Collaborative Decision-Making in Air Transportation

Robert L. Hoffman *
William D. Hall †
Michael O. Ball ‡
Amedeo R. Odoni §
Michael Wambsganss ¶

Abstract

This paper describes the collaborative decision-making (CDM) initiative being undertaken in the air transportation community. CDM's initial focus was on shortcomings of ground delay programs used to regulate traffic flow in the present air transportation system. This paper follows the past and present issues in Air Traffic Flow Management (ATFM) addressed by the CDM effort and the future extension of CDM research. This paper presents the new roles forming between industry and government in the air transportation system, as well as the algorithms, methods, and technologies employed by the CDM initiative to realize this new way of doing business. The paper also describes CDM efforts to improve several important aspects of the present air transportation system, including aircraft routing and arrival-departure resource allocation, and it discusses convergence and quality of solution issues with respect to those efforts.

1 Introduction

The traffic demand placed on the national air transportation system has increased steadily since the early days of flight. Until recently, the system capacity was adequate to handle this demand, but according to several studies (see [?], [?] and [?]) the delays incurred during routine operations will become very large in the next ten years. Indeed, the average gate-to-gate time for a flight between Chicago's O'hare and Boston's Logan airports is the same

*University of Maryland
†Draper Fellow, Massachusetts Institute of Technology
‡University of Maryland
§Massachusetts Institute of Technology
¶Metron Sciences, Inc

today as it was in 1950 [?]. The advantage of the tremendously increased speed of travel enroute has been nullified by increased delays elsewhere in the system.

The delays in the system are primarily due to limited capacity at the airports. Whereas enroute capacity is limited, the effects of limits in enroute capacity are dwarfed by the effects of limits in airport capacity. For example, over 90% of the delays in air travel are caused by capacity shortfall and subsequent congestion at one of only 22 of the 18,224 airports in the country [where did this statistic come from?][?]. Not surprisingly, these 22 locations coincide with the major metropolitan areas, which are prime destinations and points of origin for passengers. Furthermore, air traffic demand at these airports is projected to grow very rapidly over the next several years. For these reasons, limited airport capacity is a major concern to the airlines, the FAA, the aircraft manufacturing sector, and the public.

One perspective of the recurring capacity shortfall is that demand at the major airports is too great. The amount of demand is large and growing, and it is distributed unevenly through the day. The mean level of demand is considerably less than the mean capacity, but the variations in demand cause it to exceed capacity regularly.

There are several reasons for the peaks in demand. One reason for the peaks is that the flights departing closest to the times requested by a travel agent show up first on his or her screen. To capture more passengers, the airlines frequently schedule flights to depart at the precise times people are most likely to request departures. Seldom will a person request to leave at 7:17 AM. As a result, the airlines schedule many flights to depart at even hours and half hours. For example, on Monday December 15, 1997, of the 24 flights scheduled to depart Boston between 7:00 AM and 7:30 AM, eight flights were scheduled to depart at exactly 7:00 AM, four flights were scheduled to depart at exactly 7:30, and of the remaining eight flights, only two were not scheduled to depart on an even five minutes.

Another cause of demand peaks is the hub-and-spoke organization around which many carriers design their schedules. The hub-and-spoke system is a simple and effective way for a carrier to provide connectivity between every pair of airports it serves. Rather than having a separate flight for each airport pair, the carrier provides flights between many of the airports and a specific hub airport. Passengers can take a flight to the hub airport, change aircraft, and depart for their destinations on the new aircraft.

The most efficient way for a hub-and-spoke operation to work, from the passenger's point of view, would be for all incoming flights to arrive at the same time in one bank of aircraft, followed about a half hour later by a departing bank of aircraft. The airlines try to realize this ideal by scheduling banks of aircraft arriving and departing at the hub airport at very similar times. This practice creates spikes in demand that often cannot be accommodated by the airport resources. For instance, on Monday December 15, 1997, thirteen United Airlines flights were scheduled to depart Chicago's O'Hare International Airport at exactly 2:14 PM, far more than the airport can handle in one minute!

Since there would be little support for any plan that asked the airlines to curtail well-established operational paradigms or that asked the general public to travel at odd times of the day to unpopular locations, it seems more promising to focus efforts on increasing airport capacity. Long range plans to increase airport capacity include the construction of new facilities and the expansion of existing ones. Unfortunately, such plans are hindered by the fact that real estate in major metropolitan areas is often unavailable or prohibitively expensive and that expansion of facilities is often blocked by surrounding communities concerned with the noise and automotive traffic congestion such expansion would bring. A more cost-effective way to increase capacity may be to use the existing system resources more efficiently. Several efforts to do so have been undertaken recently. Notable among them is Collaborative Decision-Making (CDM). Section 1.1 briefly discusses the various approaches to improving system efficiency and explains the merits of the CDM approach. Section 1.2 describes in more detail how CDM can capitalize on its merits, and Section 1.2 lays out the remainder of the paper.

1.1 Efficient Use of System Resources

In the quest to use system resources efficiently, it would be nice to know that one is operating the system as efficiently as possible. Optimization-based methods have the potential to produce the best use of the resources possible, where "best" is defined by an objective function. Optimization-based methods frequently find solutions that are both intuitive (in hindsight) and far better than solutions found by other methods for problems in a wide range of application areas, from telecommunications to airline fleet scheduling.

Optimization-based solutions to a wide variety of problems have had tremendous impact in the airline industry in recent years. For example, Sabre Decision Technologies estimates that its clients are improving their profitability by more than $500 million annually through use of Sabre's optimization-based yield management software [Sab]. The profitability of a logistically complicated business often can be improved dramatically by changing a few innocent-seeming decisions.

Buoyed by the success of optimization-based approaches applied to real problems in industry, several researchers have developed optimization-based models to solve the FAA's decision problems related to air traffic flow management. The models presented in [Ter90], [Vra92b], [Vra92a], and [Ric91], among many others, each solve different aspects of the ground holding problem. The model presented in [BS96] addresses the ground holding problem combined with the routing problem.

None of the optimization methods designed to solve the FAA's problems has been accepted by the air transportation community. The difficulty is that these optimization methods rely on models of the problems they are to solve. The problem models consist of objectives and constraints that are designed to approximate the real objectives and constraints in the problem. Therefore, such optimization methods may improve operation of a system only to the extent that the models used accurately reflect the system characteristics. When optimization methods are applied within a single company, the company is motivated by the bottom line to provide accurate data to the method. Unfortunately, application of optimization methods to many of the problems faced by the FAA involves data from many businesses, local governments, and private concerns.

In addition to the technical difficulties involved in gathering the data from the disparate sources, there are fundamental game-theoretic issues that must be addressed to ensure the accuracy of the data. Numerous interests compete for use of the NAS, each of which would like the FAA's objectives to serve it at the expense of its competitors. If providing inaccurate data to the FAA could work to a party's advantage, it is quite likely that the data from the party will contain some inaccuracies. Indeed, this is a problem with the present system used by the FAA to schedule Ground Delay Programs (GDP's), as is thoroughly discussed in [Wam96]. An effort begun in the summer of 1993 to improve the data used by the FAA in scheduling GDP's, the FAA-Airline Data Exchange (FADE) program, discovered that improving the FAA's incomplete data would result in significantly better decision-making. FADE also found that the method used by the FAA to implement GDP's penalized those airlines that provided the additional information. The FADE participants invented a modification to the method used by the FAA to implement GDP's. The modification is embodied in an algorithm dubbed *Compression*, and its implementation has been the initial focus of the CDM effort.

The CDM program has expanded its focus to include other problems in the air transportation system. The challenge for CDM and the FAA in all their efforts is to find a way to combine information from the many competing interests to produce solutions that all users can agree are good. The information from the competing interests must be used in a way that motivates them to provide accurate information; in the literature of game theory, this is called *incentive compatibility* (see, for example, [FT91]).

1.2 The Collaborative Decision-Making Effort

The driving philosophy behind CDM is that improved data exchange and communication between aviation transportation organizations will lead to better decision making in air traffic flow management and that, whenever practical, those decisions which have potential economic impact on airline operations should be decentralized and made in collaboration with the airlines.

As CDM has grown, so has its participation. The passenger airlines and FAA have been joined by the package carriers such as Federal Express and United Parcel Service (UPS), who have a vested interest in the outcomes of new methodologies being established by CDM. The technical challenges posed by CDM have solicited the expertise of scientific consulting firm Metron Inc., the Mitre Corporation and a host of contractors. In 1995, Congress created a National Center of Excellence for Aviation Operations Research (NEXTOR) which centers around a consortium of four universities: the University of Maryland, the Massachusetts Institute of Technology, the University of California-Berkeley, and Virginia Polytechnic Institute. (See [Van97] for a reference on NEXTOR.) Representatives from the CDM participating organizations, which now includes NEXTOR, gather monthly for working-group meetings. Many of the airlines have devoted full time representatives to CDM.

The primary focus of CDM has been ground delay enhancements and the improvement of the cancellation and substitution process used by the airlines. At the time of this writing, CDM methodologies have not been put into place. Prototype operations involving ground delay program enhancements are scheduled to go in effect at San Francisco and Newark airports in early 1998. CDM efforts are extending to any constrained situation in the NAS. Given the current momentum of CDM and its growing support, it seems inevitable that CDM methodologies will have a major impact on air traffic flow management, particularly in the context of ground holding strategies. It is appropriate that the current body of aviation research should reflect, if not be tailored to, the CDM philosophy. In this section, we examine the procedures and technologies that have arisen out of CDM efforts.

[Bob, some of the above three paragraphs might be redundant – we should read over it again and clean up once more after your stuff is added here.]

[outline of paper – should this be a separate subsection?]

2 Ground Delay Programs

The Air Traffic Control Systems Command Center (ATCSCC) monitors airports throughout the United States for capacity shortfall. Whenever it is predicted that the number of flights arriving at an airport within a 15-minute time interval will exceed the number of flights scheduled to land, the ATCSCC is required by FAA regulation to take some form of action.[1] Short-term periods of capacity shortfall are alleviated by airborne tactics such as re-routing and variations in airborne speed. Longer-term periods of capacity shortfall are met by the ATCSCC with ground holding strategies in which aircraft are held at their departure gates in lieu of costly and dangerous airborne delay.

In some cases, the ATCSCC will issue a ground stop in which all flights destined for an

[1]Much of the information in this section concerning the operation of the ATCSCC was obtained through meetings with ATCSCC personnel.

afflicted airport are held on the ground at their departure gates until airport arrival capacity rises above demand. These ground stops are reserved for extreme cases in which arrival capacity was severely underestimated or dropped suddenly without warning.

The primary tool of the ATCSCC for addressing arrival capacity shortfall is a ground delay program (GDP). In a GDP, each flight scheduled to arrive at an afflicted airport over a fixed time period is held at its departure gate long enough to ensure that it will be able to land without delay. For instance, if flight f is due to arrive at airport A at 12:00 and it is known that f will not be able to land until 12:30 due to limited arrival capacity at A, then f would be held at its departure gate for 30 minutes. The construction of a GDP requires the assignment of both a controlled time of departure (CTD) and a controlled time of arrival (CTA) to each incoming flight . Since en route travel times can be predicted with reasonable accuracy, the CTD of each flight is easily calculated once its CTA is known. The CTA is set once the flight has been assigned an arrival slot. [what about MAR?]

The implementation of a GDP is generally met with trepidation by the airlines because it represents government intervention and exposes airline operations to inaccurate estimations on the part of the ATCSCC. For instance, if the future arrival acceptance rate of an airport is overestimated, then arrival demand will exceed capacity and planes will absorb delay in costly airborne holding patterns rather than on the ground at their departure gates. On the other hand, if future AAR's are underestimated, then arrival capacity will exceed demand, arrival slots become a wasted resource, and flights absorb unnecessary ground delay. The airlines find it particularly disruptive to their operations when a GDP is aborted in mid-operation. This happens whenever the ATCSCC has clear evidence that the original capacity and weather forecasts were overly pessimistic. Estimates vary between the airlines and the FAA as to the percentage of GDP's that are aborted, but these estimates are as high as 60%.

To a large degree, a GDP is based upon the economic premise that ground holding is cheaper than airborne holding. While this is certainly true on average, it does not hold for every aircraft in every delay situation. Certainly, the airlines recognize the need for centralized control of air traffic into an airport with limited arrival capacity but these situations serve as a prime example of a major objection of the airlines to the current procedures. The objection is that the FAA is making economic decisions on behalf of the airlines. These decisions are generally well- intended but it has been firmly established that the FAA lacks the timely data on daily airline operations and the private objective information of the airlines required to make these types of decisions effectively.

2.1 Grover Jack

Currently, the ATCSCC assigns arrival slots during capacity-limited periods by a first-come, first-served algorithm known as *Grover Jack*. In Grover Jack, a list of incoming flights is formed, ordered by the most recent estimated time of arrival (ETA) of each flight. The time horizon for the GDP is divided into hourly periods $t = 1, 2, ..., T$. Usually, $T = 4$. The arrival acceptance rate (AAR) of a time period t, denoted X_t, is defined as the number of aircraft that can be accepted during period t. (Strictly speaking, this is not a rate. Nonetheless, it is the established terminology in the air traffic community.) A value of X_t is set for each time period t by a specialist at the ATCSCC. In spirit, at least, the Grover Jack algorithm assigns controlled arrival times as follows. The first X_1 flights on the list are assigned to the first time period, the next X_2 flights are assigned to the second time period, and so on, preserving order of the list. The net effect of the Grover Jack algorithm is to stretch out the list of incoming flights over time.

In reality, there are several complications that need to be addressed in the Grover Jack algorithm. Since a flight cannot be assigned to a time slot earlier than its ETA, some time slots will be passed over during the assignment process and have no flight assigned to them. Additionally, international flights, general aviation, and flights airborne at the time of formulation of a GDP are exempt from the program, meaning that they cannot be issued a ground delay. In addition, the specialist may choose to exclude other categories of flights from ground delay, usually based on geographical location of point of origination. The arrival of these flights must be taken into account when assigning CTA's to flights.

2.2 Substitution and Cancellation

[need stuff]

Although the airlines agree that a first-come, first-served algorithm is an equitable method for distribution of arrival slots, they object to the use of ETA as the criterion for 'first-come'. They have cited the following incentive incompatibility, known as the *double penalty issue.*[Wam96] Suppose that flight f is scheduled to arrive at airport A at 10:00 hours. If f is delayed for 30 minutes with, say, mechanical problems, then the ETA of f would be updated to 10:30. If a GDP is implemented in which f is assigned a 30 minute delay, then f would be given a CTA of 11:00. Overall, f has suffered a total delay of $30+30 = 60$ minutes. Had the airline not reported the mechanical problem until after the ground delay program was implemented, its flight would have seen only 30 minutes of delay. In essence, the airline is penalized for providing updated information about flight f to the FAA. As a result, it has been observed that many airlines have withheld cancellation and delay information from the FAA. In some instances, the FAA has had to implement a GDP that would have been prevented had the FAA known the accurate demand information. This motivates use of the original scheduled time of arrival rather than ETA to determine landing order in a GDP.

Regardless of the criterion used for first-come in the formulation of a GDP, it is highly unlikely that any simple ordering of flights will be preserved when the program is finalized. One reason, already mentioned, is that a significant portion of the flights are not subject to the ordering criteria because they are exempt or excluded from the program. Also, flights can be delayed (or canceled) by the airlines for reasons such as mechanical difficulties.

2.3 Modeling Ground Holding

The problem of assigning ground delay to flights bound for a single airport can be mathematically modeled as an assignment problem known as the *ground holding problem* (GH). The model requires the following assumptions.

Assumption 1: (Discrete time horizon) There is a fixed time horizon which has been discretized into T equally-sized contiguous time periods, $t = 1, 2, ..., T$.

Assumption 2: (Deterministic demand) The number of incoming flights is known in advance; for each flight f, there is a scheduled time (period) of arrival, denoted a_f (this is the earliest arrival time that can be assigned to the flight).

Assumption 3: (Deterministic capacity) For each time period, t, let b_t be the arrival acceptance rate (AAR) of the airport, meaning the maximum number of flights that can be accepted by the airport during that time interval. Then we assume that b_t is known in advance for each time period t. Strictly speaking, this does not hold in practice because the AAR's are dependent upon weather conditions and runway configurations, which are stochastic in nature. However, the specialist who formulates the GDP fixes these numbers in accordance with the current best estimate, so, for purposes of this formulation, we will assume that they are deterministic and known in advance.

Let F be the set of incoming flights that require arrival slots. We define for each f and each t, a binary variable, $X_{f\,t}$, such that

$$X_{f\,t} = \begin{cases} 1, & \text{if flight } f \text{ is assigned to time interval } t \\ 0, & \text{otherwise.} \end{cases}$$

Each flight must be assigned to exactly one time interval so for each f we have one constraint of the form $\sum_{t=a_f}^{T} X_{f\,t} = 1$. The number of flights that are assigned a CTA within time period t cannot exceed the capacity b_t, so for each time interval t there is one capacity constraint of the form, $\sum_f X_{f\,t} \leq b_t$.

Let C_f be the cost of delaying flight f for one time period and let $\sigma > 1$ be a fixed parameter. Then the expression $X_{f\,t}\ C_f\,(t - a_f)^\sigma$ represents the cost of assigning flight f to time t and the summation of this term over all t and all f is the total delay cost. The parameter σ yields super-linear growth in the tardiness of a flight as t increases. This favors the assignment of a moderate amount of delay to each of two flights rather than the assignment of a small amount of delay to one and a large amount to the other.

In all, we have the following integer program.

(GH)

$$\text{Minimize} \sum_{f \in F} \sum_{t=1}^{T} X_{f\,t} C_f\,(t - a_f)^\sigma \tag{1}$$

subject to

$$\sum_{t=a_f}^{T} X_{f\,t} = 1 \tag{2}$$

$$\sum_{f} X_{f\,t} \leq b_t \tag{3}$$

$$0 \leq X_{f\,t} \leq 1 \tag{4}$$

$$X_{f\,t} \in \{0, 1\} \tag{5}$$

Let GH_{LP} be the LP relaxation of GH, that is, the problem that results from relaxing constraint set (5). GH_{LP} is a transportation problem. Since it can be shown that there is an integer optimal solution to every transportation problem, LP solvers or specialized transportation codes can be applied to GH_{LP} to obtain the (integer) solution to GH rapidly.

GH was first systematically described by Odoni in [Odo87]. Andreatta and Romanin-Jacur [ARJ87] treated the stochastic version of GH for an airport with constrained arrival capacity in (at most) one time period. In [TO93], Terrab and Odoni developed a dynamic programming formulation for the stochastic version of GH as well as heuristics to handle the larger cases. Using stochastic linear programming with recourse, Richetta and Odoni expanded this work to include the dynamic case, in which ground holdings are updated as time progresses (see [RO93]). Although their dynamic solution yielded considerable savings over the static solution, the speed of solution proved to be too slow for realistic cases. See [BDST93].

Ideally, the ground holding problem should be solved on a network-wide level, taking into account the connectivity of flights. Flights can be connected in one of three ways: by passenger, crew or aircraft. In the former sense, passengers are scheduled to travel from airport A to C by taking a flight from A to B, then B to C. The arrival of the first flight should coincide (roughly) with the departure of the second flight. In the latter two senses, a single crew or aircraft may be scheduled to traverse many flight legs, e.g., from city A to city B to city C, and so on. The delay of even a single flight can propagate throughout the entire system.

Both GH and air traffic flow management in general have been treated on a network-wide level (taking multiple airports and flight connectivity into account) in Attwool [Att77], Sokkapia [Sok85], Andreatta and Romanin-Jacur [ARJ87], Wang [Wan91] and by Vranas, et. al., in [VBO94a] and [VBO94b], and, more recently, by Bertsimas and Stock [BS96].

In practice, there are several problems associated with applying network models of the air transportation system. The first is that the models are generally difficult to solve (NP-hard) integer programs. Much progress has been made toward solving these models in real time, although some of the progress has required loss of fidelity. More importantly, however, the models require extensive, timely information about all the constraints and objectives of every user of the system. Without this information, the solution produced by the model cannot be expected to serve the best interests of every user, yet much of the information is combinatorial in nature and cannot be effectively communicated in real time. In essence, the information is inherently unknowable by the FAA – the airline can evaluate the value of different solutions to the problem, but cannot accurately describe the objective and constraint functions a priori. Furthermore, even if this information were communicable, the airlines would be wary of sharing it since this information captures an airline's trade secrets.

The fact that it is impossible for the FAA to understand each airline's business accurately makes any attempt by the FAA to help solve the airline's problems likely to fail. The airlines have learned this through experience. It is this lesson that fuels the movement for distribution of decision-making to the users of the National Airspace System and away from the FAA.

3 Roles and Procedures under CDM

3.1 New Roles

The responsibility of the FAA is to maintain the safety of the users of the NAS and to manage air traffic flow in a manner that makes efficient and equitable use of resources. The current structure imposed on the system stems from the belief that the FAA and its agencies should act as a centralized authority and decision maker. This paradigm, as well as the many of the procedures and standards for air traffic management, were developed just after World War II. Despite the enormous growth in air traffic in the United States in the last fifty years, this paradigm remains in effect today, largely unaltered.

There are several problems with maintaining the central decision-making paradigm and current operating procedures. First, the National Airspace System has grown too large to be efficiently managed by one authority. [Bill: can you add some neat sentence(s)] [cite growth statistics]

Secondly, the airlines have been given little input into the decision making process. From their perspective, it is a black box. [cite instances/examples of planes being diverted, rerouted without explanation, etc.]

Lastly, the FAA has spilled over into the area [word choice] of making economic decisions on the behalfs of the airlines [word choice]. The trouble with this is that they do not have the expertise nor timely data necessary to understand airline operations on a detailed level. [give examples] Also, airlines function by different philosophies.

For these resons, the airlines have become increasingly disgruntled wth the decision making process of the FAA and the entire central authority paradigm.

The efforts of CDM have been guiding air traffic management in the direction of decentralized decision making, thus giving the scheduled carriers input to air traffic management and greater control over their operations.

A novel relationship between the scheduled carriers and the FAA was formalized in the document, "Roles and Responsibilities", written by the FADE program manager and airline representatives in early 1995. It specifies that the ATCSCC should remain a neutral party and act as a service provider to the users of the NAS. The responsibility of the ATCSCC is to alert the users to situations within the NAS that place constraints upon their operations. The users are responsible for responding to those constraints with actions, intents and preferences that lie within those constraints.

CDM has put this into effect on a prototype basis. (see next section). even if doesn't work all that well, paved the way for new levels of cooperation and collaboration between government and industry in air traffic management.

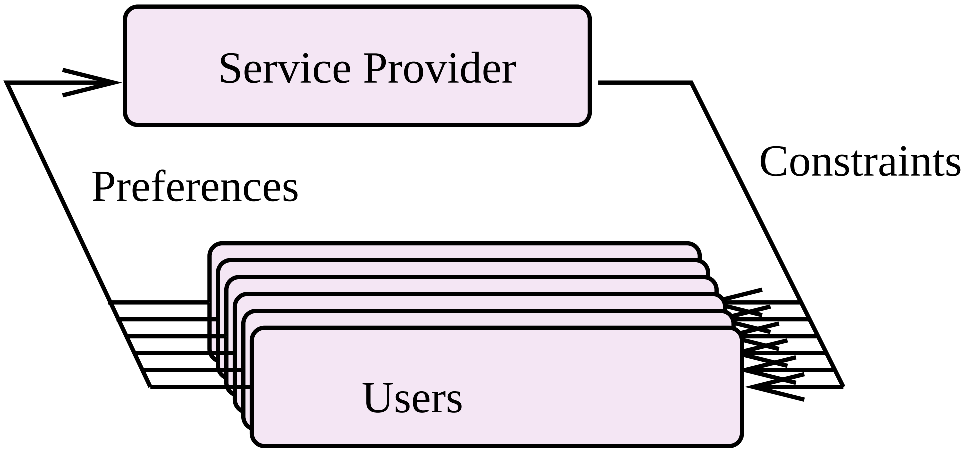

Figure 1: Constraint Dissemination and User Preference Loop (CDUP)

3.2 Underlying Procedure for CDM

The ubiquitous operational procedure for handling a constrained situation in the NAS is a cycle of feedback between the service provider and the users called constraint dissemination and user preference (CDUP). A single iteration of the cycle begins with an analysis of the constrained situation and the dissemination of constraints from the service provider to the users. For example, in the event of a ground delay program, this would consist of predicting a drop in AAR at an airport and notifying the airlines that a GDP will be in effect. Next, the users (airlines) asses their role in the constrained situation and the potential impact on their operations. Each user reacts with a set of preferences. In the case of a GDP, these preferences would be a list of cancellations and substitutions. Then, the service provider updates the information, honoring preferences when possible, and reanalyzes the situation. This completes one iteration of the cycle. The cycle repeats, perhaps at regular intervals, either until the situation is played out or until it is alleviated through user reaction.

CDUP allows for any degree of decentralization of decisionmaking. At the decentralized extreme, the service provider provides an aggregate picture of the traffic situation, and the users are responsible for changing their parts of the solution to achieve overall efficiency. For instance, suppose that under the free flight plan [have we introduces this yet?], enough airlines have filed a flight plan over the same sector of airspace to make it overloaded [reword this]. Then the service provider would issue the projected demand and capacity of sector, highlighting the inequity. Hopefully, enough of the flights would re-file their flight plan over another sector to bring demand back in line with capacity for the sector in question (this might require several iterations of the cycle). This type of collaboration is attractive because it gives the users maximum control over their operations.

At the other extreme, the users are given a set constraints by the service provider that effectively reduces their options to a single option. The user "preference" that is relayed back to the service provider is just an acknowledgement of that option. In this case, CDUP reduces to the centralized decision making.

The problem with the former use of a CDUP is that it assumes that user preferences and intentions are in accordance with the overall alleviation of constraints. This could fail when users are competing for the highly valuable resources. On the other hand, the centralized decision making paradigm is exactly what CDM is trying to avoid. For most situations, the optimal amount of collaboration lies somewhere between these two extremes. A blend of centralized decision making and user decision making can be achieved by housing additional constraints imposed by the service provider inside the set of the constraints disseminated to the users.

These constraints can be used to be employ a resource allocation scheme. For example, suppose that the arrival capacity of an airport has been reduced from 70 to 40 flights per hour. This reduction is a natural constraint on the situation. If in addition, the ATCSCC allocates say, 20 of those arrival slots to American Airlines, then the service provider has imposed an artificial constraint in the interest of overall safety and fairness.

CDUP can be somewhat cumbersome. It requires time for both the service provider to aggregate the situation and for the users of analyze their options. Each airline must be in sync with iterations of the CDUP loop. It assumes that the users are willing to respond to constraints and that their is some reliable mechanism for data exchange between the service provider and the users. The incentives and mechanisms for data exchange that have been developed by the CDM working group to be employed during ground delay programs are the explained in the next two sections.

[RBS and Compression] [FSM]

Routing Chapter:

CDUP cnn be extended to routing of aircraft how so?

this could simply consist of transmitting a traffic picture to the airlines.

3.3 Algorithms for CDM

The incentives for the airlines to provide timely data to the FAA during the formulation of a GDP have been housed in the two algorithms, **ration-by-schedule** *(***RBS***)* and **compression**. We now give brief descriptions of those algorithms and their purpose. More detailed analyses and formal presentations can be found in Appendix **??**.

The RBS algorithm: The purpose of this algorithm is to ration arrival slots according to original scheduled arrival times of flights and to serve as an initial assignment of CTA's for subsequent rounds of collaboration between the airlines and the FAA. The original scheduled arrival time of a flight is determined by the original gate time of arrival (OGTA) minus a standard ten-minute taxi time.

The primary difference between RBS and the Grover Jack algorithm currently used by the FAA is that RBS uses OGTA−Taxi time as the ordering criterion whereas Grover Jack uses ETA. Otherwise, the algorithms are very similar. Details aside, virtual arrival slots are created over a time horizon based on AAR predictions and the first flight is assigned to the first time slot, the second flight is assigned to the second time slot and so on. Under the RBS algorithm, airlines can effectively reserve slots at an airport by scheduling flights weeks ahead of time in the OAG (official airline guide). This removes the fear that of forfeiting an arrival slot by reporting a delay, as in the double penalty issue.

The initial assignment of CTA's made by RBS plays an important role in the subsequent cancellation/substitution process. Suppose that flight $AL100$ belonging to airline AL has been assigned to the t^{th} time slot. Then AL is said to "own" slot t. During a round of cancellation/substitution, AL is free to move another one of their flights, say $AL200$, from a less desirable time slot, $(t+k)$, into t, provided that t is a feasible arrival time for $AL200$. Flight $AL100$ will have to be cancelled, of course, or moved down into slot $(t+k)$, because the total number of slots owned by AL is fixed by RBS. By this process, each airline has the opportunity to minimize the damages of flight delays in a GDP.

The Compression algorithm: After a round of cancellation/substitution, the total number of flights in the program has probably been reduced through cancellations. This creates "holes" in the assignment schedule, meaning there are valid arrival slots with no flights assigned to them. The purpose of the compression algorithm is to move flights up in the schedule (earlier in time) to fill these slots.

There are two ways for an arrival slot to become empty. Either (1) the airline assigned control/ownership of the slot by RBS has simply declined to substitute a flight into the slot or (2) the slot is too early for any flight to be assigned to it by the controlling airline . In either event, the controlling airline would *release* the time slot to the compression algorithm which tries to fill the slot with another flight of the controlling airline or to make an appropriate compensation.

The abbreviated version of the compression algorithm is as follows.

1. For increasing values of t, the status of time slot t is checked. If t is filled, the algorithm moves on to slot $t+1$.

2. If t is released, then it has a controlling airline, AL. The algorithm first scans down a list of flights from AL for a feasible assignment. If no AL flight is found, it searches for a feasible assignment on a list of all flights not in AL. If a flight f from AL (or another airline) is found for t, then f has been moved up from its current slot t'. Then, t' is declared released with controlling airline AL and the algorithm is applied to t'. This creates a stream of substitutions (possibly none) of flights upward through the schedule. Eventually, this stream stops (see the full algorithm in Appendix A for stopping criteria) and the algorithm resumes at slot $t+1$, where it left off.

   The algorithm terminates when every slot has been filled with a flight or declared unusable.

The important features of the compression algorithm are that (i) arrival slots are filled whenever possible, (ii) flights from the controlling airline of a slot t are considered before all others when t is released, (iii) if the controlling airline cannot use a slot it is (eventually) compensated since it receives control of the slot vacated by the flight which moves into its slot and (iv) there is no way for an airline to involuntarily lose a slot reserved by RBS (provided that the criteria for the termination of a substitution stream are properly set). Issues such as the ordering of lists and the criteria for feasibility of a flight to be moved to a time slot are discussed in Appendix A.

3.4 Technological Implementation

In order to coordinate all of the activities surrounding the cancellation/substitution process and the foregoing algorithms, the participating airline operational centers (AOC's) and the ATCSCC have been networked by the newly established "AOC Net". The AOC Net carries the airline data to the ATCSCC and an aggregated demand list to each airline .

Collaborative decision making and common situational awareness of airport arrivals are made possible by a decision support tool called the *Flight Schedule Monitor (FSM).* The major features of FSM are listed below.

- Common situation awareness: When used on-line, FSM displays both to the NAS user and the ATCSCC the latest projections of airport arrival demand at the user-selected airport. The demand is displayed in a time line fashion in which each tick mark is a minute of time and incoming flights are represented by icons on the time line. The position of the icon corresponds to the most recent ETA of the flight. Airlines can be distinguished by an icon coloring scheme.

- Analytical tools and strategy evaluations: To explore alternative ground holding strategies, both the NAS user and the ATCSCC can formulate hypothetical ground delay programs to generate delay statistics and measure equity amongst the airlines. Also, FSM has an (off-line) historical mode in which the user can conduct analyses on archived flight data.

- Information processing: FSM (1) executes both the RBS and Compression algorithms, (2) allows users to make cancellations and substitutions and (3) processes an updated demand list together with user-specified parameters such as AAR's to generate a complete ground holding strategy (such as GDP or ground stop). In particular, controlled times of arrival and departure are output for each flight in the program. However, only the ATCSCC has the capability to institute a program.

- Compliance monitoring: FSM allows the NAS users and service provider (ATCSCC) to verify compliance with established rules of conduct within the system. For instance, an alarm is tripped whenever a flight departs that was reported cancelled by an airline.

Although FSM was developed as the primary flow tool for the FAA, the airlines have made significant contributions and improvements to FSM through experimental sessions (war games), user feedback and concept design. FSM is in the final stages of testing and development and is now operating on-line at the ATCSCC and 14 airlines.

One can see that, from beginning to end, the CDM process for addressing reduced arrival capacity at an airport is rather involved. It promises to be an improvement, however, on the current system in which the ATCSCC acts as sole decision maker operating with incomplete information. By making the users aware of their own role in a constrained situation, it is possible that the need for government intervention can be greatly diminished.

4 The Routing Problem

Sectors have capacities, blah blah blah, interesting statistics ,etc.

Under current operations, the airline specifies a route for each flight to follow when filing the flight plan for the flight prior to takeoff [how true is this? how often is it rejected?].

If a sector along the flight's path becomes overloaded, the controllers in that sector usually issue miles-in-trail restrictions on the aircraft that have yet to enter the sector. Miles-in-trail restrictions specify that aircraft entering the sector along the same route must be separated by a certain number of miles along the path of travel. In this way, the rate of aircraft entering the sector is reduced so that the controllers are able to maintain safety. However, the miles-in-trail restrictions force the aircraft that have not yet entered the sector to sustain delay in the queue of aircraft waiting to enter. Anecdotal evidence indicates that, in hindsight, the aircraft waiting to enter the sector could in many cases have reduced their overall delay had

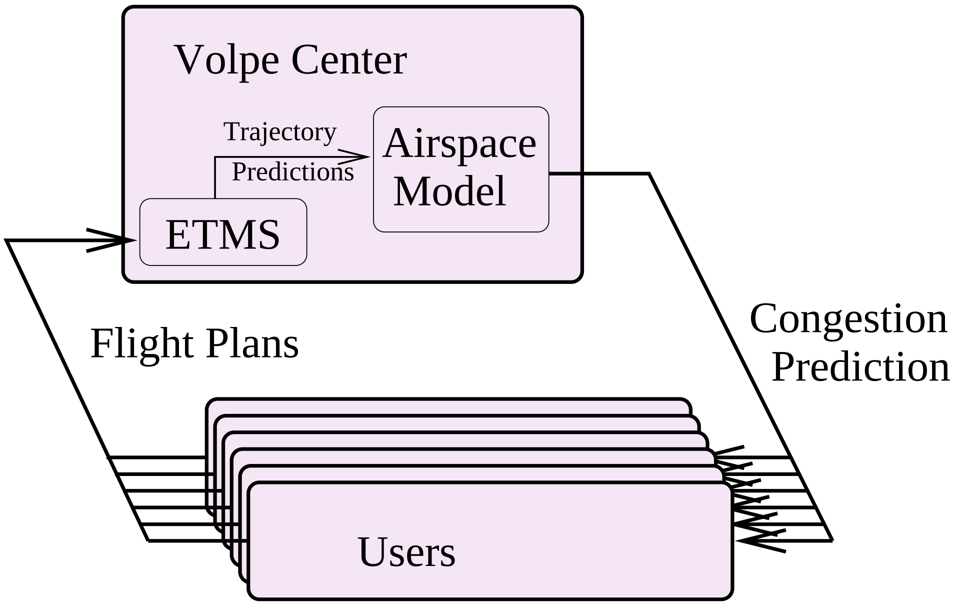

Figure 2: En Route CDUP

they chosen a less congested route [Gro97]. Unfortunately, in the current system, there is little information to help an airline decide which routes are likely to be congested a priori.

4.1 Extensions of CDM to Routing

The CDM community is currently considering benefits that could be produced by applying the underlying CDM procedure, CDUP, to the flight routing problem. The envisioned system would use data from the Enhanced Traffic Monitoring System (ETMS), a tool developed by the Volpe National Transportation Systems Center that predicts aircraft trajectories, to predict when and where the congestion problems will occur. These predictions of sector loading would be distributed to the airlines over the AOCNet. The airlines then have the ability to change flight plans in response to the predicted congestion, resulting in an update to the predictions, and so on. Figure 2 depicts this process.

4.2 Ball Method

A modification to this arrangement was proposed by Mike Ball. Basic idea is to relieve the airlines of the burden of examining the congestion picture every time it is updated, allowing them to specify their preferences for various routes as a function of the predicted delay along the route and have the FAA determine the flight plan based on those preferences. In essence, Ball's proposal allows the users to file intelligent flight plans – file a plan to take the fastest route, or shortest route, or any combination of preferences involving the length of time it will take to travel the route.

4.3 Brief Analysis

The En Route CDUP arrangement being investigated by the CDM community and described in Section 4.1 can be considered in the context of price-decomposed hierarchical optimization problems (of which Dantzig-Wolfe decomposition is the textbook example. See, for example, [ST78] for more background). In this context, the Volpe Center solves the master level problem, enforcing the separation and sector capacity constraints, and the subproblem solved by each user is the routes that the subset of flights owned by the user should take. The congestion predictions passed down from the Volpe Center to the users contain nonlinear price information in the form of the delay imposed on each flight due to congestion. The flight plans resubmitted by the users to the FAA and Volpe Center are the subproblem solutions. Of course, the controllers still maintain feasibility at every step; if a proposed flight plan would overload a sector, the FAA may still reject it or issue miles-in-trail restrictions upon execution of the plan to maintain safety.

The recognition of this solution technique as an optimization decomposition problem allows us to apply the results of the theory to it. Doing so illustrates some of the qualities we might expect to observe in the solutions generated.

The solution will always be feasible. At each iteration, the FAA maintains feasibility through the same methods used today. For example, if demand for a sector exceeds capacity during a given time period, the FAA might issue miles-in-trail restrictions during that time period. The resulting predicted delays would encourage some of the traffic to find alternate routes.

In general, it is possible for the planning process to be unstable, meaning that the subproblem solutions never converge. For example, consider the situation shown in Figure 3.

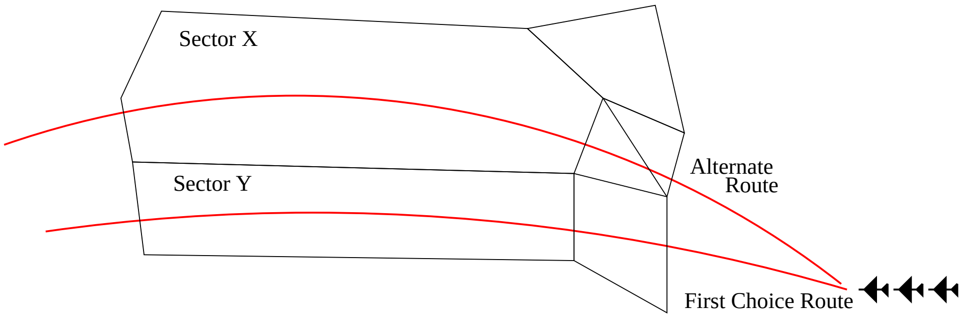

Figure 3: Two-Route Situation

In this example, a large number of flights have filed flight plans along the "First Choice Route" shown.

too many users switch upon seeing congestion.

Optimality of the generated solution is not likely duality gap price update method not right to guarantee that even if no gap most importantly, price to user is not a lagrangean relaxation example: first flight in queue sees little cost for going through the congested sector, yet he costs the guys behind him collectively quite a bit. Ball's approach could help solve these problems.

Problem: no guarantee that the solutions generated will be better than they would without collaborative routing. Example: airline A can slow down airline B enroute by adding aircraft to the head of the queue, resulting in airline A getting the landing slots it wants.

Conditions for this a bit extreme. Better information likely to improve solution, not degrade it. Still, possible.

5 Research topics

collaborative routing collaborative arrival/departure mix set stochastic model of GDP (weather, airport config, etc.) algorithms/opt models for replacing existing algorithms

References

[ARJ87] G. Andreatta and G. Romanin-Jacur. Aircraft flow management under congestion. *Transportation Science*, 21:249–253, 1987.

[Att77] V. W. Attwool. Some mathematical aspects of air traffic systems. *Journal of the Institute of Navigation*, 30:394–411, 1977.

[BDST93] M. Ball, R. Dahl, L. Stone, and T. Thompson. Optiflow build-i design document, version 1.5. Project report, What institution?, December 1993.

[BS96] D. Bertsimas and S. Stock. Air traffic flow management problem with enroute capacities. *Operations Research*, 1996.

[FT91] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, 1991.

[Gro97] CDM Working Group. Cdm meeting minutes, July 1997.

[Odo87] A. R. Odoni. *Flow Control of Congested Networks*, chapter The Flow Management Problem in Air Traffic Control, pages 269–288. Springer-Verlag, 1987.

[Ric91] Octavio Richetta. *Ground Holding Strategies for Air Traffic Control Under Uncertainty*. PhD thesis, Massachusetts Institute of Technology, 1991.

[RO93] O. Richetta and A. R. Odoni. Solving optimally the static ground-holding policy problem in air traffic control. *Transportation Science*, 27:228–238, 1993.

[Sab] Yield management solutions. Sabre Decision Technologies web page. Available as of 1 June 1997 at http://www.sabre.com/sabr_grp/sdt/solution/yield.htm.

[Sok85] B. G. Sokkapia. Arrival flow management as a feedback control system. Technical report, The Mitre Corporation, Washington, D.C., 1985.

[ST78] Madan G. Singh and Andre Titli. *Systems: Decomposition, Optimisation, and Control*. Pergamon International Library of Science, Technology, Engineering, and Social Studies. Pergamon Press, Oxford ; New York, 1978.

[Ter90] Mostafa Terrab. *Ground Holding Strategies for Air Traffic Control*. PhD thesis, Massachusetts Institute of Technology, 1990.

[TO93] M. Terrab and A. R. Odoni. Strategic flow management for air traffic control. *Operations Research*, 1993.

[Van97] Anthony R. Vanchieri. Aviation dream team. *ORMS Today*, pages 39–41, April 1997.

[VBO94a] P. Vranas, D. Bertsimas, and A. R. Odoni. The multi-airport ground-holding problem in air traffic control. *Operations Research*, 42:249–261, 1994.

[VBO94b] P. D. Vranas, D. Bertsimas, and A. R. Odoni. Dynamic ground-holding policies for a network of airports. *Transportation Science*, 28:275–291, 1994.

[Vra92a] Peter B. Vranas. Dynamic ground-holding policies for a network of airports. In *Working Papers*. MIT Sloan School of Management, 1992.

[Vra92b] Peter B. Vranas. The multi-airport ground-holding problem in air traffic control. In *Working Papers*. MIT Sloan School of Management, 1992.

[Wam96] Michael Wambsganss. Collaborative traffic flow management. Draft Available as of 1 June 1997 at Web Site http://www.metsci.com/faa/gen_doc.html, July 1996.

[Wan91] H. Wang. A dynamic programming framework for the global flow control problem in air traffic management. *Transportation Science*, 25:308–313, 1991.