Collaborative Approaches to the Application of Enroute Traffic Flow Management Optimization Models

Michael O. Ball
R.H. Smith School of Business and Institute for Systems Research
University of Maryland
College Park, USA
mball@rhsmith.umd.edu

David J. Lovell
Department of Civil and Environmental Engineering and Institute for Systems Research
University of Maryland
College Park, USA
lovell@umd.edu

Charles Glover
Institute for Systems Research
University of Maryland
College Park, MD USA
cnglover@math.umd.edu

Abstract—**Recent research has produced stochastic optimization models for adjusting traffic flows in response to predicted congestion in the enroute airspace. These models simultaneously consider a set of options for each flight that includes both the possibility of ground delay and reroutes. They take into account a distribution of possible weather scenarios and their outputs include alternate courses of action based on weather outcomes. The direct application of such models has been challenging because they do not provide the decision making flexibility to flight operators that has now become standard for collaborative air traffic management. In this paper we propose two different recipes for incorporating collaborative features into a meta-framework for the application of one such model. This paradigm for the combined use of optimization and collaboration is specifically geared for use within airspace flow programs, which are now in widespread use within the U.S. We provide computational results that demonstrate the effectiveness of this approach.**

Keywords: ATM; Air Traffic Manegement; Collaborative Decision Making; Rerouting; Stochastic Programming; Ground Delay; Airborne Delay.

I. INTRODUCTION

The initial development of Collaborative Decision Making (CDM) concepts for air traffic management (ATM) took place in the mid-1990's. In the U.S., CDM-based ground delay program (GDP) planning and control started in 1998 (see Ball et al. (2001) and Wambganss (1996) for background). The initial decision support tools and operational concepts have been modified and enhanced in many ways and are now used throughout the U.S. Further CDM-based tools are now used for enroute ATM, most specifically for the planning and control of airspace flow programs (AFP's). In Europe, "airport-CDM" systems and concepts have been developed and are moving toward widespread implementation (Eurocontrol, 2010). By now it is safe to say that any new

major traffic flow management system proposed either in the U.S. or Europe must adhere to Collaborative Air Traffic Management (CATM) philosophical principles and contain CATM features. In fact, CATM is a key tenet in both NextGen and SESAR.

Of course, as CATM has developed, there has been a stream of scientific research that has supported it and in some cases been fundamental to newly implemented systems. On the other hand, there is a well-established line of research into ATM optimization models that do not include a CDM viewpoint. Such models may determine a disposition for each of a large group of flights. Specifically, these models might determine a detailed departure time and route for all flights under consideration. Such models implicitly assume that a type of "central planner" can make and implement such decisions for all flights within the geographic area and time window of interest. In today's ATM environment, it is rare that either the air navigation service provider (ANSP) or the flight operator can independently exercise such universal control. Rather, in one way or another, the various entities involved collaborate to reach a comprehensive ATM plan.

In this paper, we consider a recent line of research (Ganji at al., 2009a, 2009b; Glover, 2010) that has produced powerful stochastic optimization models that are able to create a traffic flow management plan for a set of flights, whose preferred flight plans intersect a volume of airspace (a flow constrained area or FCA) where a severe capacity reduction is predicted over an extended period of time. This is in fact the typical scenario that motivates the use of Airspace Flow Programs (AFP's) within the U.S. These new stochastic optimization models determine for each flight an appropriate amount of ground delay and also whether a reroute (with or without ground delay) may be appropriate. They do so while considering a distribution of possible weather scenarios. The direct implementation of these models, however, would require that the ANSP exercise complete control over all flights in question, which, as was discussed above, is not appropriate. In this paper we develop and analyze two alternative approaches for incorporating these models in a

CDM-like setting. In both cases, the ANSP allocates certain resources to the flight operators, and the flight operators then optimize the use of the resources they are given. Specifically, the flight operators have the final decision on which flights are rerouted and they also have the flexibility to adjust the times of ground delayed flights based on their overall allocation of slots at the FCA. At the same time, the overall paradigm makes use of the power of the optimization models (either by the ANSP or the flight operator).

II. RELATED RESEARCH

From the perspective of rationing ANSP resources, by far the most attention has been paid to optimizing the assignment of delays in GDPs. The ground holding problem was first analyzed in Odoni (1987), where a deterministic optimization model was defined. Richetta and Odoni (1993) provided a stochastic optimization model that formally treated capacity variability (usually caused by weather). Whereas these two references as well as other early ones implicitly assume that the ANSP had total control over the disposition of all flights considered, Ball et al. (2003) introduced a stochastic integer programming that (only) set a planned airport arrival rate, thereby allowing other processes, e.g. those developed under CDM, to determine how delay is allocated among individual flights.

These results have been extended in a number of directions. Mukherjee and Hansen (2007) provide a general stochastic, dynamic model, allowing for real-time consideration of the evolution of the stochastic capacity profile. Other recent work has provided simple rationing methods that explicitly take into account existing CDM procedures, e.g. Ball and Lulli (2004) and Ball et al. (2010). A common theme is that the basic paradigm for fairness in rationing is first-scheduled, first-served, also known as ration-by-schedule (RBS).

The more challenging problem, and the one that is the focus of this paper, is rationing of en route airspace resources. A major new feature in this setting is that re-routing flights around the congested resource becomes a possible control option, in addition to assigning ground delay. A body of research has led to the development of large-scale deterministic integer programming models, e.g. Bertsimas and Stock Patterson (1998, 2000). Mukherjee and Hansen (2009) provide a general stochastic programming model for these problems. Nilim et al (2001) also model problem dynamics, albeit in a somewhat stylized setting.

Recent work of Ganji et al. (2009a, 2009b) develops models that explicitly take into consideration the problem addressed by AFP's, which focus on a capacity reduction associated with an FCA. In Ganji et al. (2009b), the basic framework is introduced, whereby every flight anticipated to use the FCA under consideration is assigned a nominal disposition under the GDP, which could be either following the original flight plan (with or without some ground delay), or taking a re-route around the FCA. The weather uncertainty is captured in a scenario tree of possible time-dependent capacity profiles, and the two-stage stochastic program produces both an optimal initial assignment of flight dispositions, as well as recourse

decisions to be taken once the specific capacity scenario is realized. Ganji et al. (2009a) clarifies that this approach is conservative in the sense that even the nominal assignment has to satisfy the capacity constraints at the FCA, despite some expectation that the weather event could clear by the time flights reach its boundary. An alternative, therefore, is to allow those constraints to be violated in the nominal (first stage) program, hedging against different possible capacity outcomes, and offsetting the burden of ameliorating those constraint violations to the individual recourse solutions in the second stage.

As mentioned in Section 1, the main hindrance to implementing such a solution is that it implicitly assumes the ANSP has control over the actions of all flights considered. This paper seeks to make the model compatible with current CDM practices.

To do so, a number of building blocks are assembled into two primary solution mechanisms. The fundamental building block is the model of Ganji et al. (2009b), with the additional strengthening of the IP formulation of that model developed by Glover (2010) (throughout the paper we call this the GG model). The primary change to these models is that we relax the assignment of slots to flights that characterized these models originally, and replace it with an assignment of slots to air carriers. The flights within any individual carrier's bundle can be thought of as fungible from the perspective of the ANSP, but the carrier may have internal value propositions for flights that invite internal optimization. The basic fairness mechanism in these models is still RBS. As explained in the next section, one version of the new model enforces strict feasibility constraints on the carriers' assignments of flights to slots. The interaction between the ANSP and a single carrier can be thought of as a one-way allocation of a bundle of slots, with which the carrier carries out an internal optimization (this is functionally equivalent to the CDM cancellation and substitution process). In the second variant, an inter-carrier slot exchange mediated by the ANSP is carried out (this is functionally equivalent to the CDM compression algorithm). In our experiments, we employ an optimization-based version of compression described in Vossen and Ball (2006b). There are various subtleties related to the relative benefits and practicality of these two approaches, which we explore in the paper.

III. PROBLEM FRAMEWORK

Both the research we build on, and the current research, employ the structures used in AFP planning today. Specifically, an FCA (volume of airspace) is designated, together with a time window of reduced capacity. That time window is broken down into a sequence of slots and a capacity for each slot is designated. These capacities effectively meter the traffic entering the FCA at its boundary. Thus, if there were 30 slots defined over a one hour time window with a capacity of one flight per slot, then one flight would be allowed to enter the FCA every two minutes. Clearly, this capacity model is a simplification, because the true capacity of a volume of airspace in general depends on a much more

complex set of considerations. This simple model reflects the fact that AFP planning models represent a more or less direct carry-over of the models used for GDP planning. At the same time, AFP's, even with these relatively rudimentary capacity models, have proved to be very effective tools for traffic flow management in the U.S.

Using the capacity models just described, AFP planning can be carried out in a way very analogous to GDP planning. That is, whereas a GDP plan consists of an assignment of flights to arrival slots at an airport, an AFP plan consists of an assignment of flights to arrival slots at the FCA. The principal initial mechanism for making such assignments for GDP is the ration-by-schedule (RBS) algorithm, which prioritizes flights based on their scheduled time of arrival. At first the application of this mechanism to AFP's might seem problematic since no scheduled arrival time exists for an FCA. However, this can be overcome by translating the scheduled arrival time at the destination airport to a scheduled arrival time at the FCA using the filed flight plan. In fact, current decision support tools for AFP planning use this approach.

The optimization models that this paper builds upon use this same capacity model. That is, the options considered for each flight are to assign it to an available FCA slot or to avoid the FCA by routing around it. Each possible slot assignment has an associated cost (the cost of the associated required ground delay) and the reroute option also has a cost in excess flight time. The models consider all of these costs and find a solution that minimizes total system cost.

We now begin to describe our two proposed mechanisms for combining the optimization approaches with CDM capabilities. As a first step, both approaches employ RBS to determine an initial slot-to-flight assignment. Our two mechanisms differ in exactly how this initial assignment is later used and in the information the flight operators are required to provide. As is also done with GDP planning, this slot-to-flight assignment is not binding from the flight perspective, but is really just a mechanism for determining an assignment of slots to flight operators, $SLOTS(RBS,a)$. In the GDP case, flight operators, given a set of slots, then solve the problem of determining which flights to swap or cancel in order to maximize their internal objective functions. In the FCA case, the operators decide which flights should remain on their primary routes ($FLIGHTS'_a$) and take a ground delay, and which should be rerouted around the FCA ($FLIGHTS_a - FLIGHTS'_a$). The primary purpose of the Ganji et al. (2009) model was to solve this latter problem optimally, albeit from the perspective of a benevolent monopoly ANSP. In a real situation, cancellations are also possible reactions to AFP's, although we do not model those here, since these decisions, in practice, depend strongly on real-time load factors, connection opportunities, and other passenger information not representable in a planning model.

In this paper, we relax the implicit assumption that the optimization is performed globally by the ANSP, and instead model the decisions that might be made by each independent carrier, in a CDM-enabled environment. Each flight operator solves its own problem using the resources (slots) allocated to

it by the ANSP. We introduce two possible mechanisms for interaction between the ANSP and the operators, called MECH1 and MECH2. The key difference between the two mechanisms is that in MECH1, each operator must assign specific flights to each of its slots, while MECH2 requires only that the flight operator provide a prioritized list of flights. Following is a description of each of the mechanisms:

MECH1:
1: ANSP: execute RBS; assign a set of slots, $SLOTS(RBS,a)$ to each flight operator a.
2: for each flight operator a: considering the available slots $SLOTS(RBS,a)$, determine the subset of flights to be left on their primary routes, $FLIGHTS'_a$, and the subset of flights to be rerouted, $FLIGHTS_a - FLIGHTS'_a$.
Assign an available slot to each flight in $FLIGHTS'_a$; deliver this assignment $FLIGHTS'_a \rightarrow SLOTS(RBS,a)$ to the ANSP.

MECH2:
1: ANSP: execute RBS; assign a set of slots, $SLOTS(RBS,a)$ to each flight operator a.
2: for each flight operator a: considering the available slots $SLOTS(RBS,a)$, determine the subset of flights to be left on their primary routes, $FLIGHTS'_a$, and the subset of flights to be rerouted, $FLIGHTS_a - FLIGHTS'_a$.
Determine a priority order for $FLIGHTS'_a$: $ORD(FLIGHTS'_a)$ and deliver that list to the ANSP.
3: ANSP: Determine final reassignment of flights to slots based on flight operator priorities.

As discussed above, step 2 in both procedures can employ the recent research on stochastic optimization models we build upon. However, there are subtle differences in these flight-operator-specific models when compared to the earlier models. First, in the earlier work, when the model had "control" over all available capacity, it could optimize the reassignment of rerouted flights to their primary routes in the event of a weather clearance. In the present context, when each flight operator solves this problem separately, it is possible that the models will propose to "overuse" the newly available FCA capacity. This is not a problem in the practical implementation of the models in the sense that the ANSP operational systems would only approve such reassignments when feasible. However, it does imply that the individual flight operator models would be using an approximate objective function, essentially making their best estimate of the ANSP behavior in the future. The second contrast with the earlier models can be seen in the differences between MECH1 and MECH2. The problem to be solved for step 2 of MECH1 specifically assigns flights to slots available to that flight operator. Under MECH2, however, such a specific assignment is not required. In fact, the approach we use to solve this problem does "associate" flights with slots, however, this association implies only that

the flight in question should be assigned to a slot no earlier than the associated slot.

We should also provide some comments on step 3 of MECH2. The key difference in the results passed from step 2 of MECH1 or MECH2 back to the ANSP is that it may not be feasible to assign the prioritized list of flights (from MECH2) provided by each flight operator to the slots it "owns". That is, some flight might have an earliest arrival time later than the slot with which it is associated. This is precisely the challenge that the compression algorithm used for GDP planning was designed to address. In this paper, we apply the optimization-based version of compression proposed by Vossen and Ball (2006b) to solve this problem. This version is able to simultaneously compress several slots and has certain advantages over the iterative execution of compression steps.

We now provide an example comparing MECH1 and MECH2.

Example 1:
The following two tables give an airline's flight list and the slots allocated to it. Here *arr* denotes the earliest arrival time at the FCA.

FLIGHTS$_a$:

flight	*arr*
f1	1600
f2	1610
f3	1620
f4	1650
f5	1710

SLOTS$_a$:

Slot	time
S2	1610
S3	1620
S4	1630
S9	1720
S12	1750

Note that MECH1 and MECH2 differ in the input transferred from an airline to the ANSP at the end of Step 2. Below we give examples of these two inputs.

Sample Airline Inputs:

MECH1 Input:

flight	slot
f1	*s2*
f2	*s3*
f3	*s4*
f4	*s9*
X	*s12*

MECH2 Input:

flight
f1
f2
f4
f5

MECH 2 Implicit Assignment:

flight	slot
f1	*s2*
f2	*s3*
f4	*s4*
f5	*s9*
X	s12

Thus, for the MECH1, *FLIGHTS'$_a$* = {f1, f2, f3, f4} and for the MECH2, *FLIGHTS'$_a$* = {f1, f2, f4, f5} so that the airline decides to reroute *f5* under MECH1 and *f3* under MECH2. Under MECH1, this input would be justified if the cost of rerouting *f5* is less than the cost of the ground delay it would incur if assigned to slot *s12*. The change under MECH2 could be economically justified in the case where *f4* is a high-value flight. The closest slot to *f4* later than its *arr* is *s9* so under MECH1 this is the best slot it can be assigned to. However, under MECH2, the airline decides to reroute *f3* and assign *f4* to *s4*. While *f4* cannot use *s4*, under MECH2, this assignment

is allowed since a compression step will be executed to give *f4* the best available slot after *s4* that is compatible with its *arr* (1650).

IV. MODELS

A. Review of GG Model

Suppose that an FCA has been identified, and that an AFP has been declared by the ANSP that encapsulates the following details:

    a. The geographic boundary of the FCA
    b. The start time of the FCA, and the set of possible termination times *T*
    c. The probabilities associated with each of the possible termination times $\{p_t\}, t \in T$
    d. The value of the capacity reduction in place during the AFP
    e. The set of flights *FLIGHTS$_a$* whose flight plans would have them present in area (a.) during the broadest interval present in (b.), where the subscript *a* denotes ownership by an individual flight operator.

The RBS algorithm would take the capacity (d.) and convert that into an equivalent set of slots and slot capacities through the FCA. Next, for the flights in (e.), the scheduled departure times are translated into scheduled arrival times at the FCA. RBS then assigns flights to slots based on that chronological ordering. To complete step 1 of either MECH1 or MECH2, the RBS algorithm then identifies, for each flight operator *a*, the set of slots *SLOTS$_1$* to which flights from operator *a* had been assigned. Once the slots are identified, the temporary mapping from flights to slots is forgotten. Each operator *a* now has a set of flights *FLIGHTS$_a$* and a set of slots *SLOTS$_1$* with which to develop their internally optimal assignment.

The notation and formulation for the operator optimization problem is based on the two-stage stochastic integer program in Ganji et al. (2009b) and Glover (2010). For every flight $f \in FLIGHTS_a$, the operator *a* wants to determine a nominal disposition represented by the binary variables $\{x_{f,i}^p\}$ and $\{x_f^s\}$, where $x_{f,i}^p = 1$ if flight *f* from that operator is assigned to its primary (intended) path through the FCA at time slot $i \in SLOTS_1$, and 0 otherwise, and $x_f^s = 1$ if flight *f* is assigned to a secondary route that detours the flight around the FCA, and 0 otherwise.

The following constraints are enforced for each operator. A flight can only be assigned to either its primary or secondary path, and if to its primary path, at exactly one slot time:

$$\sum_{\substack{i \in SLOTS_1 \\ time(i) \geq dep(f) + en(f)}} x_{f,i}^p + x_f^s = 1 \quad \forall f \in FLIGHTS_a \qquad (1)$$

Here the scheduled departure time of flight *f* is denoted $dep(f)$, its origin-FCA enroute time by $en(f)$, and the clock time associated with slot *i* in the reduced capacity

situation by $time(i)$. A slot capacity constraint must be enforced:

$$\sum_{\substack{f \in FLIGHTS_a \\ time(i) \geq dep(f)+en(f)}} x^p_{f,i} \leq cap_1(i) \quad \forall i \in SLOTS_1 \qquad (2)$$

where $cap_1(i)$ is the ANSP-provided AFP capacity for slot i allotted to operator a. We note here that the sum of $cap_1(i)$ does not necessarily exceed or equal the number of flights in $FLIGHTS_a$; the ANSP allocation to the operator may necessitate a minimum number of reroutes (or, in practice, cancellations).

These variables and constraints represent the first stage of the two-stage program, which is the set of flight dispositions that are intended for the nominal situation when the AFP termination time has not yet been determined. Because we assume that the capacity caps provided by the ANSP to the operators do not exceed the system-wide capacity of the FCA, it is important to notice here that the collection of first-stage decisions made by the carriers is feasible system-wide by design. Thus, in a worst-case scenario, if the weather clears early but the agglomeration of the carriers' proposed second stage decisions violates the good-weather capacity of the FCA, the ANSP can always fall back on the carriers' first-stage decisions as a system feasible (although perhaps not efficient) plan, or they can approve the second-stage propositions piecemeal. It is because of this uncertainty in the ANSP decision process (and the independence afforded to the carriers themselves) that the consideration of second stage resources in the operators' objective functions must be thought of as an estimation of their costs, subject to random decisions by other carriers and by the ANSP.

The second stage variables and constraints represent the recourse decisions proposed by the carriers to be taken when the specific value of the stochastic AFP termination time is realized. Any possible second-stage termination scenario can be represented by its termination time t, a new set of slot capacity values $\{cap_2(i)\}$, and a new set of slots $SLOTS_2$, with slot times $\{time(j,t)\}$ where $time(j,t)$ is the clock time associated with slot j under scenario t. This flexibility in notation allows for a wide variety of different capacity windfalls after the AFP terminates, although the capacity would probably be set to the good-weather capacity of the FCA.

Each carrier can pre-compute some conditional flight data that are necessary for the second stage decision. Specifically, if scenario t were to materialize, then for each flight f, we denote by $earliest(f,i,t)$ the earliest FCA arrival slot that a flight f could be reallocated to in scenario t, had it been assigned a primary path slot of i in the first stage. Depending on the values of i and t, the flight could be a) already en route, in which case $earliest(f,i,t)$ would be $time(i)$, b) not yet departed but also not ground delayed, in which case

$earliest(f,i,t)$ would be $dep(f)+en(f)$, or c) ground delayed, in which case $earliest(f,i,t)$ would be $t+en(f)$.

Importantly, at the time the AFP terminates, some of the flights in the program may already have departed on their secondary routes. Thus, in addition to the previous primary and secondary dispositions, there is also an option in the recourse to revert a flight on its secondary path back to its primary path, which we call a hybrid path. The second stage binary decision variables are $\{y^s_{f,t}\}$ and $\{y^h_{f,i,t}\}$, where $y^s_{f,t}=1$ if flight f from operator a is assigned to its secondary route in scenario t, and 0 otherwise; and $y^h_{f,i,t}=1$ if flight f from operator a is assigned to a hybrid route that uses the FCA at slot i, under scenario t.

For a given scenario t, we initiate the second stage decision-making by building a queue for each second stage slot j amongst the primary FCA slots available in that scenario. We assume that an important factor influencing the decisions of carriers is the size of the aircraft. Thus, we set up a separate queue for each aircraft size in the model. The inputs to the aircraft c queue for slot j are flights f with primary first stage dispositions that are capable of using that slot, flights with secondary first stage dispositions that have not departed on their secondary route, and flights from earlier slot queues that could not be accommodated due to second stage capacity constraints. The outputs from the queue for slot j are the flights that get assigned to that slot, the number of which is denoted $u_{c,j,t}$. We denote by $z_{c,j,t}$ the number of flights that are forwarded from slot $j-1$ to slot j in the aircraft c queue of scenario t. The binary decision variable $s^p_{f,t}=1$ if flight f was assigned to its secondary route in stage one but reassigned to its primary route in scenario t of stage two. The following flow conservation equation holds at the queue:

$$\sum_{(f,i,t)|earliest(f,i,t)=time(j,t)} x^p_{f,i} + \sum_{f|dep(f)+en(f)=time(j,t)} s^p_{f,t} \qquad (3)$$
$$+z_{c,j,t} - z_{c,j+1,t} - u_{c,j,t} = 0 \quad \forall j \in SLOTS_2, t \in T$$

When $j=1$, $z_{j,t} \equiv 0$, and when $j=|SLOTS_2|$, $z_{j+1,t} \equiv 0$.

We must prevent flights that have not yet departed from being assigned to a hybrid route:

$$\sum_{j \in SLOTS_2} y^h_{f,j,t} = 0 \quad \forall f, t \ni t \leq dep(f) \qquad (4)$$

and slights that have already departed on secondary routes cannot be reassigned to primary routes:

$$s^p_{f,t} = 0 \quad \forall f, t \ni t > dep(f) \qquad (5)$$

The operator is also expected to pre-compute the hybrid diversion time $\tau^d_{f,j}$, which is defined as the time at which flight f, if assigned a secondary disposition in stage one, would have to divert via a hybrid route in order to reach the FCA to meet slot j in stage two. Using this information, we can prevent ineligible hybrid routes:

$$y^h_{f,j,t} = 0 \quad \forall f, t, j \ni \tau^d_{f,j} < t \qquad (6)$$

All flights that had secondary assignments in stage one must be told in stage two either to stay with their secondary assignments, to use a hybrid route, or to get re-assigned back to their primary route:

$$\sum_{j \in SLOTS_2} y^h_{f,j,t} + y^s_{f,t} + s^p_{f,t} = x^s_f \quad \forall f, t \qquad (7)$$

The stage two FCA capacities must be respected:

$$u_{j,t} + \sum_{f \in FLIGHTS_a} y^h_{f,j,t} \leq cap_2(j) \quad \forall j, t \qquad (8)$$

Finally, because stage two is supposed to represent a capacity increase, it would be politically unpalatable to give any flight a worse disposition in stage two than it had in stage one, even if the system objective was better met by doing so. Thus, we require each flight to depart the system in stage two do so no later than it would have done so in stage one:

$$\sum_{j=earliest(f,i,t)}^{time(i)} u_{j,t} \geq \sum_{\substack{f' \in FLIGHTS_a \\ earliest(f',i,t)=earliest(f,i,t)}} x^p_{f',i} \quad \forall f, i, t \qquad (9)$$

The structural constraints on the decision variables are:

$$x^p_{f,i}, x^s_f, y^h_{f,j,t}, y^s_{f,t}, s^p_{f,t} \in \{0,1\} \quad \forall f, i, j, t \qquad (10)$$

$$u_{j,t}, z_{j,t} \geq 0 \text{ and } u_{j,t}, z_{j,t} \in \mathbf{Z} \qquad (11)$$

B. Compression Using the OPTIFLOW Model

As discussed in Section 3, we execute a compression step in MECH2 using the OPTIFLOW model. Here we explain this approach. First we illustrate the basic concept behind compression and then give the model formulation.

Example 2: Compression

RBS Assignment: Compression Reassignment after cancellation or rerouting of A-f3:

flight	arr	slot time
C-f1	1555	1600
A-f1	1600	1610
A-f2	1610	1620
A-f3	1620	1630
C-f2	1625	1640
B-f1	1630	1650
B-f2	1635	1700
C-f3	1640	1710
A-f4	1650	1720

flight	Arr	slot time
C-f1	1555	1600
A-f1	1600	1610
A-f2	1610	1620
C-f2	**1625**	**1630**
B-f1	**1630**	**1640**
A-f4	**1650**	**1650**
B-f2	1635	1700
C-f3	1640	1710
		1720

The RBS assignment given is based on the assumption is *arr* is the scheduled arrival time. We assume that airline *A* cancels or reroutes flight *A-f3*. Since it "owns" slot 1630 it could sub another flight into this slot but since it does not have a flight that can be placed there, compression is executed. Flights *C-f2* and *B-f1* are moved up, freeing up slot 1650, which is late enough to accommodate *A-f4*. Note that this example is consistent with Example 1 with information on other airlines added (the airline in Example 1 is airline A in this example.

The OPTIFLOW approach to compression employs a simple assignment model with an appropriately defined objective function. The input data include a set of flights, *FLIGHTS* and a set of slots, *SLOTS*. Each flight *f* has an earliest arrival time, *arr(f)* and a goal slot *g(f)*; each slot *i* has a slot time *time(i)* and a capacity *cap(i)*. Recall from Section 3, that, under MECH2, an ordered list of flights was input by each flight operator and this was used to associate a slot assigned to that flight operator with each flight. The goal slot *g(f)* is the slot associated with flight *f*. Thus, in Example 1, for the application of MECH2, the goals for flights *f1, f2, f4, f5* would be *s2, s3, s4, s9* respectively. We employ the variable set: $w_{fi} = 1$ if flight *f* is assigned to slot *i*; 0 otherwise. The OPTIFLOW model is defined by:

$$Min \sum_{f \in FLIGHTS} \sum_{i \in SLOTS} cst(f,i) w_{f,i}$$

$$s.t. \sum_{f:earliest(f) \leq time(i)} w_{f,i} \leq cap(i) \qquad \forall i$$

$$\sum_{i:time(i) \geq earliest(f)} w_{f,i} = 1 \qquad \forall f \qquad (12)$$

$$w_{t,t} \in \{0,1\} \qquad \forall f, i$$

This is a simple assignment model, however, the subtlety comes in the definition of the objective function coefficients, *cst(f,i)*. The goals serve as attraction points with the cost of deviation from the goal increasing at a greater than linear rate. Thus, assuming a flight is always assigned a slot with a time later than its goal, the cost function for assigning *f* to *i* would be: $(time(i) - time(goal(f)))^{1+e}$ where ε is a small positive constant. This objective function must be modified slightly to allow for the assignment of flights to slots earlier than their goals. To do this, the maximum possible (early) deviation from a goal is computed:

$$md = Max_f \{time(goal(f)) - earliest(f)\} \qquad (13)$$

This can then be used to define the cost function:

$$cst(f,i) = (time(i) - (time(goal(f)) + md))^{1+e} \qquad (14)$$

Here we implicitly assume $time(i) \geq time(goal(f)) + md$; if this is not the case appropriate modifications can be made. Vossen and Ball (2006b) show that the OPTIFLOW model with this cost function provides the basic compression functionally with the added flexibility of being able to vary goal assignments. Thus, this model is quite appropriate for use in MECH2.

C. Modes of Operation of GG Model

In our experiments the GG model is used in three different modes. Its description in Section IV-A focuses on its use in MECH1. There is a modification in the cost function for MECH2. In addition, we use it in its original form to compute a system-wide optimal solution. These differences are described below.

1) System-Wide Optimal

As discussed earlier the prior descriptions of the GG model implicitly assumed that the ANSP or a single monopoly airline had total control over all flights. In our experiments, we apply the model in this mode. However, we do not compute this solution to imply that it would actually be used, but rather the cost of this solution represents a system-wide minimum that can be used as a basis for comparison. Specifically, it is common in the economics literature to compare the total value or cost achieved by a market mechanism or distributed control mechanism against a social- or system-optimal value or cost. This solution represents such a system optimal and so we can a measure of the quality of our proposed CDM approaches (MECH1 and MECH2) relative to how close they come to this system optimal. In this case, the set of flights input to the model is the set of all flights and the capacities are not rationed in any way; they are estimates of the true capacities.

2) Mechanism 1 vs Mechanism 2

The model as described applies to its use under MECH1. MECH2 requires a subtle change. Specifically, under MECH2, the final slot to which a flight is assigned is determined by the OPTIFLOW model. Thus, it is possible for a flight to be associated with a slot to which it cannot be assigned. For example, in Example 1, flight *f4* was assigned to slot *s4* even though *time(s4) < arr(f4)*. To model this possibility we allow a flight *f* to be assigned to a slot earlier that *arr(f)* but in such cases assign the cost to be the cost of using *arr(f)*, i.e. 0. The second stage costs are modified in a similar way.

V. RESULTS

Our computational experiment was set up to compare these two alternative approaches. For comparison, the system-wide optimal solution is also produced, although as stated earlier, this solution is not practical. The input to the model consisted of a data set of 400 flights and an FCA of 5 hours in duration. During the associated AFP, the FCA had a reduced capacity of one arrival per four minutes. The 5 hour duration was the maximum time of reduced capacity as the FCA was restored to its nominal capacity and the AFP terminated at a random time. Upon restoration, the capacity of the FCA increased to one arrival per minute. There were five possible cancellation times for the AFP, coinciding with each hour of duration, each occurring with probability 0.2.

Since the objective of the mechanisms we tested was to allow airlines to internally optimize their individual costs functions, we simulated a set of airlines including their fleet and flight cost data. We considered a set of 400 flights and randomly assigned each flight a carrier, and an aircraft type. We assumed seven different carriers and four different aircraft types (ERJ-170, 737-300, 757-200, 767-400). The assumed

aircraft passenger counts were: 70, 128, 205, and 245 passengers respectively. The distribution of aircraft sizes to carriers was as described in Table 1.

	ERJ-170	737-300	757-200	767-400
Carrier 1	64	64	16	16
Carrier 2	30	12	6	12
Carrier 3	0	70	0	0
Carrier 4	20	8	0	12
Carrier 5	0	10	10	0
Carrier 6	20	0	0	0
Carrier 7	12	9	9	0

Table 1: Distribution of aircraft sizes to carriers.

Our airline cost models used as a starting point the model presented in Vakili and Ball (2009), which is based on ATA data and models from Metron Aviation. The aircraft operatin cost per minute of airborne delay is \$64 and for ground delay \$32. The cost per minute of delay per passenger is \$0.6. It is assumed that for $1/6^{th}$ of this cost in making economic tradeoffs. Thus, the cost per minute of ground delay and airborne delay is $(32+0.1Psn(f))$ and $(64+0.1Psn(f))$ respectively, where $Psn(f)$ represents the number of passengers on the flight *f*.

The expected airborne delay for a flight *f* is the amount the flight is expected to be deterred by taking a secondary route, *cs(f)*. This value can be reduced by the time saved by this flight being rerouted through the FCA on a hybrid route. The amount saved is *sv(f, j, t)* if the flight *f* is rerouted to the slot *j* in scenario *t*.

If the flight *f* arrives at the FCA slot *j* in scenario *t* on its primary route, then *x(f, i) = 1* for some *i* such that $arr(f) \leq time(i)$ and *u(c, j, t) = 1*, where *Psn(f) = Size(c)*. The ground delay that this flight has served is *(time(j) – arr(f))*, where *arr(f) = dep(f) + en(f)*. This can be represented in the objective function placing *(– arr(f))* as the ground delay coefficient on each *x(f, i)* variable, where $arr(f) \leq time(i)$. Similarly *time(j)* can be placed as the ground delay coefficient on each *u(c, j, t)* variable. Similar considerations can be made for the $s_{k,t}^{p}$ variables. The ANSP can then be seen to have a system-wide objective function of:

$$OBJ_{ANSP} = \sum_{(f,i)|arr(f)\leq time(i)} (0.1Psn(f)+32)(-arr(f))x(f,i)+\sum_{f}(0.1Psn(f)+64)x_{f}^{s}$$
$$\sum_{(f,t)}p(t)\left[(Psn(f)\cdot0.1+32)(-arr(f))-(Psn(f)\cdot0.1+64)cs(f)\right]s_{f,t}^{p}+$$
$$\sum_{\substack{j\in SLOTS_{2}\\t\in T\\c\in Aircraft}}(0.1Size(c)+32)p(t)time(j)\cdot u_{c,j,t}+\sum_{(f,j,t)|arr(f)\leq time(j)}(0.1Psn(f)+64)y_{f,j,t}^{h}$$

Each carrier has a slight modification of this objective function, where they are only concerned with minimizing the costs to their individual flights. This can be stated as

$$OBJ_{CARR(a)} = \sum_{\substack{(f,i)|arr(f)\leq time(i)\\ carrier(f)=a}} \left(0.1Psn(f)+32\right)\left(-arr(f)\right)x(f,i)+ \sum_{f|carrier(f)=a} \left(0.1Psn(f)+64\right)x_f^s$$

$$\sum_{(f,t)|carrier(f)=a} p(t)\left[\left(Psn(f)\cdot0.1+32\right)\left(-arr(f)\right)-\left(Psn(f)\cdot0.1+64\right)cs(f)\right]s_{f,t}^p+$$

$$\sum_{\substack{j\in SLOTS_2\\ t\in T\\ c\in Aircraft}} \left(0.1Size(c)+32\right)p(t)time(j)\cdot u_{c,j,t}+ \sum_{\substack{(f,j,t)|arr(f)\leq time(j)\\ carrier(f)=a}} \left(0.1Psn(f)+64\right)y_{f,j,t}^h$$

As we described in Section IV-C, the OPTIFLOW approach to compression gives cost coefficients $cst(f,i)$ for the $x(k, i)$ variables that can be used as an inter-airline exchange of slots that are unusable by the airlines themselves.

Each objective function provided a different feasible solution to stage one. To compare these solutions, we used the model to determine the optimal stage two solution for each of these feasible solutions to stage one. This provided us with the total expected cost of each of these solutions. Table 2 measures the total expected cost of MECH1, MECH2 in comparison to the system optimal cost over several trial runs. In each run, the flights were randomly assigned to carriers and given one of the four sizes.
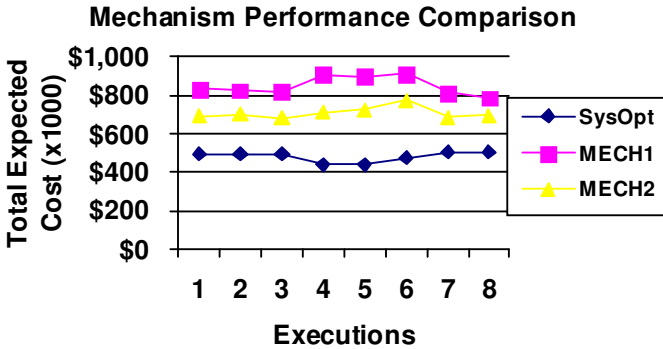
Mechanism Performance Comparison

Table 2: Total Expected Costs of MECH1 and MECH2 compared to the system optimal cost.

We note both MECH1 and MECH2 have costs substantially higher than system optimal. This difference is commonly referred to as the "price of anarchy". That is by allowing a distributed or market-based mechanism to allocate resources, there is a penalty paid over the social or system optimal. It is noteworthy that this cost penalty is quite substantial: the costs for MECH1 are generally around twice the system optimal. Clearly MECH2 provides a very significant improvement – thus, it has distinct advantages over MECH1.

MECH1 and MECH2 have other important differences. The main advantage of MECH1 is that it provides an airline with very specific control over its slots and the allocation of flights to slots. This generally would give MECH1 a higher rating relative to measures of predictability. On the other hand, the airline input under MECH2 is much simpler than under MECH1. Specifically, under MECH2, the airlines only need to provide a prioritized list of flights. Under MECH1 the airlines must provide an explicit mapping of flights to slots. This requires the airlines to know the time of arrival of its flights at

the FCA boundary (which is not necessarily a task readily supported by their TFM tools). Also, as aspects of the FCA dynamically change slot times and FCA capacity could change. MECH1 would require that the airlines dynamically update slot assignments as appropriate. On the other hand, under MECH2, the simple prioritized list could serve as a robust statement of an airline's priority that is effective under changing conditions. Given the cost advantage of MECH2 and the advantage just described it certainly would seem that it would be the preferred option under most circumstances.

VI. CONCLUSIONS

In this paper, we have described two approaches to embedding the GG model within a CATM environment. The role of the ANSP has now been reduced from having complete control over all decisions regarding flights to one that consists of allocating of a set of slots to each airline and allowing and the flight operators to then optimize the use of the resources they are given. MECH1 gives the carrier the final decision on which flights to reroute and the amount of delay for their flights. MECH2 includes an additional step where the ANSP helps reduce system-wide costs by performing a step similar to compression. Our experimental results this provides a substantial advantage.

REFERENCES

1. Ball, M.O., R. L. Hoffman, D. Knorr, J. Wetherly and M. Wambsganss, 2001. Assessing the benefits of collaborative decision making in air traffic management. In: *Air Transportation Systems Engineering*, G. Donohue and A. Zellweger (eds.), American Institute of Aeronautics and Astronautics, Inc., Reston, VA., 239-250 (a preprint appeared in *Proceedings of the 3rd USA/Europe Air Traffic Management R&D Seminar*, 2000).
2. Ball, M.O., R. Hoffman and A. Mukherjee, 2010. Ground delay program planning under uncertainty based on the ration-by-distance principle. *Transportation Science,* **44**, 1-14.
3. Ball, M.O., R. Hoffman, A. Odoni, and R. Rifkin, 2003. A stochastic integer program with dual network structure and its application to the ground-holding problem. *Operations Research*, **51**, 167-171.
4. Ball, M. O., and G. Lulli, 2004. Ground delay programs: optimizing over the included flight set based on distance. *Air Traffic Control Quarterly*, **12**, 1-25.
5. Bertsimas, D. and S. Stock Paterson, 1998. The air traffic flow management problem with en route capacities. *Operations Research,* **46**, 406-422.
6. Bertsimas, D. and S. Stock Paterson, 2000. The traffic flow management rerouting problem in air traffic control: a dynamic network flow approach. *Transportation Science,* **34**, 239-255.
7. Eurocontrol, 2010. *Airport CDM Implementation Manua,l vers 3.1*, available at: http://www.euro-cdm.org/
8. Ganji, M., M. Ball, M. and D. Lovell, 2009a. Resource allocation in flow-constrained areas with stochastic termination times: Optimistic approch. In: *Proceedings of the 8th USA/Europe Air Traffic Management R&D Seminar*.

9. Ganji, M., D. J. Lovell, M. O. Ball, and A. Nguyen, 2009b. Resource allocation in flow-constrained areas with stochastic termination times. *Transportation Research Record: Journal of the Transportation Research Board*, **2106**, 90-99.

10. Glover, C., 2010. Computationally tractable integer programming models for air traffic flow management. Ph.D. Dissertation, University of Maryland, College Park.

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

12. Mukherjee, A., and M. Hansen, 2009. A dynamic rerouting model for the air traffic flow management. *Transportation Research Part B: Methodological*, **43(1)** 159-171.

13. Nilim, A., L. El-Ghaoui, V. Duong, and M. Hansen, 2001. Trajectory-based air traffic management (TB-ATM) under weather uncertainty. In: *Proceedings of the 4th USA/Europe Air Traffic Management R&D Seminar*.

14. Odoni, A. R., 1987. The flow management problem in air traffic control. In: *Flow Control of Congested Networks*, A.R. Odoni, L. Bianco, and G. Szego (eds.), Springer Verlag, New York, 269-288.

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

16. Vossen, T. and Ball, M.O., 2006a. Slot trading opportunities in collaborative ground delay programs. *Transportation Science*, **40**, 29 - 43.

17. Vossen, T. and Ball, M.O., 2006b. Optimization and mediated bartering models for ground delay programs. *Naval Research Logistics*, **53**, 75 – 90.

18. Vossen, T., Ball, M.O., Hoffman, R. and Wambsganss, M., 2003. A general approach to equity in traffic flow management and its application to mitigating exemption bias in ground delay programs. *Air Traffic Control Quarterly*, **11**, 277-292. (a preprint was published in *Proceedings of 5th USA/Europe Air Traffic Management R&D Seminar*, 2003).

19. Wambsganns, 1996, M. Collaborative decision making through dynamic information transfer, *Air Traffic Control Quarterly,* **4**, 107-123.

AUTHOR BIOGRAPHY

Michael Ball is the Orkand Corporation Professor of Management Science in the Robert H. Smith School of Business at the University of Maryland. He also holds a joint appointment within the Institute for Systems Research in the Clark School of Engineering. Dr. Ball received his Ph.D. in Operations Research in 1977 from Cornell University. He is co-Director of NEXTOR, the National Center of Excellence for Aviation Operations Research, and he leads the NEXTOR Collaborative Decision Making project.

Charles Glover is a post-doctoral associate with the National Center of Excellence for Aviation Operations Research (NEXTOR) at the University of Maryland. Dr. Glover received his Ph.D. in Applied Mathematics in 2010 from the University of Maryland.

David J. Lovell is an Associate Professor in the Department of Civil and Environmental Engineering at the University of Maryland. He holds a joint appointment with the Institute for Systems Research and an affiliate appointment with the Applied Mathematics, Statistics, and Scientific Computing Program. Dr. Lovell received his B.A. degree in Mathematics from Portland State University, and his M.S. and Ph.D. degrees in Civil Engineering from the University of California, Berkeley.