Apart from the computational difficulties, network models have not successfully transitioned into practice because of the lack of an acceptable notion of fairness among airlines. Barnhart et al. (2012) develop a fairness metric that measures deviation from first-scheduled, first-served (FSFS) and propose a discrete optimization model that approximately minimizes this metric. They further develop an exponential penalty approach and report encouraging computational results using simulated regional and national scenarios. In contrast, we provide an exact method to model overtaking within an optimization framework and introduce another, although related, metric of fairness—controlling the number of reversals. In addition, our approach consists of a slot reallocation phase that enables airlines to implicitly trade their internal objective functions to derive more utility.

We review next other important concepts in the ATFM literature as well: *collaborative decision-making* (CDM) and *ration-by-schedule* (RBS). The decision-making responsibilities in ATFM initiatives are shared between a number of stakeholders (primarily, airlines and the FAA). This poses a major challenge because their actions are highly interdependent and demand real-time exchange of information between the FAA and the airlines. This realization of enhanced cooperation between the various stakeholders led to the adoption of CDM philosophy (Ball et al. 2001; Wambsganss 1996) by the FAA in the 1990s. Under CDM, all ATFM initiatives are conducted in a way that gives significant decision-making responsibilities to airspace users (see Hoffman, Mukherjee, and Vossen 2012 for details on CDM). All recent efforts to improve ATFM have been guided by this philosophy. In the United States, RBS is the fundamental principle for all of the CDM initiatives. Under this paradigm, arrival slots at airports are assigned to flights in accordance with a FSFS priority discipline (see Ball et al. 2001 and Wambsganss 1996 for details on rationing). In the case of GDP and AFP planning, all stakeholders have agreed that this principle is fair to all parties. Based on the slots allotted, an airline is allowed to swap flights among the set of slots it owns. This allocation process is followed by a *compression algorithm*, which fills open slots created by flights that are cancelled. The compression procedure gives airlines an incentive to report accurate flight information by rewarding them for reporting cancellations. The combined process, RBS plus compression (formally called RBS⁺⁺) is the policy currently in use for slot allocations. Despite the use of RBS in a GDP setting, there have been no network models that satisfy the RBS principle in a multiairport setting. This is because applying RBS to each of the airports individually might not lead to a schedule that preserves time, sector, and flight connectivities. In addition, the imposition of a maximum permissible delay on each flight (as required by any tractable optimization model) would mean that a feasible solution under RBS might not even exist if the capacity reduction at some airports is significant. Hence, there is no straightforward extension of RBS from a single-airport setting to an airspace context. (We give a concrete example in §5 to elaborate on this point.)

As part of the CDM philosophy, researchers have also explored dynamic interaction with airlines. Toward this aim, Vossen and Ball (2006a, b) have studied opportunities for slot trading in a single-airport setting, where the aim is to formalize an optimization problem for the FAA given the offers to trade from various airlines. Similarly, Sherali et al. (2011) study slot exchange mechanisms in an AFP through a mediated bartering of assigned slots to allow airlines to improve flight efficiencies.

In this paper, we develop a two-stage approach for network ATFM that incorporates fairness and airline collaboration. In Stage 1, we propose a discrete optimization model that attempts to incorporate the equitable distribution of delays among airlines by introducing a notion of fairness in network ATFM models—controlling the number of reversals and total amount of overtaking, which is a natural generalization of RBS. In Stage 2, we allow for airline collaboration by proposing a network model for slot reallocation. More specifically, our overall approach consists of the following two stages:

Stage 1: Network ATFM model incorporating fairness. We augment the classic ATFM models (Bertsimas and Stock-Patterson 1998) to incorporate fairness considerations for airlines by generalizing the RBS principle (currently operational in a GDP and AFP setting). The objective function used in the existing network ATFM models is to minimize the total delay costs across all flights—i.e., the focus is on overall system efficiency. A disadvantage of such an approach is that the solution to such models can have a large number of reversals, meaning the resulting order of flight arrivals could be quite different than the published flight schedules. Specifically, for two flights f and f' arriving at the same destination airport, a *reversal* occurs if f was scheduled to arrive before f' but f' arrives before f in the actual sequence. Moreover, across these reversals, there might be a different number of time periods of overtaking. Concretely, the duration between the arrival of flights f' and f constitutes the amount of *overtaking*. Consequently, the total overtaking across these reversals might be large. Because of this deviation from the original flight ordering, it becomes difficult to implement such a solution because of the coupling in the crew assignments and the use of hub and spoke networks. Thus, the appropriate fairness criterion is to control the number of reversals and amount of overtaking. We propose discrete optimization models that add these fairness controls. The key output in this stage is the assignment of flights to different time periods.

Stage 2: Slot reallocation through airline collaboration. We generalize the notion of intra-airline exchange of arrival slots, a key component of the current CDM practice in a single-airport setting to network-wide slot reallocation among airlines. Specifically, we propose an optimization model that takes as input the assignment of flights to different time periods from Stage 1 and permits the airlines to trade these assigned slots across different airports, thereby resulting in improved internal objective functions. In our model, airlines submit AMAL (at-most, at-least) trade offers (initially proposed by Vossen and Ball 2006a and also used recently in Sherali et al. 2011). An AMAL offer is of the form $(f_1, t_1; f_2, t_2)$, which means

that the airline is willing to move flight f_1 to a later time period, but no later than t_1 ; in return for moving flight f_2 to an earlier time period, but no later than t_2 . An AMAL offer is attractive because of simplicity in data input and the flexibility of an individual offer that allows multiple combinations of viable trades. (We elaborate on this later in §4.) The model proposed for Stage 2 of our approach allows airlines to react to the schedule determined in Stage 1 by taking into account their flights in the entire network and making appropriate trade-offs.

Given the wide acceptance and implementation of the CDM philosophy, it becomes imperative to design new ATFM models that are amenable to the CDM framework. In this spirit, we briefly describe here how our approach can be integrated in a practical operational framework (like the current RBS⁺⁺ procedure for GDP planning). A more detailed example is provided in §5.

1.1. Integration with Current CDM Practice

Currently, there are three key phases involved in the coordination of various ATFM initiatives (like GDPs and AFPs):

1. *RBS for each ATFM program.* The FAA invokes the RBS policy to allocate arrival slots to the airlines for each ATFM program based on the original schedule ordering.
2. *Airline response to schedule disruption.* Based on the slots allotted, an airline is allowed to make changes to the schedule by cancelling flights and swapping the slots of two or more of its own flights if they are compatible with the scheduled departure times.
3. *Final coordination by the FAA.* The FAA coordinates with the airlines to come up with an overall feasible schedule by combining their proposed changes. This is further complemented by compression (wherein the FAA attempts to fill in any holes created by cancellations to further optimize the final schedule).

Our overall approach is integrable in the above framework as follows:

- *Stage 1: Control reversals/overtaking.* This stage of our approach interfaces with phase 1 of the CDM framework. Rather than applying RBS to each ATFM program, we control the reversals and overtaking in the resulting flight sequences. The input requirements for our models, namely, the set of feasible times that a flight can be in a sector and the capacity inputs are readily available from the flight schedule monitor (FSM). Furthermore, the output of our models can be readily converted to a slot assignment for each flight (by following the scheduled order of the slots allotted for flights during each time period).

• *Stage 2: Airline collaboration.* This stage of our approach interfaces with the last two phases of the CDM framework.

—*Intra-airline Substitutions and Cancellations.* Given an assignment from Stage 1, the airlines can still swap flights among its allotted slots (as long as the resulting schedule is feasible in terms of capacity). Thus, the intra-airline substitution phase in current CDM practice goes through as is in our approach. Note that a major attraction of the RBS principle was the incentive it gave the airlines to report cancellations and delays without worrying about losing slots. As our approach is an extension of RBS, the incentives to the airlines to report cancellations are still present in our framework. For instance, an airline that reports a cancellation will be given precedence in moving up one of the flights it owns.

—*Interairline Slot Reallocation.* The input required from the airlines on the offers to trade are readily available because the airlines know the slots allotted to them from Stage 1. The final exercise of compression goes through as is done presently to fill in gaps resulting from empty slots or flight cancellations.

1.2. Organization of the Paper

Section 2 summarizes an adaptation of the Bertsimas and Stock-Patterson (1998) model to accommodate our approach. Section 3 introduces models of fairness. Section 4 introduces our model of slot reallocation. Section 5 compares and discusses integration issues with the current CDM practices. Section 6 reports computational results of the proposed optimization models on six days of national-scale, real-world data sets. Section 7 summarizes our conclusions, and the appendix reports polyhedral analysis that illuminates the strength of our formulations.

2. Bertsimas Stock-Patterson Model: Notation and Solutions

In this section, we review the widely studied Bertsimas and Stock-Patterson (1998) model for the ATFM problem, which provides the starting point for all of the models presented herein. We use an extended version of the notation used in that paper to accommodate fairness and slot reallocation considerations. Finally, we illustrate difficulties relative to fairness considerations in the solutions obtained from this model.

The Decision Variables. The decision variables are as follows:

$$w_{j,t}^f = \begin{cases} 1, & \text{if flight } f \text{ arrives at sector } j \text{ by time } t, \\ 0, & \text{otherwise.} \end{cases}$$

This definition of the decision variables, using by instead of at , is critical to the understanding of the formulation. The variables are defined only for the set of

sectors an aircraft may fly through on its route to the destination airports. In addition, variables are used for the departure and arrival airports to determine the optimal times for departure and arrival. Because we do not consider flight cancellations, at least two variables can be fixed a priori for each flight: each aircraft has to take off by the end of a feasible time window and has to land, as well, within a feasible time window, which is determined by the time of departure.

Notation. The model's formulation requires a definition of the following notation:

\mathcal{K} :	set of airports,
\mathcal{F} :	set of flights,
\mathcal{T} :	set of time periods,
\mathcal{W} :	set of airlines,
$\mathcal{F}_w \subseteq \mathcal{F}$:	set of flights belonging to airline w ,
\mathcal{S} :	set of sectors,
$\mathcal{S}^f \subseteq \mathcal{S}$:	sequence of sectors flown by flight f ,
\mathcal{C} :	set of pairs of flights that are continued,
\mathcal{R}^j :	set of pairs of flights that are reversible in resource j (definition in §3),
\mathcal{R}^S :	set of pairs of flights that are reversible in sectors (definition in §3),
\mathcal{R}^A :	set of pairs of flights that are reversible at airports (definition in §3),
\mathcal{P}_i^f :	sector i 's preceding sector in the path of flight f ,
\mathcal{L}_i^f :	sector i 's subsequent sector in the path of flight f ,
$D_k(t)$:	departure capacity of airport k at time t ,
$A_k(t)$:	arrival capacity of airport k at time t ,
$S_j(t)$:	capacity of sector j at time t ,
d_f :	scheduled departure time of flight f ,
a_f^j :	scheduled arrival time of flight f in resource j ,
s_f :	turnaround time of an airplane after flight f ,
orig_f :	airport of departure of flight f ,
dest_f :	airport of arrival of flight f ,
l_{fj} :	minimum number of time units that flight f must spend in sector j ,
M :	maximum permissible delay for a flight,
T_j^f :	set of feasible time periods for flight f to arrive in resource j ,
\underline{T}_j^f :	first time period in the set T_j^f ,
\bar{T}_j^f :	last time period in the set T_j^f ,
$T_{f,f',j}^r$:	set of time-periods common for flights f and f' where a reversal could occur in resource j (f arrives before f' in the published schedule),
$O_{f,f',j}^{\max}$:	maximum amount of overtaking possible between flights f and f' in resource j (definition in §3),
\mathcal{O} :	set of all possible trades (definition in §4),
$\mathcal{O}^f \subseteq \mathcal{O}$:	set of offers containing flight f ,

c_f : time period assigned to flight f from
Stage 1 optimization.

The key additions relative to the notation used in Bertsimas and Stock-Patterson (1998) are \mathcal{W} , \mathcal{F}_w , \mathcal{R}^j , \mathcal{R}^S , \mathcal{R}^A , $T_{f,f',j}^r$, $O_{f,f',j}^{\max}$, \mathcal{G} , $\mathcal{G}^f \subseteq \mathcal{G}$, and c_f . This new notation is elaborated on in §§3 and 4.

The Objective Function. We use an adapted expression for the total delay (TD) cost for each flight introduced recently in Bertsimas, Lulli, and Odoni (2011). The total delay (TD) cost is a combination of the costs of airborne delay (AD) and ground-holding delay (GD). ($TD = GD + \alpha \cdot AD$, where $\alpha > 1$, because AD is typically more costly than GD.) By substituting AD in terms of TD (i.e., $AD = TD - GD$), the objective can be rewritten as $\alpha \cdot TD - (\alpha - 1) \cdot GD$.

Consequently, the objective function consists of two terms: a first term that takes into account the cost of the total delay assigned to a flight and a second term that accounts for the cost reduction obtained when a part of the total delay is taken as ground delay at the origin airport. The objective function cost coefficients are a super-linear function of the tardiness of a flight of the form $(t - a_f^k)^{1+\epsilon}$, where ϵ is close to zero. Hence, for each flight f and for each time period t , we define the following two cost coefficients:

$$\begin{aligned} c_{\text{total}}^f(t) &= \alpha(t - a_f^k)^{1+\epsilon}: \text{ total cost of delaying flight } f \\ &\quad \text{for } (t - a_f^k) \text{ units of time,} \\ c_g^f(t) &= (\alpha - 1)(t - d_f)^{1+\epsilon}: \text{ cost reduction obtained} \\ &\quad \text{by holding flight } f \text{ on the ground for } (t - d_f) \\ &\quad \text{units of time.} \end{aligned}$$

The motivation of using super-linear cost coefficients is that it will favor a moderate assignment of total delays between two flights rather than assigning a much larger delay to one versus another flight. In view of the above, the delay cost function for each flight f takes the following form:

$$\begin{aligned} &\sum_{t \in T_{\text{dest}_f}^f} c_{\text{total}}^f(t) \cdot (w_{\text{dest}_f, t}^f - w_{\text{dest}_f, t-1}^f) \\ &\quad - \sum_{t \in T_{\text{orig}_f}^f} c_g^f(t) \cdot (w_{\text{orig}_f, t}^f - w_{\text{orig}_f, t-1}^f). \end{aligned}$$

The TFMP model. The complete description of the model, referred to as TFMP (traffic flow management problem), is as follows:

$$\begin{aligned} \min \quad &\sum_{f \in \mathcal{F}} \left(\sum_{t \in T_{\text{dest}_f}^f} c_{\text{total}}^f(t) \cdot (w_{\text{dest}_f, t}^f - w_{\text{dest}_f, t-1}^f) \right. \\ &\quad \left. - \sum_{t \in T_{\text{orig}_f}^f} c_g^f(t) \cdot (w_{\text{orig}_f, t}^f - w_{\text{orig}_f, t-1}^f) \right) \\ \text{subject to:} \quad &\sum_{f \in \mathcal{F}: \text{orig}_f = k} (w_{k, t}^f - w_{k, t-1}^f) \leq D_k(t), \\ &\forall k \in \mathcal{K}, t \in \mathcal{T}. \quad (1a) \end{aligned}$$

$$\begin{aligned} \sum_{f \in \mathcal{F}: \text{dest}_f = k} (w_{k, t}^f - w_{k, t-1}^f) &\leq A_k(t), \\ &\forall k \in \mathcal{K}, t \in \mathcal{T}. \quad (1b) \end{aligned}$$

$$\begin{aligned} \sum_{f \in \mathcal{F}: j \in \mathcal{F}_f, j' = \mathcal{F}_j^f} (w_{j, t}^f - w_{j', t}^f) &\leq S_j(t), \\ &\forall j \in \mathcal{S}, t \in \mathcal{T}. \quad (1c) \end{aligned}$$

$$\begin{aligned} w_{j, t}^f - w_{j', t-l_{ff'}}^f &\leq 0, \quad \forall f \in \mathcal{F}, t \in T_j^f, \\ j \in \mathcal{S}^f: j \neq \text{orig}_f, j' = \mathcal{P}_j^f. \quad (1d) \end{aligned}$$

$$\begin{aligned} w_{\text{orig}_f, t}^f - w_{\text{dest}_{f'}, t-s_f}^{f'} &\leq 0, \\ \forall (f, f') \in \mathcal{C}, \forall t \in T_{\text{orig}_f}^f. \quad (1e) \end{aligned}$$

$$\begin{aligned} w_{j, t-1}^f - w_{j, t}^f &\leq 0, \\ \forall f \in \mathcal{F}, j \in \mathcal{S}^f, t \in T_j^f. \quad (1f) \end{aligned}$$

$$w_{j, t}^f \in \{0, 1\}, \quad \forall f \in \mathcal{F}, j \in \mathcal{S}^f, t \in T_j^f.$$

Note that $w_{k, T_k^f-1}^f$ is set to 0 ($\forall k \in \mathcal{K}$) at the outset to capture the case when flight f departs in the first time period T_k^f . The first three sets of constraints take into account the capacities of the various elements of the system. Constraints (1a) ensure that the number of flights that may take off from airport k at time t will not exceed the departure capacity of airport k at time t . Likewise, constraints (1b) ensure that the number of flights that may arrive at airport k at time t , will not exceed the arrival capacity of airport k at time t . Finally, constraints (1c) ensure that the total number of flights that may feasibly be in Sector j at time t will not exceed the capacity of Sector j at time t .

The next three sets of constraints capture the various connectivities: sector, flight, and time connectivity. Constraints (1d) stipulate that a flight cannot arrive at Sector j by time t if it has not arrived at the preceding sector by time $t - l_{ff'}$. In other words, a flight cannot enter the next sector on its path until it has spent at least $l_{ff'}$ time units (the minimum possible) traveling through one of the preceding sectors on its current path. Constraints (1e) represent connectivity between flights. They handle the cases in which a flight is continued, i.e., the flight's aircraft is scheduled to perform a subsequent flight within some user-specified time interval. The first flight in such cases is denoted as f' and the subsequent flight as f , whereas s_f is the minimum amount of time needed to prepare flight f for departure, following the landing of flight f' . Constraints (1f) ensure connectivity in time. Thus, if a flight has arrived at element j by time \tilde{t} , then $w_{j, t}^f$ has to have a value of 1 for all later time periods ($t \geq \tilde{t}$).

Solutions from TFMP. Here, we illustrate the difficulties relative to fairness considerations in the solutions obtained from the TFMP formulation. We report

a solution from TFMP for one of the six realistic data sets (derived from Lincoln Lab databases and elaborated later on in §6) on which we have performed our experiments in this paper. This data set comprises 5,092 flights, five airlines, 55 airports, and 100 sectors. The analysis is carried over 96 time periods, each with 15 minute durations. First, the total number of reversals in the resulting solution is 915. Moreover, there are 1,492 units of overtaking across these reversals. To put the number of reversals in perspective, the maximum possible number of reversals is around 9,500. (This is attained when maximum elements of \mathcal{R}^A are reversed with the resulting schedule remaining feasible.) This indicates that the sequence of flight arrivals in the solutions from TFMP differ significantly from the scheduled sequence of flight arrivals. (Note that, if M was large enough, the schedule with 0 reversals would be considered most fair.) These two observations in the solutions obtained from the formulation TFMP are present across all six data sets. In §3, we introduce discrete optimization models that control the number of reversals and the amount of overtaking.

3. Network Models That Incorporate Concepts of Fairness

We first elaborate on the notation introduced to model fairness considerations.

The sets \mathcal{R}^j , \mathcal{R}^S , and \mathcal{R}^A . We first define \mathcal{R}^j (a set of pairs of flights that are reversible in resource j). We make a distinction between the case when the resource j is a sector and when it is an airport.

DEFINITION 1. For an airport $k \in \mathcal{K}$, a pair of flights (f, f') belongs to \mathcal{R}^k if the following two conditions are satisfied:

1. $\text{dest}_f = \text{dest}_{f'} = k$, i.e., the destination airport of both flights f and f' is the same.
2. $a_f^k \leq a_{f'}^k \leq a_f^k + M$, i.e., the scheduled time of arrival of flight f' at the destination airport lies between the scheduled time of arrival of flight f and the last time period in the set of feasible time periods that the flight f can arrive at its destination airport.

DEFINITION 2. Analogously, for a sector $j \in \mathcal{S}$, a pair of flights (f, f') belongs to \mathcal{R}^j if $j \in \mathcal{S}^f \cap \mathcal{S}^{f'}$ and $a_f^j \leq a_{f'}^j \leq a_f^j + M$.

For each pair of flights $(f, f') \in \mathcal{R}^j$, we count a *reversal* if, in the resulting solution, flight f' arrives before flight f in resource j (i.e., there exists a time t such that $w_{j,t}^{f'} > w_{j,t}^f$). We refer to the reversals occurring in sectors as *sector reversals* and the reversals occurring at the airports as *airport reversals*. This clustering is motivated from fairness considerations in a network setting and is elaborated on later in §3.2.

DEFINITION 3. The set of pairs of flights that are reversible in sectors (\mathcal{R}^S) and at airports (\mathcal{R}^A) is defined as

1. $\mathcal{R}^S = \bigcup_{j \in \mathcal{S}} \mathcal{R}^j$,
2. $\mathcal{R}^A = \bigcup_{k \in \mathcal{K}} \mathcal{R}^k$.

The set $T_{f,f',j}^r$ and parameter $O_{f,f',j}^{\max}$.

DEFINITION 4. The set $T_{f,f',j}^r$ (a set of time periods common for flights f and f' where a reversal could occur) is defined as $\{\underline{T}_{j,t}^{f'}, \dots, \bar{T}_{j,t}^f - 1\}$.

To elaborate on Definition 4, $T_{f,f',j}^r$ is the set of time periods t , such that it is possible to have the following assignment: $w_{j,t}^f = 0$ and $w_{j,t}^{f'} = 1$. This assignment would imply that a reversal occurs at time t .

DEFINITION 5. The parameter $O_{f,f',j}^{\max}$ is defined as $|T_{f,f',j}^r|$ (cardinality of the set $T_{f,f',j}^r$) and hence is equal to $\bar{T}_{j,t}^{f'} - \underline{T}_{j,t}^f - 1$.

To elaborate on Definition 4, $O_{f,f',j}^{\max}$ is the maximum amount of overtaking possible between flights f and f' and would be attained when $w_{j,\bar{T}_{j,t}^{f'}-1}^f = 0$ and $w_{j,\underline{T}_{j,t}^f}^{f'} = 1$.

EXAMPLE 1. Figure 1 depicts a reversible pair of flights $(f, f') \in \mathcal{R}^A$. Let $\text{dest}_f = \text{dest}_{f'} = k$. In this example, the arrows correspond to the set of time periods common for both flights. Moreover, the set of time periods marked by these arrows (except for the last one) constitute $T_{f,f',k}^r$. This is because the model enforces $w_{k,a_f^k+M}^f = 1$ at the outset, and hence, it is not possible to have a reversal at $a_f^k + M$. Finally, $O_{f,f',k}^{\max} = |T_{f,f',k}^r| = 6$.

3.1. Minimizing the Total Amount of Overtaking

A notion of fairness widely agreed on by the airlines is to have a schedule that preserves the order of flight arrivals at an airport according to the published schedules. As mentioned in §1, this is known as RBS. However, given capacity reductions at airports, it might not always be possible to have a feasible solution under RBS in a network setting. A good proxy to the RBS solution would be one that

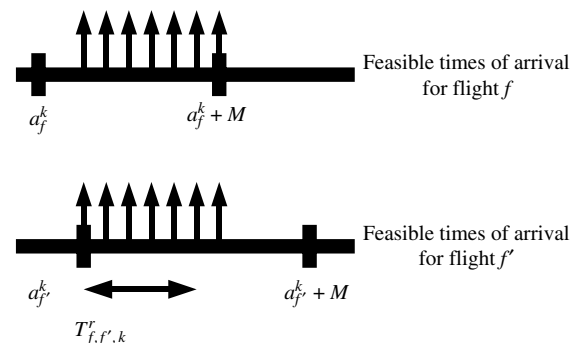


Figure 1 A Reversible Pair of Flights $(f, f') \in \mathcal{R}^A(\text{dest}_f = k)$

has a small amount of overtaking. Hence, in such a scenario, a plan that minimizes the total overtaking while keeping the total delay cost small might be a more desirable solution. As the first approach, we present a model that achieves this objective. It is appropriate to remark that this fairness criterion has considerable intuitive appeal because it naturally captures the priorities inherent in the original airport arrival schedules. For every reversible pair of flights $(f, f') \in \mathcal{R}^j$, let $g_{f,f',j}$ denote the total amount of overtaking between flights f and f' . Then, we need to define the following set of variables to express $g_{f,f',j}$.

$$s_{f,f',j}^i = \begin{cases} 1, & \text{if flight } f' \text{ arrives but } f \text{ does not} \\ & \text{arrive by time } \underline{T}_j^{f'} + i \text{ in resource } j, \\ 0, & \text{otherwise.} \end{cases}$$

The definition above implies the following:

$$s_{f,f',j}^i = 1 \iff \{w_{j, \underline{T}_j^{f'} + i}^f = 0, w_{j, \underline{T}_j^{f'} + i}^{f'} = 1\}.$$

Table 1 summarizes the various feasible combinations of these variables under the above definition. Thus, an alternative way to express $s_{f,f',j}^i$ is as follows:

$$s_{f,f',j}^i = \max\{w_{j, \underline{T}_j^{f'} + i}^{f'} - w_{j, \underline{T}_j^{f'} + i}^f, 0\}. \quad (2)$$

Equation (2) implies the following constraints necessary to express $s_{f,f',j}^i$ in a mathematical programming framework if the objective is to minimize $s_{f,f',j}^i$.

$$s_{f,f',j}^i \geq w_{j, \underline{T}_j^{f'} + i}^{f'} - w_{j, \underline{T}_j^{f'} + i}^f, \quad (3a)$$

$$s_{f,f',j}^i \geq 0. \quad (3b)$$

The set of variables $s_{f,f',j}^i$ are defined for $i \in \{0, \dots, O_{f,f',j}^{\max}\}$. Now, $g_{f,f',j} \in \{0, \dots, O_{f,f',j}^{\max}\}$ can be defined as follows:

$$g_{f,f',j} = \sum_{i=0}^{O_{f,f',j}^{\max}} s_{f,f',j}^i. \quad (4)$$

We shall work with an alternative description of Equation (3a) to make the exposition on overtaking clearer. We substitute $i = t - \underline{T}_j^{f'}$ in Equation (3a) to rewrite it as follows:

$$w_{j,t}^{f'} \leq w_{j,t}^f + s_{f,f',j}^{t - \underline{T}_j^{f'}}. \quad (5)$$

We prove next that the following set of constraints are required to model overtaking between $(f, f') \in \mathcal{R}^j$

if we use an objective function to minimize $g_{f,f',j}$ in addition to $s_{f,f',j}^{t - \underline{T}_j^{f'}} \geq 0, \forall t \in T_{f,f',j}^r$:

$$w_{j,t}^{f'} \leq w_{j,t}^f + s_{f,f',j}^{t - \underline{T}_j^{f'}}, \quad \forall t \in T_{f,f',j}^r. \quad (6)$$

THEOREM 1. *If we use an objective function of minimizing $g_{f,f',j}$ (the total amount of overtaking for $(f, f') \in \mathcal{R}^j$) in addition to $s_{f,f',j}^{t - \underline{T}_j^{f'}} \geq 0, \forall t \in T_{f,f',j}^r$, then constraint (6) correctly captures the semantics of overtaking.*

PROOF. If there is no reversal, i.e.,

$$w_{j,t}^{f'} \leq w_{j,t}^f, \quad \forall t \in T_{f,f',j}^r,$$

then constraint (6) becomes redundant. Because we minimize the total amount of overtaking (and $s_{f,f',j}^{t - \underline{T}_j^{f'}} \geq 0, \forall t \in T_{f,f',j}^r$), it forces

$$s_{f,f',j}^{t - \underline{T}_j^{f'}} = 0, \quad \forall t \in T_{f,f',j}^r,$$

leading to $g_{f,f',j} = 0$. On the contrary, if there are i units of overtaking, then $\exists t \in \{\underline{T}_j^{f'}, \dots, \underline{T}_j^{f'} + O_{f,f',j}^{\max} - i\}$ such that

$$w_{j,t-1}^{f'} = 0, \quad w_{j,t}^{f'} = 1,$$

$$w_{j,t+i-1}^f = 0, \quad w_{j,t+i}^f = 1.$$

Now, the time-connectivity constraints (constraints (1f)) imply that

$$w_{j,t+m}^{f'} = 1, \quad w_{j,t+m}^f = 0, \quad \forall 0 \leq m \leq i - 1.$$

Constraint (6) then enforces

$$s_{f,f',j}^k = 1, \quad \forall t \leq k \leq t + i - 1.$$

Again, because we minimize the total amount of overtaking (and $s_{f,f',j}^{t - \underline{T}_j^{f'}} \geq 0, \forall t \in T_{f,f',j}^r$)

$$s_{f,f',j}^k = 0, \quad \forall 0 \leq k < t, t + i - 1 < k \leq O_{f,f',j}^{\max},$$

leading to $g_{f,f',j} = i$. In summary, constraint (6) (in addition to $s_{f,f',j}^{t - \underline{T}_j^{f'}} \geq 0$), correctly models overtaking if the objective function is to minimize $g_{f,f',j}$. \square

The proof of Theorem 1 relied critically on the assumption that we use an objective function that minimizes $g_{f,f',j}$. Next, we propose a formulation to model overtaking that is independent of the objective function used. We propose a set of constraints that capture the convex hull of the four feasible integer points enumerated in Table 1, namely $(0, 0, 0)$, $(0, 1, 1)$, $(1, 0, 0)$, and $(1, 1, 0)$. Figure 2 depicts the convex hull of these four points. We introduce the following set of constraints to model overtaking:

$$w_{j,t}^{f'} \leq w_{j,t}^f + s_{f,f',j}^{t - \underline{T}_j^{f'}}, \quad \forall j \in \mathcal{P} \cup \mathcal{R}, (f, f') \in \mathcal{R}^j, t \in T_{f,f',j}^r. \quad (7a)$$

Table 1 Modeling the Overtaking Variables

$w_{j, \underline{T}_j^{f'} + i}^f$	$w_{j, \underline{T}_j^{f'} + i}^{f'}$	$s_{f,f',j}^i$
0	0	0
0	1	1
1	0	0
1	1	0

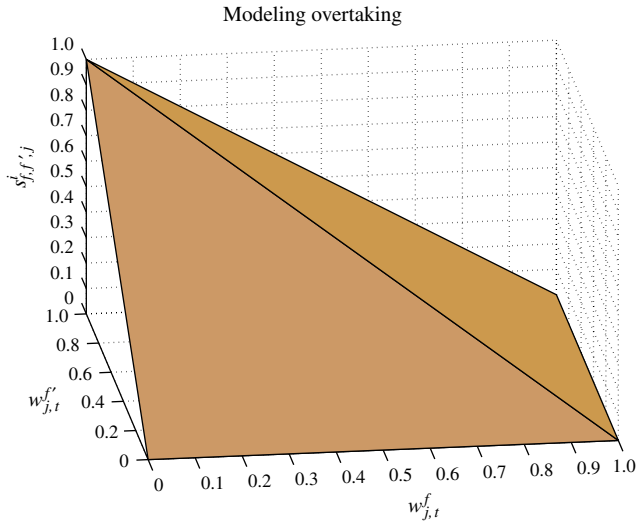


Figure 2 (Color online) Convex Hull of the Integer Points in Table 1 to Model Overtaking ($i = t - T_j^{f'}$)

$$w_{j,t}^f \leq w_{j,t}^{f'} + 1 - s_{f,f',j}^{t-T_j^{f'}}, \quad \forall j \in \mathcal{P} \cup \mathcal{K}, (f, f') \in \mathcal{R}^j, t \in T_{f,f',j}^r. \quad (7b)$$

$$w_{j,t}^f + s_{f,f',j}^{t-T_j^{f'}} \leq 1, \quad \forall j \in \mathcal{P} \cup \mathcal{K}, (f, f') \in \mathcal{R}^j, t \in T_{f,f',j}^r. \quad (7c)$$

$$-w_{j,t}^{f'} + s_{f,f',j}^{t-T_j^{f'}} \leq 0, \quad \forall j \in \mathcal{P} \cup \mathcal{K}, (f, f') \in \mathcal{R}^j, t \in T_{f,f',j}^r. \quad (7d)$$

The TFMP model with the additional control on total amount of overtaking (referred to as TFMP-Overtake henceforth) is as follows:

$$\begin{aligned} \min \quad & \left\{ \sum_{f \in \mathcal{F}} \left(\sum_{t \in T_{\text{dest}_f}^f} c_{\text{total}}^f(t) \cdot (w_{\text{dest}_f,t}^f - w_{\text{dest}_f,t-1}^f) \right. \right. \\ & \left. \left. - \sum_{t \in T_{\text{orig}_f}^f} c_g^f(t) \cdot (w_{\text{orig}_f,t}^f - w_{\text{orig}_f,t-1}^f) \right) \right. \\ & \left. + \lambda_s^0 \cdot \left(\sum_{j \in \mathcal{P}, (f, f') \in \mathcal{R}^j} \sum_{i=0}^{O_{f,f',j}^{\max}} s_{f,f',j}^i \right) \right. \\ & \left. + \lambda_a^0 \cdot \left(\sum_{k \in \mathcal{K}, (f, f') \in \mathcal{R}^k} \sum_{i=0}^{O_{f,f',k}^{\max}} s_{f,f',k}^i \right) \right\} \end{aligned}$$

subject to: (1a)–(1f).

(7a)–(7d).

$$s_{f,f',j}^i \in \{0, 1\}, \quad \forall j \in \mathcal{P} \cup \mathcal{K}, (f, f') \in \mathcal{R}^j, i \in \{0, \dots, O_{f,f',j}^{\max}\}.$$

3.2. Minimizing the Total Number of Reversals

The model introduced in §3.1 took into account the magnitude of overtaking within each reversal. In this section, we introduce a model that controls the total number of reversals.

For each element $(f, f') \in \mathcal{R}^j$, we introduce the following new variable:

$$s_{f,f',j} = \begin{cases} 1, & \text{if there is a reversal,} \\ 0, & \text{otherwise.} \end{cases}$$

Next, we relate the variables $s_{f,f',j}$ (used to model a reversal) to the variables $s_{f,f',j}^i$ (used to model overtaking). It is evident that a reversal occurs if and only if there is at least one time period of overtaking. Mathematically, it translates to the following:

$$s_{f,f',j} = 1 \iff \{\exists i \in \{0, \dots, O_{f,f',j}^{\max}\}, s_{f,f',j}^i = 1\}. \quad (8)$$

Building on Equation (8), we have the following:

$$\begin{aligned} s_{f,f',j} &= \max_{t \in T_{f,f',j}^r} \{s_{f,f',j}^{t-T_j^{f'}}\}, \\ s_{f,f',j} &= \max_{t \in T_{f,f',j}^r} \{\max\{w_{j,t}^{f'} - w_{j,t}^f, 0\}\}, \\ s_{f,f',j} &= \max \left\{ \max_{t \in T_{f,f',j}^r} \{w_{j,t}^{f'} - w_{j,t}^f\}, 0 \right\}. \end{aligned} \quad (9)$$

Equation (9) implies that the following constraints suffice to express $s_{f,f',j}$ in a mathematical programming framework if the objective is to minimize $s_{f,f',j}$:

$$s_{f,f',j} \geq w_{j,t}^{f'} - w_{j,t}^f, \quad \forall t \in T_{f,f',j}^r. \quad (10a)$$

$$s_{f,f',j} \geq 0. \quad (10b)$$

Equation (10a) can be rearranged as follows:

$$w_{j,t}^{f'} \leq w_{j,t}^f + s_{f,f',j}, \quad \forall t \in T_{f,f',j}^r. \quad (11)$$

THEOREM 2. If we use an objective function of minimizing $s_{f,f',j}$, then constraint (11) in addition to constraint (10b) correctly captures the semantics of modeling a reversal.

PROOF. If there is no reversal, i.e.,

$$w_{j,t}^{f'} \leq w_{j,t}^f, \quad \forall t \in T_{f,f',j}^r,$$

then constraint (11) becomes redundant. Because we minimize $s_{f,f',j}$ (and $s_{f,f',j} \geq 0$), it forces $s_{f,f',j} = 0$. On the contrary, if there is a reversal, then there exists a time $t \in T_{f,f',j}^r$ such that

$$w_{j,t}^{f'} = 1, \quad w_{j,t}^f = 0.$$

Constraint (11) then implies that $s_{f,f',j} \geq 1$. Again, minimizing $s_{f,f',j}$ makes $s_{f,f',j} = 1$ ensuring that constraint (11) indeed models a reversal correctly. \square

The proof of Theorem 2 relied critically on the assumption that we use an objective function that minimizes $s_{f,f',j}$. Here, we present a formulation that models a reversal correctly independently of the

objective function used. For each element $(f, f') \in \mathcal{R}^j$, we introduce the following constraints to TFMP:

$$w_{j,t}^{f'} \leq w_{j,t}^f + s_{f,f',j}, \quad \forall j \in \mathcal{J} \cup \mathcal{K}, (f, f') \in \mathcal{R}^j, t \in T_{f,f',j}^r. \quad (12a)$$

$$w_{j,t}^f \leq w_{j,t}^{f'} + 1 - s_{f,f',j}, \quad \forall j \in \mathcal{J} \cup \mathcal{K}, (f, f') \in \mathcal{R}^j, t \in T_{f,f',j}^r. \quad (12b)$$

If there is a reversal between flights f and f' in resource j , i.e., $s_{f,f',j} = 1$, then constraint (12a) becomes redundant, and constraint (12b) stipulates that if flight f has arrived by time t , then flight f' has to arrive by that time, hence ensuring that flight f cannot arrive before flight f' . Similarly, if there is no reversal, i.e., $s_{f,f',j} = 0$, then constraint (12b) becomes redundant, and constraint (12a) stipulates that if flight f' has arrived by time t , then flight f has to arrive by that time, hence ensuring that flight f' cannot arrive before flight f . Thus, we are able to model a reversal with the addition of only one variable ($s_{f,f',j}$).

Given this additional set of constraints, the model then minimizes a weighted combination of total delay costs and total number of reversals. The parameters λ_s^r and λ_a^r are chosen appropriately to control the degree of fairness in sector reversals and airport reversals, respectively.

The TFMP model with the additional control on reversals is as follows:

$$\begin{aligned} \min \quad & \left\{ \sum_{f \in \mathcal{F}} \left(\sum_{t \in T_{\text{dest}_f}^f} c_{\text{total}}^f(t) \cdot (w_{\text{dest}_f,t}^f - w_{\text{dest}_f,t-1}^f) \right. \right. \\ & \left. \left. - \sum_{t \in T_{\text{orig}_f}^f} c_g^f(t) \cdot (w_{\text{orig}_f,t}^f - w_{\text{orig}_f,t-1}^f) \right) \right\} \\ & + \lambda_s^r \cdot \left(\sum_{j \in \mathcal{J}, (f,f') \in \mathcal{R}^j} s_{f,f',j} \right) \\ & + \lambda_a^r \cdot \left(\sum_{k \in \mathcal{K}, (f,f') \in \mathcal{R}^k} s_{f,f',k} \right) \end{aligned}$$

subject to:

(1a)–(1f).

(12a)–(12b).

$$s_{f,f',j} \in \{0, 1\}, \quad \forall j \in \mathcal{J} \cup \mathcal{K}, (f, f') \in \mathcal{R}^j.$$

For each element $(f, f') \in \mathcal{R}^j$, let $IP_R(f, f', j)$ denote the set of all feasible binary vectors satisfying constraints (12a) and (12b). We show in the appendix that the polyhedron induced by constraints (12a) and (12b) is the convex hull of solutions in $IP_R(f, f', j)$.

$$IP_R(f, f', j) = \{w_{j,t}^f \in \{0, 1\}, s_{f,f',j} \in \{0, 1\} \mid$$

$$\begin{aligned} w_{j,t}^{f'} - w_{j,t}^f - s_{f,f',j} &\leq 0, \quad t \in T_{f,f',j}^r, \\ w_{j,t}^f - w_{j,t}^{f'} + s_{f,f',j} &\leq 1, \quad t \in T_{f,f',j}^r \}. \end{aligned}$$

REMARK 1 (RBS POLICY—A SPECIAL CASE OF TFMP-REVERSAL). When there is sufficient capacity at all airports, such that a feasible solution under RBS exists (i.e., there are no reversals), this model is capable of generating that solution (using a sufficiently high λ_a^r) while minimizing the total delay costs. Hence, a solution under RBS policy is a special case of our model. Because a solution under RBS preserves the order of flight arrivals, for every pair of flights $(f, f') \in \mathcal{R}^A$, the variable $s_{f,f',\text{dest}_f} = 0$, and constraints (12a) and (12b) reduce to constraint (13), which ensures that flight f' cannot arrive before flight f

$$w_{\text{dest}_{f'},t}^{f'} \leq w_{\text{dest}_f,t}^f, \quad \forall (f, f') \in \mathcal{R}^A, t \in T_{f,f',\text{dest}_f}^r. \quad (13)$$

REMARK 2 (DISCUSSION OF AIRPORT AND SECTOR REVERSALS). Currently RBS is the principle used to maintain the sequencing of flights at capacitated elements of the National Airport System (NAS) (both for GDPs and AFPs) and indirectly allocate delays to airlines. By optimizing over reversals, we control the degree of disruptions to such sequences when capacity decreases. This allows us to decrease the overall delay (relative to RBS) at the expense of increasing reversals. That is, we allow more flexibility on the trade-off between efficiency (i.e., delays) and fairness (i.e., reversals). This trade-off is controlled by parameters λ_a^r and λ_s^r in the objective function that control the airport and sector reversals, respectively. In §6.6, we provide empirical evidence on the trade-off between fairness and efficiency.

Size of the Formulations. Denoting with

$$N = \max_{f \in \mathcal{F}} |\mathcal{S}^f|,$$

the total number of decision variables and constraints for the various models can be bounded as listed in Table 2.

To get a feeling of the size of the formulations, let us consider an example that adequately represents the U.S. network.

Table 2 Upper Bound on the Size of the Models

Model	No. of decision variables	No. of constraints
TFMP	$ \mathcal{F} MN$	$2 \mathcal{K} \mathcal{T} + \mathcal{S} \mathcal{T} + 2 \mathcal{F} MN + 2 \mathcal{F} N + M \mathcal{C} $
TFMP-Reversal	$ \mathcal{F} MN + \mathcal{R}^A + \mathcal{R}^S $	$2 \mathcal{K} \mathcal{T} + \mathcal{S} \mathcal{T} + 2 \mathcal{F} MN + 2 \mathcal{F} N + M \mathcal{C} + 2(\mathcal{R}^A + \mathcal{R}^S)M$
TFMP-Overtake	$ \mathcal{F} MN + M \mathcal{R}^A + M \mathcal{R}^S $	$2 \mathcal{K} \mathcal{T} + \mathcal{S} \mathcal{T} + 2 \mathcal{F} MN + 2 \mathcal{F} N + M \mathcal{C} + (\mathcal{R}^A + \mathcal{R}^S)M$

Table 3 Numerical Example: Upper Bound on the Size of the Models

Model	No. of decision variables	No. of constraints
TFMP	300,000	768,000
TFMP-Reversal	350,000	1,368,000
TFMP-Overtake	600,000	1,068,000

EXAMPLE 2. Let $|\mathcal{K}| = 50$, $|\mathcal{T}| = 100$, $|\mathcal{S}| = 100$, $|\mathcal{R}^A| = 50,000$, $|\mathcal{R}^S| = 0$, $|\mathcal{F}| = 10,000$, $|\mathcal{C}| = 8,000$, $M = 6$, and $N = 5$. The duration of each period is 15 minutes, implying a planning horizon of almost a day. For this example, the upper bound on the number of variables and constraints are listed in Table 3.

Because we introduce only one class of variables $s_{f,f',j}$ for all elements $(f, f') \in \mathcal{R}^i$, the number of variables in the model (TFMP-Reversal) are comparable to the original model (TFMP).

Alternative Notion of Fairness. An alternative model of fairness is to equalize the resulting reversals and overtaking among airlines taking into account the number of flights they operate. Let d_w denote the number of reversals per flight for airline w and η denote the mean of the d_w 's across all airlines. Then, this can be achieved as follows:

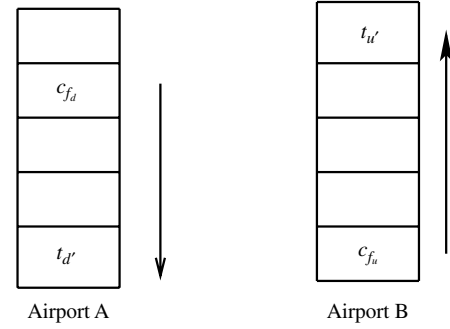
$$d_w = \left(\sum_{j \in \mathcal{S} \cup \mathcal{K}} \sum_{(f, f') \in \mathcal{R}^j, f, f' \in \mathcal{F}_w} s_{f,f',j} \right) / |\mathcal{F}_w|,$$

$$\eta = \left(\sum_{w \in \mathcal{W}} d_w \right) / |\mathcal{W}|.$$

Then, we add $|d_w - \eta|$ term to the objective function of minimizing the total delay cost with an appropriate trade-off parameter.

4. Network Model for Slot Reallocation That Incorporates Airline Collaboration

The Set \mathcal{O} (Set of Airline Offers). We next give the definition of \mathcal{O} (set of airline offers). We use a structure proposed by Vossen and Ball (2006a) and also used recently by Sherali et al. (2011) that allows the airlines to submit AMAL offers. An AMAL offer is of the form $(f_d, t_d; f_u, t_u)$, which means that the airline is willing to move flight f_d to a later time period, but no later than t_d , in return for moving flight f_u to an earlier time period, but no later than t_u . The destination airports of flights f_d and f_u are allowed to be distinct. An AMAL offer is attractive because (a) simplicity in data input (only a four-tuple needs to be specified) and (b) the flexibility of an individual offer (several slots of delay for a flight are acceptable in lieu of a reduction in the delay of another flight), which allows multiple viable combinations of

**Figure 3** Illustration of an AMAL Offer: $(f_d, t_d; f_u, t_u)$

Notes. $\text{dest}_{f_d} = A$, $\text{dest}_{f_u} = B$, $t_d' = c_{f_d} + 3$, and $t_u' = c_{f_u} - 4$. The offer states that the airline is willing to delay flight f_d by at most three slots if in return flight f_u is moved earlier by at least four slots.

executable trades. Figure 3 gives an example to illustrate the semantics of the offer structure. The set \mathcal{O} contains all AMAL offers submitted by the airlines after a schedule is generated from Stage 1 of our approach. Note that c_{f_d} and c_{f_u} denote the slots allotted to the two flights from Stage 1, and hence, for such an offer to be useful, we must have $c_{f_d} < t_d'$ and $c_{f_u} > t_u'$. Finally, $\mathcal{O}^f \subseteq \mathcal{O}$ defines the set of offers containing flight f .

4.1. A Model for Slot Reallocation in a Network Setting (TFMP-Trading)

Here, we introduce our model of slot reallocation in a network setting. The model only introduces one additional variable per offer in addition to $w_{j,t}^f$, which are the variables used in the TFMP model of Bertsimas and Stock-Patterson (1998). The key difference from the models proposed in Vossen and Ball (2006a) and Sherali et al. (2011) is that our model considers a macroscopic setting (multiple airports and sectors), whereas the latter operate in a more microscopic context. The key commonality is the notion of AMAL in addition to the idea of slot reallocation.

The Decision Variables.

- $o_{dd'uu'} \in \{0, 1\} = 1$ if offer $(f_d, t_d; f_u, t_u)$ is executed.

Constraints.

$$(1a) - (1f).$$

$$o_{dd'uu'} \leq w_{\text{dest}_{f_d}, t_d'}^{f_d}, \quad \forall (f_d, t_d; f_u, t_u) \in \mathcal{O}. \quad (14a)$$

$$o_{dd'uu'} \leq w_{\text{dest}_{f_u}, t_u'}^{f_u}, \quad \forall (f_d, t_d; f_u, t_u) \in \mathcal{O}. \quad (14b)$$

$$\sum_{j \in \mathcal{O}^f} o_j \leq 1, \quad \forall f \in \mathcal{F}. \quad (14c)$$

$$w_{\text{dest}_f, c_f}^f - w_{\text{dest}_f, c_f-1}^f \geq 1 - \sum_{j \in \mathcal{O}^f} o_j, \quad \forall f \in \mathcal{F}. \quad (14d)$$

$$w_{j,t}^f \in \{0, 1\}, \quad \forall f \in \mathcal{F}, j \in \mathcal{S}^f, t \in T_j^f.$$

$$o_{dd'uu'} \in \{0, 1\}, \quad \forall (f_d, t_d; f_u, t_u) \in \mathcal{O}.$$

Constraints (14a) and (14b) enforce that when an offer $o_{dd'uu'}$ is executed (i.e., $o_{dd'uu'} = 1$), then $w_{dest_{f_d}, t_{d'}}^{f_d} = 1$ and $w_{dest_{f_u}, t_{u'}}^{f_u} = 1$ —i.e., flights f_d and f_u cannot arrive after the respective time periods in the offer, namely, $t_{d'}$ and $t_{u'}$. This ensures that the semantics of the structure of an offer are satisfied. Constraint (14c) enforces that, for each flight, at most one offer will be executed. Moreover, constraint (14d) stipulates that if no offer for a flight f is executed (i.e., $o_j = 0, \forall j \in \mathcal{O}^f$), then the flight will arrive at the time period allotted from Stage 1 (c_f).

Objective Function. In the model presented above, we have not explicitly stated the objective function that should be used. It is evident that fairness in the number of executed offers across airlines would again be relevant in this stage of our approach.

Let n_w denote the number of trades executed corresponding to airline w , and let γ denote the mean of the trades executed across all airlines.

$$n_w = \sum_{f \in \mathcal{F}_w, j \in \mathcal{O}^f} o_j,$$

$$\gamma = \left(\sum_{w \in \mathcal{W}} n_w \right) / |\mathcal{W}|.$$

In §5, we report computational results based on the following two objective functions:

- **Objective 1:** maximize the total number of trades ($\max \sum_{(f_d, t_{d'}; f_u, t_{u'}) \in \mathcal{O}} o_{dd'uu'}$).
- **Objective 2:** minimize the difference in the number of trades executed for each airline from the mean ($\min \sum_{w \in \mathcal{W}} |n_w - \gamma|$).

5. Comparison with Current CDM Practice

In this section, we present a concrete example to highlight the utility of our overall approach.

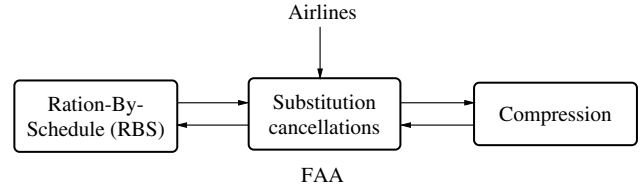


Figure 4 Schematic of a GDP Under CDM

Example of GDP Operations. To concretely illustrate the three-phase CDM paradigm described in §1, we give a detailed example of a GDP implementation. Figure 4 shows the schematic of the various phases involved in GDP planning.

Figure 5 depicts an example illustrating the current operational details of a GDP. There are three airlines A, B, and C operating seven flights between them. Airline A has three flights (A1, A2, and A3), Airline B has two flights (B1 and B2) and Airline C has two flights (C1 and C2) in the GDP. The original published ordering is shown in the leftmost table. The next table shows the output of applying the RBS principle to the original sequence. Assume that the capacity is reduced by half, thereby leading to an airport acceptance rate (AAR) of one flight every two slots (compared with a flight every slot under nominal conditions). As a result, each flight is assigned a controlled time of arrival (CTA), which is twice the slot allotted in the original sequence (the exact control is exerted through an expected departure clearance time (EDCT) to meter the demand at the destination airport). The RBS sequence is followed by the substitution-cancellation phase, wherein the airlines are given the flexibility of changing the order of the flights assigned to them as well as cancelling flights as long as the resulting schedule remains a capacity that is feasible. In this example, Airline C substitutes

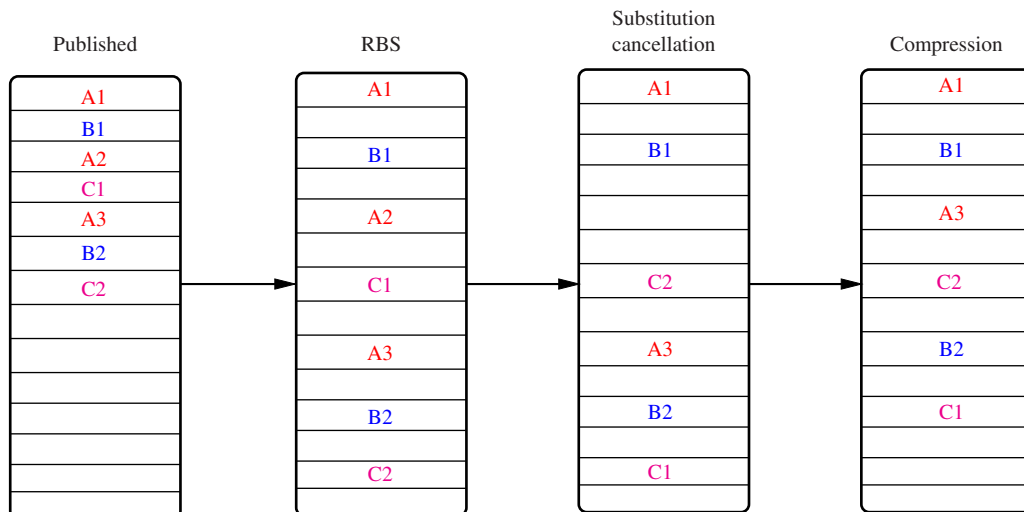


Figure 5 (Color online) An Example Demonstrating the Current Operational Details of a GDP

flight C1 by C2 and Airline A cancels flight A2. The final step in this sequence is the application of the compression procedure, wherein flights are moved up to fill up the holes created by cancelled flights (as long as the final allotment is compatible with the earliest arrival times of all flights). Note how compression moves up the canceling airline's flight (flight A3 for Airline A) before moving up C2. This is an important feature that gives incentives to airlines to report cancellations.

5.1. Comparison: Example Illustration

To demonstrate the utility of our overall approach, consider the scenario depicted in Figure 6. Time is represented on the horizontal axis. For each flight, the arrowhead points to the new time of arrival under reduced capacity. There are two airports and two airlines. Airline A operates seven flights (A1 to A7) and Airline B operates six flights (B1 to B6). Flight A1 (arriving at Airport 1) is continued by flight A5 (arriving at Airport 2). The turnaround time is one unit followed by two units of flight time for the subsequent

flight. Similarly, flight B2 (arriving at Airport 1) is continued by flight B5 (arriving at Airport 2).

The capacity at Airport 1 gets reduced to one flight arriving every three slots as opposed to one flight every slot (under nominal conditions), whereas the capacity at Airport 2 gets reduced to two flights arriving every three slots. Let us first consider the utility of a schedule that controls reversals compared with the RBS schedule. Because of the coupling between A5 and A1, flight A5 is unable to utilize the slot allotted to it under the RBS paradigm because it is not compatible with the earliest slot it can have given the slot assigned to A1. A similar situation holds for flight B5. Based on the internal airline objectives, there are multiple allocations possible after the end of the RBS pass based on how airlines resolve the infeasibility over aircraft connectivities. In a scenario where both airlines do not want to cancel flights, one possible allocation after the RBS pass is as shown in Figure 6, which preserves aircraft connectivity. In this RBS schedule, there is available capacity that is not utilized (this unutilized capacity is denoted by *). By

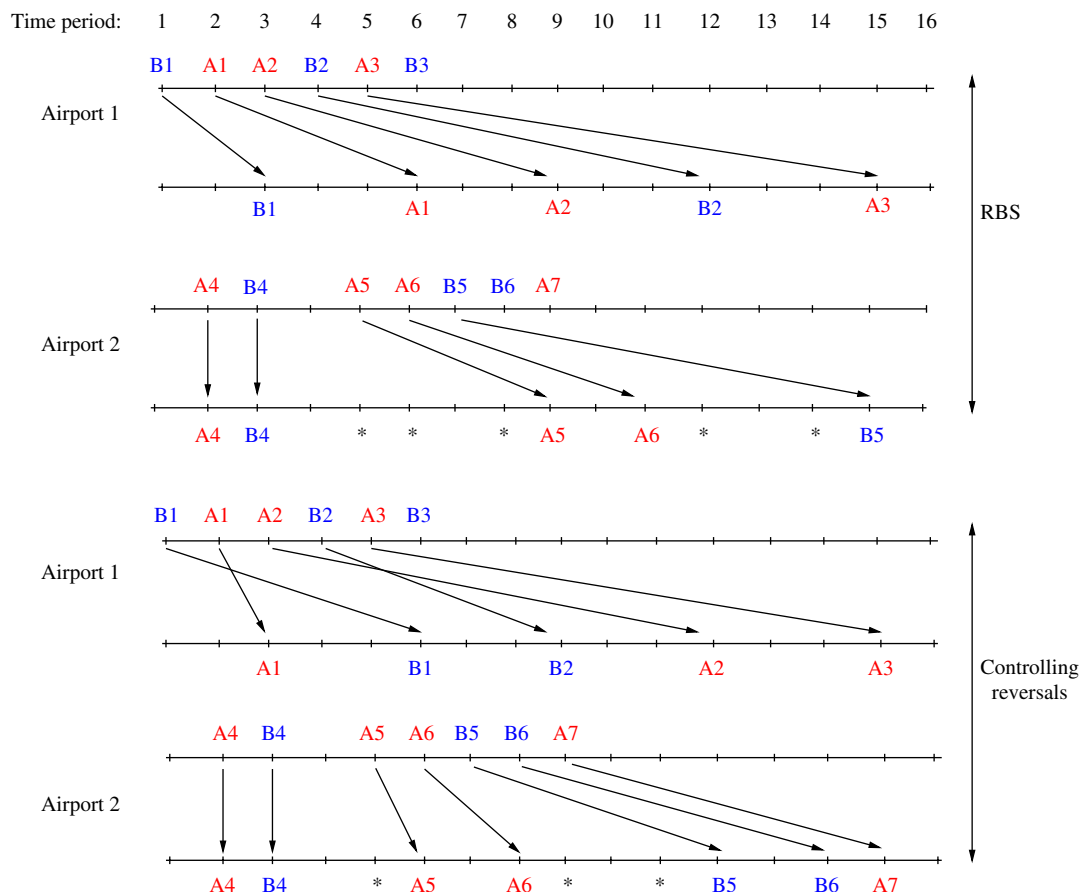


Figure 6 (Color online) Example Demonstrating the Utility of Controlling Reversals Over the RBS Solution

Note. The capacity of Airport 1 gets reduced by two-thirds, whereas that of Airport 2 by one-third.

*Denotes the capacity that remains unutilized despite being available.

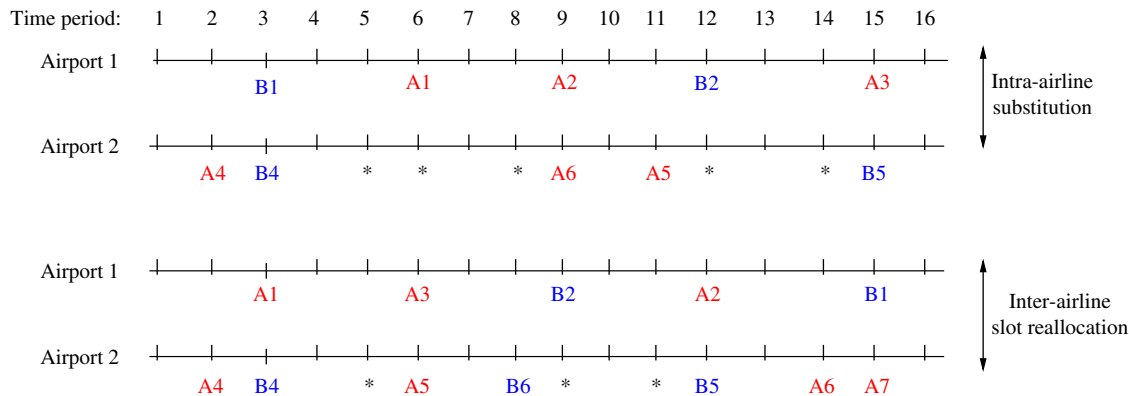


Figure 7 (Color online) Example Demonstrating the Utility of Our Slot Reallocation Phase Over Intra-Airline Substitution

*Denotes the capacity that remains unutilized despite being available.

allowing reversals, the unused capacity (in the RBS schedule) can be utilized, thereby leading to a more efficient schedule. For instance, in this example, by reversing flights B1 and A1 at Airport 1, flight A5 arrives earlier during the sixth slot as compared with ninth in the RBS allotment. Using a similar argument, by reversing flights B2 and A2, flight B5 can utilize an earlier unused slot of the RBS schedule. Furthermore, by allowing reversals, the final schedule can overcome cancellations that might occur in the RBS schedule (for instance, flights B6 and A7 get cancelled in the RBS schedule).

Given the allocation from the fairness stage, we now contrast the utility of the interairline slot reallocation phase of our approach compared with the intra-airline substitution phase in current CDM practice. Figure 7 depicts the resulting schedule. In the example we consider, each airline is concerned with maximizing the number of flights arriving on time (defined as arrival within one time period) at its hub airport. Airport 1 is the hub of A, and Airport 2 is the hub of B. If Airline A submits an offer (A6, 6; A3, 9) and Airline B submits another offer (B1, 9; B6, 6), then a trade is executed. The key observation is that by allowing inter-airline substitution (with airline offers across multiple airports), we are able to have two more flights (B6 and A3) arrive on time. In contrast, with the intra-airline substitution phase, it is not possible to have flights A3 and B6 arrive on time.

Table 4 summarizes the advantages of our approach relative to RBS⁺⁺ on the performance metrics of total

delay, capacity (under) utilization, reversals, and on-time performance. Allowing reversals leads to the dual benefits of reduced delays and better utilization of the available capacity relative to the RBS schedule. Finally, by allowing interairline (across airports) flight substitutions, we have two more flights arriving on time, which is not possible in the intra-airline substitution phase of current CDM practice.

6. Computational Results

In this section, we report computational results from the optimization models introduced in §§3 and 4 on national-scale, real-world data sets spanning six days. The data set for each day encompasses 55 major airports in the United States and covers operations of the top five airlines. Each data set contains data on the actual flight arrival and departure times for that particular day, which lets us compute the actual delays.

6.1. Statistics of the Data Sets

Table 5 summarizes the statistics of six days of flight data. These six days correspond to a set of randomly selected summer days when the delays were high. The selected days were July 14, 2004; August 4, 2004; May 13, 2005; July 16, 2005; July 27, 2005; and July 27, 2006. Hereafter, we refer to these days simply by {1, ..., 6}. These correspond to the operations

Table 4 Utility of Our Overall Approach Compared with RBS⁺⁺

Metric	Our approach	RBS ⁺⁺
Total delay (no. of slots)	60	72
Slots unused	3	5
Reversals	2	0
On-time performance (no. of flights)	2	0

Table 5 Summary of the Data Sets

Day	No. of flights	\mathcal{CF}	Total delay (units of 15 min.)	Reversals	Overtaking
1	5,092	2,691	4,438	9,944	30,652
2	5,844	3,298	4,926	7,756	19,003
3	5,780	3,310	3,079	4,797	12,231
4	4,590	2,301	3,907	6,488	18,825
5	5,128	2,728	3,326	6,399	19,998
6	4,781	2,504	3,101	4,351	12,954

Note. \mathcal{CF} denotes the number of connecting flights.

at the 55 major airports of the United States. We filter in the flights corresponding to the operations of the top five airlines (measured by the number of flight operations)—Southwest (SWA), American (AAL), Delta (DAL), United (UAL), and Northwest (NWA)—to enable us to better analyze the results.

6.2. Experimental Setup

In our experimental setup, the airspace is subdivided into sectors of equal dimensions (10 by 10) that form a grid, thereby forming a total of 100 sectors. The 55 major airports of the United States are then mapped to one of these sectors based on its geographical coordinates. For each flight, we fix its flight trajectory (i.e., the sequence of sectors in its path) based on a randomized shortest path from the origin to the destination airport. Using the information on its flight time, we compute the minimum amount of time that each flight has to spend in a sector. This value is then used to calculate the set of feasible times that a flight can be in a sector. By tracking the tail number of an aircraft, we form the set of connecting flights. We believe this setup represents a reasonably approximate testbed of the actual NAS operations to study the implications of incorporating various fairness criteria and the resulting impacts on efficiency.

For a sample day, we know the scheduled departure and arrival times of a flight as well as what actually happened on that day. We use this to compute its ground and air-hold delays. Furthermore, we compute the base capacities at all of the resources (airports and sectors) by noting the actual departure and arrival times of the flights. It is important to note that this capacity corresponds to the exact number of flight departures and arrivals that occurred on that particular day and, hence, is the most conservative estimates of the capacity. The available capacities on that day has to be higher than the actual number of operations. Although data difficulties make it hard to infer the exact capacities available for the concerned day, we use educated augmented values of base capacity as capacity inputs to run the optimization models. Specifically, we increase the capacity by a certain percentage over the number of aircraft operations that occurred for each capacitated resource (we use a slack of 20%). This ensures that the weather impact across the airspace is captured.

A critical parameter of the optimization models is the maximum permissible delay for a flight (M). This value is used to define the set of feasible times that a flight can be in a particular sector. For example, the set of feasible times that a flight f can arrive at its destination airport is given by all values in a_f^k through $(a_f^k + M)$, where $\text{dest}_f = k$. The size of all of the optimization models and, hence the computational times, are directly proportional to the value of M

(e.g., for $M \geq 10$, the number of variables and constraints quickly scale up to the order of millions). The use of this parameter is not motivated from equity considerations and has no consequences on the fairness properties of the resulting schedules. We use a value of $M = 6$, which corresponds to six time periods (each 15 minutes long), hence permitting a maximum delay of 90 minutes.

REMARK 3. Whereas a small value of M will facilitate the computational efficiency of the optimization models, the framework is generic enough to allow for a larger value of M for a select set of flights for which higher delays are expected (or desired). The latter is practically relevant because airlines are willing to tolerate longer delays (on certain flights) instead of cancellations.

To compute optimal solutions, we use CPLEX 11, implemented using AMPL as a modeling language on a laptop with 2 GB RAM and Linux Ubuntu OS. The instances reported in this paper have a typical size on the order of 300,000 variables (this increases significantly for TFMP-Overtake) and 800,000 constraints (this increases significantly for TFMP-Reversal).

6.3. Performance of TFMP

To calibrate the TFMP model, we initially assign capacity the exact number of flight arrivals and departures as occurred on that particular day. In this case, the total absolute delay under optimization is not very different from the actual delays of the day under consideration. We then increase the capacity by 20%. Table 6 reports solutions from the (TFMP) model for the case when the capacity used is 20% over and above the actual number of aircraft operations. There is an average reduction of 23% in the total absolute delays. This illustrates the order of magnitude reduction of delay that one might expect from optimization. Furthermore, the number of reversals consistently range between 500 and 1,000 and amount of overtaking range between 800 and 1,500 across all days. This confirms that, although there can be significant benefits in the total delay costs by using the model (TFMP), the number of reversals and overtaking might be high.

Table 6 Performance of TFMP

Day	No. of flights	Actual delay (units of 15 min.)	TFMP		
			Delay (15 min.)	Reversals	Overtaking
1	5,092	4,438	3,385	915	1,492
2	5,844	4,926	3,492	924	1,426
3	5,780	3,079	2,242	753	1,191
4	4,590	3,907	3,053	769	1,235
5	5,128	3,326	2,648	801	1,291
6	4,781	3,101	2,542	526	822

6.4. Performance of TFMP-Overtake

Table 7 reports the computational performance of the TFMP-Overtake model on the six data sets. These results pertain to the parameter λ_a^o set to 100. The number reported under \mathcal{OV} is the total amount of overtaking, whereas \mathcal{RV} reports the number of reversals (without taking into account the relative magnitudes of overtaking within each reversal, i.e., the number of time periods by which a flight overtakes its preceding flight when a reversal occurs). The key observation is the possibility of generating schedules with overtaking in double digits (i.e., less than 100) for all days. This is a considerable improvement compared to the actual amount of overtaking, which in all cases is greater than 10,000 (see Table 5). The degradation in total delay costs from the TFMP-Overtake model over the TFMP solution range between 13% and 41% for fairness at 25 airports, the average being 24.5%. The model on average takes less than 30 minutes to converge to optimality for up to 25 airports. For the three cases when the optimizer stopped because the runtime exceeded 3,600 seconds (runs for

Table 7 Computational Performance of TFMP-Overtake

Day (no. of flights)	\mathcal{AF}	Solution time (sec.)	\mathcal{RV}	\mathcal{OV}	Delay cost (15 min.)	% increase in delay cost over TFMP
1 (5,092)	0	261	915	1,492	3,525	
	5	186	2	4	3,690	4.68
	15	727	13	27	4,243	20.36
	25	3,073	26	39	4,662	32.25
	30	3,600	39	65	4,815	36.59
2 (5,844)	0	108	924	1,426	3,604	
	5	206	1	2	3,802	5.49
	15	596	6	9	4,029	11.79
	25	806	9	12	4,080	13.20
	30	3,530	16	18	4,510	25.13
3 (5,780)	0	311	753	1,191	2,313	
	5	170	3	6	2,401	3.80
	15	397	8	18	2,584	11.71
	25	295	11	22	2,651	14.61
	30	3,394	17	28	3,096	33.85
4 (4,590)	0	51	769	1,235	3,173	
	5	150	1	1	3,628	14.33
	15	746	5	6	4,201	32.39
	25	691	13	18	4,452	40.30
	30	3,600	29	33	4,743	49.47
5 (5,128)	0	178	801	1,291	2,744	
	5	116	0	0	2,871	4.62
	15	492	10	18	3,319	20.95
	25	1,983	17	26	3,505	27.73
	30	3,600	25	36	3,804	38.62
6 (4,781)	0	49	526	822	2,637	
	5	143	5	7	2,826	7.16
	15	378	9	15	3,070	16.42
	25	479	15	22	3,145	19.26
	30	1,305	28	49	3,383	28.28

Notes. \mathcal{AF} denotes the number of airports where fairness is imposed. In particular, $\mathcal{AF} = 0$ corresponds to the TFMP solution.

Day 1, 4, and 5 with $\mathcal{AF} = 30$), the optimality gaps at termination were 4.1%, 2.1%, and 2.5%, respectively.

6.5. Performance of TFMP-Reversal

The TFMP-Reversal model minimizes a weighted combination of total delay costs and total number of reversals, where λ_a^r is the weight parameter. We study the tradeoff inherent in these conflicting objectives in two ways—(a) as a function of the trade-off parameter λ_a^r and (b) as a function of the number of airports where this fairness criterion is imposed.

The effect of the trade-off parameter. Figure 8 plots the trade-off in the number of reversals with the total delay cost as a function of λ_a^r for fairness based on controlling total reversals imposed at 25 airports. The five points on the plot for each day correspond to the result from TFMP-Reversal with $\lambda_a^r = 0, 1, 10, 100$, and 1,000. Initially, there is a significant reduction in the number of reversals at the cost of a small increase in total delay cost, but the subsequent benefits in the number of reversals come at a high cost. For all days, the model is able to achieve less than 100 reversals for a degradation of at most 10% in the total delay cost. To achieve reversals between 10 and 30, the degradation in total delay costs range between 10% and 40% across all days.

The effect of the number of airports. Table 8 reports the computational performance of the TFMP-Reversal model on the six data sets as a function of the number of airports where this fairness criterion is imposed. These results pertain to the trade-off parameter λ_a^r set to 100. As is evident from the results reported across all days, the number of reversals can be held to 10–30. The degradation in total delay costs from the TFMP-Reversal model over the TFMP solution range

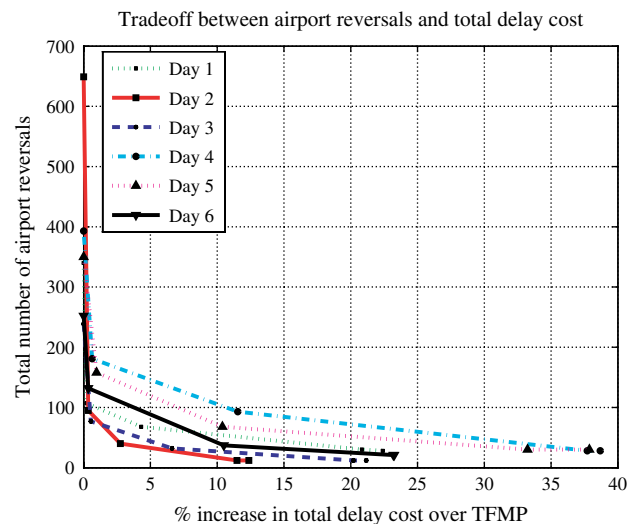


Figure 8 Effect of the Trade-Off Parameter λ_a^r

Note. The five points for each day correspond to the result from TFMP-Reversal with $\lambda_a^r = 0, 1, 10, 100$, and 1,000.

Table 8 Computational Performance of TFMP-Reversal

Day (no. of flights)	\mathcal{AF}	Solution time (sec.)	\mathcal{RV}	\mathcal{OV}	Delay cost (15 min.)	% increase in delay cost over TFMP
1 (5,092)	0	261	915	1,492	3,525	
	5	228	2	4	3,691	4.70
	15	899	13	30	4,246	20.45
	25	3,600	25	53	4,402	24.87
	30	3,600	37	84	4,586	30.09
2 (5,844)	0	108	924	1,426	3,604	
	5	213	1	2	3,785	5.02
	15	837	6	9	4,028	11.76
	25	1,024	9	12	4,077	13.12
	30	3,600	16	30	4,403	22.16
3 (5,780)	0	311	753	1,191	2,313	
	5	219	3	6	2,406	4.02
	15	282	8	19	2,583	11.67
	25	384	11	24	2,648	14.48
	30	3,600	15	35	3,048	31.77
4 (4,590)	0	51	769	1,235	3,173	
	5	165	1	1	3,627	14.30
	15	1,367	5	10	4,138	30.41
	25	2,292	12	22	4,468	40.81
	30	3,600	29	55	4,558	43.64
5 (5,128)	0	178	801	1,291	2,744	
	5	225	0	0	2,866	4.44
	15	654	10	24	3,319	20.95
	25	1,556	16	35	3,504	27.69
	30	3,600	25	38	3,925	43.03
6 (4,781)	0	49	526	822	2,637	
	5	213	5	7	2,836	7.54
	15	751	9	16	3,059	16.00
	25	654	15	25	3,125	18.50
	30	3,600	28	51	3,387	28.44

Notes. \mathcal{AF} denotes the number of airports where fairness is imposed. In particular, $\mathcal{AF} = 0$ corresponds to the TFMP solution.

between 13% and 40% for fairness at 25 airports, the average being 24.5%. The model on average takes less than 30 minutes to converge to optimality for up to 25 airports. As expected, the total amount of reversals reported in Table 8 is always less than the corresponding number in Table 7, whereas the opposite is true for the amount of overtaking.

The computational times of both TFMP-Reversal and TFMP-Overtake are consistently less than 30 minutes for up to 25 airports, but they break down when we impose fairness at 30 airports and above. We believe that this is due to memory limitations. Given the fact that for some instances the computational times are still prohibitively large, we study the times needed to achieve a solution that is within 20% of the optimal. In all cases, we could achieve feasible solutions satisfying this optimality criterion in less than 15 minutes.

6.6. Controlling Sector Reversals and Balancing with Airport Reversals

In this section, we study the interaction of the two objectives of controlling airport reversals and sector

Table 9 Balancing Sector Reversals with Airport Reversals

Day	$\lambda_r^s = 0, \lambda_r^a = 100$			$\lambda_r^s = 1, \lambda_r^a = 100$			$\lambda_r^s = 10, \lambda_r^a = 100$		
	\mathcal{SR}	\mathcal{AR}	\mathcal{DC}	\mathcal{SR}	\mathcal{AR}	\mathcal{DC}	\mathcal{SR}	\mathcal{AR}	\mathcal{DC}
1	3,174	8	3,780	253	11	3,865	127	12	4,261
2	3,455	6	3,860	169	8	3,944	59	8	4,235
3	1,289	9	2,486	58	9	2,493	14	10	2,631
4	3,242	5	3,663	326	6	3,874	136	4	4,292
5	1,646	0	2,897	268	0	2,958	75	1	3,456
6	1,615	10	2,832	144	10	2,907	46	9	3,235

Notes. \mathcal{SR} denotes sector reversals, \mathcal{AR} denotes airport reversals, and \mathcal{DC} denotes the delay cost. Fairness imposed in five sectors of the northeast region and 10 airports spatially close to these sectors.

reversals with the aim to quantify the price of controlling the latter (on airport reversals and total delay cost). Our setup for this exercise comprises of controlling reversals in five sectors of the northeast region of the United States and a set of 10 airports that lie spatially close to these sectors. Table 9 reports the sector reversals, airport reversals, and total delay cost for different combinations of the trade-off parameters λ_r^s and λ_r^a .

The left plot in Figure 9 is a box plot quantifying the percentage increase in the number of airport reversals and the delay cost by enforcing the additional control on sector reversals for the trade-off parameter $\lambda_r^s = 10$. The percentage increase in delay cost lies between 5% and 20% with a mean of around 13%, whereas the percentage increase in airport reversals is around 5%. Moreover, the sector reversals can be reduced from the order of 1,000s to 10s. In contrast, for $\lambda_r^s = 1$, the average increase in the total delay cost is 3% on average, but the sector reversals are still in the 100s. The right plot in Figure 9 depicts the reduction in the number of sector reversals possible by this explicit control (potential reduction from four digit reversals to two digit reversals).

The overall summary regarding control of sector reversals is as follows:

1. We have developed the TFMP-Reversal model, which is capable of controlling sector reversals in conjunction with airport reversals.
2. The computational experiments show that the impact of controlling sector reversals on total delay cost is substantial (13% on average) for large λ_r^s ($= 10$) and is relatively small (3% on average) for small λ_r^s ($= 1$).

6.7. Interaction with Super-Linear Cost Coefficients

We used super-linear cost coefficients in the overall objective function as an additional means to impose equity. As explained earlier in §2, this eliminates flights with extreme delays. Since our primary fairness notion is controlling reversals and overtaking,

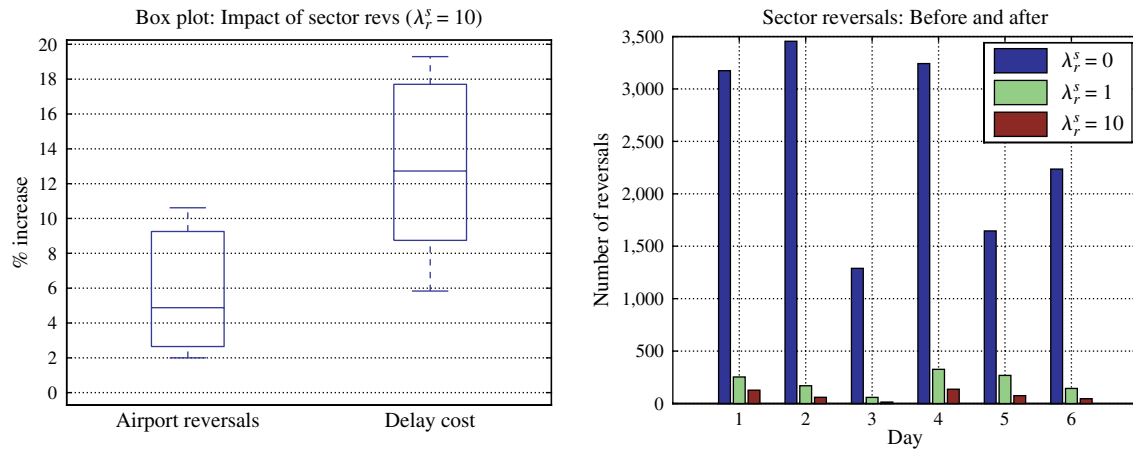


Figure 9 Impact of Controlling Sector Reversals on (i) Airport Reversals and Total Delay Cost (Left) and (ii) Sector Reversals (Right)

we study the interaction of super-linear cost coefficients with this fairness criteria.

The left plot in Figure 10 is a box plot quantifying the percentage increase in the number of airport reversals and the total delay by using super-linear coefficients over the solution obtained using linear coefficients (i.e., $\epsilon = 0$). The percentage increase in reversals consistently lies between -2% and 2% , with a mean of around 0.3% . Furthermore, the total delays never differ by more than 1% , with a mean difference of 0.1% . This highlights that using super-linear coefficients causes insignificant changes to the fairness and delay characteristics of the resulting schedules. The right side of Figure 10 is a table depicting the distribution of delays across flights. As is evident, the use of super-linear coefficients induces a more moderate assignment of delays in contrast to linear coefficients that lead to more flights with either 0 (minimum) or 6 (maximum) units of delay.

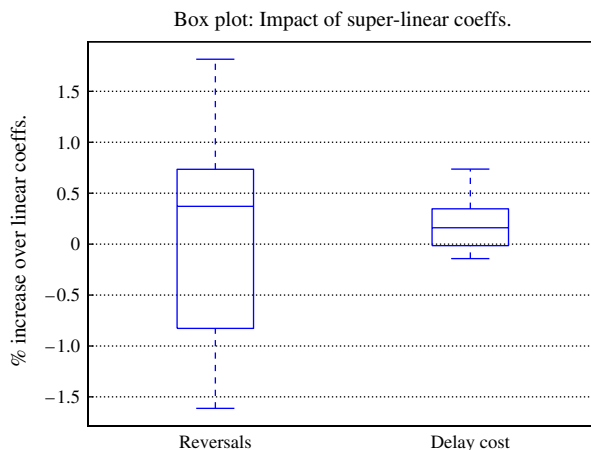
In summary, the use of super-linear cost coefficients achieves its objective of reducing flights with extreme

delays while not causing any material changes to the number of reversals and total delay cost.

6.8. Performance of Slot Reallocation Model (TFMP-Trading)

Given the assignment of flights to various time periods from Stage 1, we generate offers to trade for each airline which maximizes its on-time performance. In other words, each airline tries to maximize the number of flights with delay less than one time unit (15 minutes). We give an example to elaborate on this.

EXAMPLE 3. Suppose two flights f_1 and f_2 (belonging to the same airline) have been assigned two time units of delay each from Stage 1 optimization. Moreover, let t_1 and t_2 be the time periods assigned to the two flights, respectively. Then, the owner airline generates an offer to trade that says that it is willing to delay flight f_1 further by three time units if in return flight f_2 can arrive within one time unit of delay—i.e., it generates the offer $(f_1, t_1 + 3; f_2, t_2 - 1)$. Thus, in



Delay (units of 15 min)	Linear coefficients	Super-linear coefficients
0	4,288	4,280
1	398	402
2, 3, 4	340	346
5	51	52
6	15	12

Figure 10 (Color online) Impact of Super-Linear Cost Coefficients on (i) Reversals and Total Delay Cost (Left); and (ii) Distribution of Flight Delays (Right)

Table 10 Computational Performance of TFMP-Trading

Day	Objective 1			Objective 2		
	Objective cost	Offers executed	Sol. time (sec.)	Objective cost	Offers executed	Sol. time (sec.)
1	283	283	13	0	190	29
2	256	256	14	0	180	31
3	150	150	5	0	140	19
4	278	278	11	0	215	27
5	194	194	9	0	140	22
6	153	153	6	0	90	16

case this trade is executed and flight f_2 will arrive on time, thereby improving the internal objective function of the airline (which is to maximize the on-time performance).

Table 10 reports the results from our network slot reallocation model (TFMP-Trading). As is evident, the computational times of the model TFMP-Trading is consistently less than a minute, which is encouraging for practical deployment. Figure 11 plots the distribution of the number of executed trades across airlines from TFMP-Trading for the two objectives. Not surprisingly, when fairness is not imposed (Objective 1), the distribution is skewed (e.g., for Day 1, Airline 5 gets almost three times more trades executed than Airline 1). In contrast, when fairness is explicitly incorporated (Objective 2), all airlines get exactly the same number of trades executed.

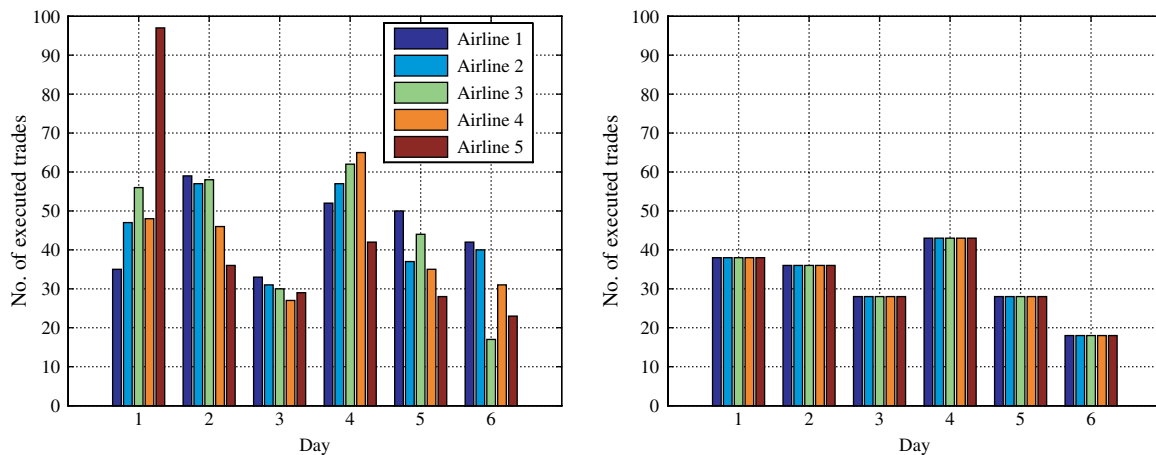
Comparison between single-airport and network-wide settings. In this section, we contrast the performance of TFMP-Trading between single-airport and network-wide settings. It is evident that in the network version, there is a trade-off between the flexibility of trading slots at different airports versus the added constraint of satisfying all network connectivities. In contrast, in a single-airport setting, there is

the advantage of not having to satisfy the network connectivity requirements at the expense of losing on trades across different airports. Thus, to compare the two settings, we divide the set of generated offers between local and network offers as defined below.

DEFINITION 6. An offer $(f_d, t_d; f_u, t_u) \in \mathcal{O}$ is

- a *local offer* if both flights involved have the same destination airport, i.e., $\text{dest}_{f_d} = \text{dest}_{f_u}$.
- a *network offer* if the destination airports of the two flights are distinct, i.e., $\text{dest}_{f_d} \neq \text{dest}_{f_u}$.

Hence, the single-airport model will not have any executed trades that correspond to the network offers. Our overall goal is to demonstrate the utility of TFMP-Trading in executing trades across offers containing distinct airports. Table 11 reports the results from the two versions under the two objectives described in §4.1. The number reported under TFMP-Trading is the number of offers executed from the network model proposed in this paper, whereas SA-Trading (SA stands for single-airport) reports the results obtained from TFMP-Trading after removing all of the network satisfiability constraints and only taking into account the local offers. The numbers reported highlight the trade-offs inherent in ignoring the network effects vis-à-vis trading slots at different airports. As the percentage of local offers increases, SA-Trading outperforms TFMP-Trading, emphasizing that network connectivities are indeed relevant, whereas for a higher fraction of network offers, TFMP-Trading performs better reinforcing the utility of network offers. Finally, the number of trades executed with Objective 2 (when fairness is included) are less than Objective 1, which is along expected lines and demonstrates the flexibility of choosing alternative objective functions.

**Figure 11** Distribution of Number of Executed Trades Across Airlines from TFMP-Trading

Note. Left: Objective 1; Right: Objective 2.

Table 11 Comparison of TFMP-Trading Between Single-Airport and Network-Wide Settings

% local offers	% network offers	No. of offers executed (Obj. 1)		No. of offers executed (Obj. 2)	
		SA-Trading	TFMP-Trading	SA-Trading	TFMP-Trading
0	100	0	281	0	190
25	75	152	255	130	90
50	50	221	241	175	125
75	25	269	238	195	120
100	0	308	258	240	175

6.9. Summary of Computations

To summarize, TFMP (the optimization model without fairness) is able to reduce the total delays by 23% on average when we take reasonable estimates on the available capacity (increasing it by 20% from the actual number of flight operations). In Stage 1 of our approach, we obtain solutions that are able to control the total reversals and overtaking at airports to less than 100 (from the models TFMP-Reversal and TFMP-Overtake, respectively). In addition, we report encouraging computational times of less than 30 minutes for up to 25 airports from both models. Furthermore, we also demonstrate the ability to control sector reversals. In Stage 2 of our approach, there are 220 trades executed on average when the objective function used is to maximize the number of trades and 160 trades when we impose fairness. This reinforces the utility of the slot reallocation phase of our approach because the airlines are able to increase the number of flights arriving on time.

7. Conclusions

In this paper, we develop an optimization-based approach for NAS-wide ATFM that explicitly incorporates fairness and airline collaboration. Specifically, in Stage 1 of our approach, we introduce models for the ATFM problem that incorporates fairness by controlling reversals and overtaking, which is a generalization of the RBS principle to a network setting. Given a schedule from this stage, in Stage 2, we propose an optimization model for slot reallocation to mimic the current substitution-cancellation process in a GDP setting. Furthermore, we report extensive empirical results of the proposed models on national-scale, real-world data sets to quantify the price of fairness and benefits of collaboration. We feel the key advantages of our approach are high quality of solutions, consideration of network effects, and promising computational times.

Acknowledgments

We would like to thank the reviewers and the associate editor for their comments that have improved the paper.

Appendix. Strength of TFMP-Reversal

Let us denote the polyhedron induced by the additional set of constraints to model a reversal for each element $(f, f') \in \mathcal{R}^j$ as $P_R(f, f', j)$.

THEOREM 3. *The polyhedron $P_R(f, f', j)$ is integral.*

PROOF. $P_R(f, f', j)$ can be written as follows:

$$P_R(f, f', j) = \{x = (w_{j,t}^f, s_{f,f',j}) \mid 0 \leq w_{j,t}^f \leq 1, 0 \leq s_{f,f',j} \leq 1, \\ w_{j,t}^{f'} - w_{j,t}^f - s_{f,f',j} \leq 0, t \in T_{f,f',j}^r, \\ w_{j,t}^f - w_{j,t}^{f'} + s_{f,f',j} \leq 1, t \in T_{f,f',j}^r\}.$$

We make use of the following two facts from discrete optimization (Bertsimas and Weismantel (2005)):

Fact 1. Let \mathbf{A} be an integral matrix. \mathbf{A} is totally unimodular if and only if $\{\mathbf{x} \mid \mathbf{a} \leq \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{1} \leq \mathbf{x} \leq \mathbf{u}\}$ is integral, for all integral vectors $\mathbf{a}, \mathbf{b}, \mathbf{1}, \mathbf{u}$.

Fact 2. A matrix \mathbf{A} is totally unimodular if and only if each collection Q of rows of \mathbf{A} can be partitioned into two parts so that the sum of the rows in one part minus the sum of the rows in the other part is a vector with entries only 0, +1, and -1.

Consider the following polyhedron P , and let \mathbf{A} be the matrix such that $P = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$:

$$P = \{x = (w_{j,t}^f, s_{f,f',j}) \mid w_{j,t}^{f'} - w_{j,t}^f - s_{f,f',j} \leq 0, t \in T_{f,f',j}^r, \\ w_{j,t}^f - w_{j,t}^{f'} + s_{f,f',j} \leq 1, t \in T_{f,f',j}^r\}.$$

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & \cdots & \cdots & 0 & -1 \\ -1 & 1 & 0 & 0 & 0 & \cdots & \cdots & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 & \cdots & \cdots & 0 & -1 \\ 0 & 0 & -1 & 1 & 0 & \cdots & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & -1 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & 1 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & -1 & 1 & 1 \end{bmatrix}.$$

The matrix \mathbf{A} has a special structure. If we remove the last column, the remaining matrix is a network matrix.

Let B_1, B_2, \dots, B_n be consecutive blocks of two rows each, i.e., block B_k contains the rows $2k-1$ and $2k$. For any collection Q of rows of the matrix \mathbf{A} , we show how to partition it into two parts J_1 and J_2 so that the sum of the rows in J_1 minus the sum of the rows in J_2 is a vector with entries 0, +1, and -1 only. Suppose Q contains both the rows of some block B_i ; then, put both these rows in J_1 . The remaining

rows (say m) in Q then come from different blocks, call them $R_{j_1}, R_{j_2}, \dots, R_{j_m}$. These m rows are partitioned as follows:

Let Q_+ be the subset of these m rows where the last element is $+1$ and Q_- is the rows where the last element is -1 . Then, put $\lceil |Q_+|/2 \rceil$ rows of Q_+ in J_1 and the remaining rows in J_2 . Similarly, put $\lceil |Q_-|/2 \rceil$ rows of Q_- in J_1 and the remaining rows in J_2 . Because the sum of two rows in the same block is all zeroes, all such blocks in J_1 do not affect the sum of all of the rows in J_1 . Let T denote the vector resulting from the sum of the rows in J_1 minus the sum of the rows in J_2 . All of the elements except the last one in T are exactly 0, $+1$, or -1 because of the structure of the matrix A . The contribution of the rows from Q_+ to the last element of T is either 0 or $+1$. Similarly, the contribution of the rows from Q_- to the last element of T is either -1 or 0. This implies that the last element of T that is the sum of these two contributions can either be $+1$, -1 , or 0.

This shows that the matrix A is totally unimodular. Using Fact 1, we conclude that $P_R(f, f', j)$ is integral. \square

References

- Airlines for America (2012) Cost of delays. <http://www.airlines.org/>.
- Ball M, Chen CY, Hoffman R, Vossen T (2001) Collaborative decision making in air traffic management: Current and future research directions. Bianco L, Dell'Olmo P, Odoni AR, eds. *New Concepts and Methods in Air Traffic Management* (Springer, Berlin), 17–30.
- Barnhart C, Bertsimas D, Caramanis C, Fearing D (2012) Equitable and efficient coordination of traffic flow management programs. *Transportation Sci.* 46(2):262–280.
- Bertsimas D, Odoni AR (1997) A critical survey of optimization models for tactical and strategic aspects of air traffic flow management. Technical report, Massachusetts Institute of Technology, Cambridge, MA.
- Bertsimas D, Stock-Patterson S (1998) The air traffic management problem with enroute capacities. *Oper. Res.* 46(3):406–422.
- Bertsimas D, Stock-Patterson S (2000) The traffic flow management rerouting problem in air traffic control: A dynamic network flow approach. *Transportation Sci.* 34(3):239–255.
- Bertsimas D, Weismantel R (2005) *Optimization Over Integers* (Dynamic Ideas, Belmont, MA).
- Bertsimas D, Lulli G, Odoni AR (2011) The air traffic flow management problem: An integer optimization approach. *Oper. Res.* 59(1):211–227.
- Churchill A, Lovell D, Ball M (2009) Evaluating a new formulation for large-scale traffic flow management. Eighth USA/Europe Air Traffic Management Res. Development Seminar, ATM2009.
- Helme M (1992) Reducing air traffic delay in a space-time network. *IEEE Internat. Conf. Systems, Man, Cybernetics* 1:236–242.
- Hoffman R, Mukherjee A, Vossen T (2012) Air traffic flow management. *Quantitative Problem Solving Methods in the Airline Industry*, Vol. 169 (Springer, New York), 385–453.
- Lindsay K, Boyd E, Burlingame R (1993) Traffic flow management modeling with the time assignment model. *Air Traffic Control Quart.* 1(3):255–276.
- Odoni AR (1987) The flow management problem in air traffic control. *Flow Control Congested Networks* 38:269–288.
- Richetta O, Odoni AR (1993) Solving optimally the static ground-holding policy problem in air traffic control. *Transportation Sci.* 27(3):228–238.
- Richetta O, Odoni AR (1994) Dynamic solution to the ground-holding policy problem in air traffic control. *Transportation Res. Part A* 28(3):167–185.
- Rios J, Ross K (2010) Massively parallel Dantzig-Wolfe decomposition applied to traffic flow scheduling. *J. Aerospace Comput., Inform. Comm.* 7(1):32–45.
- Sherali H, Hill J, McCrea M, Trani A (2011) Integrating slot exchange, safety, capacity, and equity mechanisms within an airspace flow program. *Transportation Sci.* 45(2):271–284.
- Terrab M, Odoni AR (1993) Strategic flow control on an air traffic network. *Oper. Res.* 41(1):138–152.
- Terrab M, Paulose S (1993) Dynamic strategic and tactical air traffic flow control. Technical report, Rensselaer Institute of Technology, Troy, NY.
- US Department of Transportation Bureau of Transportation Statistics (2012) Airline on-time statistics. <http://www.transtats.bts.gov/HomeDrillChart.asp>.
- Vossen T, Ball M (2006a) Optimization and mediated bartering models for ground delay programs. *Naval Res. Logist.* 53(1):75–90.
- Vossen T, Ball M (2006b) Slot trading opportunities in collaborative ground delay programs. *Transportation Sci.* 40(1):29–43.
- Vranas PB, Bertsimas D, Odoni AR (1994) The multi-airport ground holding problem in air traffic control. *Oper. Res.* 42(2):249–261.
- Wambsgans M (1996) Collaborative decision making through dynamic information transfer. *Air Traffic Control Quart.* 4(2):107–123.