Dynamic Collaborative Decision-Making in Air Traffic Flow Management

William D. Hall *
Operations Research Center
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

This paper discusses methods for distributing airport arrival capacity among the users of the National Airspace System (NAS). It is assumed that the users understand their own objectives and can communicate them to the FAA, that the FAA is responsible for distributing the arrival capacity among the users, and that the FAA has the objective of putting the capacity to the best possible use; that is, the FAA seeks to maximize the sum of objective values achieved by the users through the use of arrival capacity. The value that a specific slot of landing capacity contributes to any given flight depends in part on some random quantities that are unknown in advance. Observations of the system may reduce the uncertainty in the users' estimates of these quantities as time passes.

1 Introduction

Recently, the air transportation community has recognized that improved communication, navigation, surveillance, and computer technologies have created new opportunities for improving the air transportation system. The current design of the air transportation system is based on technology that was developed over fifty years ago. Improvements in technology since then have made it possible for the users of the NAS (referred to simply as *users* by the community) and the FAA to collect and share much more information than ever before. For example, it is now possible for aircraft to communicate their precise positions accurately enough to avoid each other even when conditions do not permit either crew to detect the other

*Research supported by a Draper Fellowship

aircraft visually. Technology has improved to the point where aircraft can be made to land without human intervention, relying solely on GPS-based navigation [PCL+95] [CLP+95]. The technology is available for the users and FAA to share and process large amounts of demand, schedule, weather, and trajectory information. The implications of these changes for all aspects of air traffic management are vast.

A large number of organizations, including the FAA, NASA, and the airlines, have recognized the potential benefits to an overhaul of the air transportation system. These groups have begun work on a number of improvement initiatives, known collectively as the "Free Flight" program, or simply Free Flight. The various initiatives of Free Flight share several characteristics. One common characteristic is that each initiative involves the potential to improve the performance of the air transportation system by taking advantage of technological advances. Another common characteristic is that each initiative involves the decentralization of decision-making, giving the users of the system more control over the process.

Decentralization has the potential to improve the system performance because the users of the system have better knowledge of the system objectives and economic motivation to improve the system performance. For example, the crew of an aircraft has more detailed information about the costs and benefits of flying at various speeds than does the FAA, and the users have more information about the costs, benefits, and logistical down-line effects of delays to any of their aircraft than does the FAA. Decentralization allows the various users of the system to account for details of the problems of running the air transportation system that cannot be communicated practically to the FAA.

However, decentralization poses the problem of coordinating the desires of the various users of the system, each of which is economically motivated to improve the value of the portion of the service serving it, even if it is at the expense of the overall performance of the system. Any decentralization effort that is to achieve the full potential system performance must cause the optimal system-wide solution to be optimal for each of the subproblems solved by the users with their own objective functions.

This paper considers the initiative to decentralize the decision-making process in the assignment of arrival slots to flights during periods in which demand for slots exceeds supply. This initiative is currently the primary focus of a group of FAA and airline representatives studying what they call *collaborative decision-making* (CDM). The CDM group's efforts to design a collaborative slot allocation method have yielded a series of proposed refinements to the existing slot allocation method; a refinement to the process is proposed, then its weaknesses are examined and a refinement to the refinement is proposed, and so forth.

The CDM design has yet to be implemented. At present, the assignment of slots to flights is made by the FAA according to a method that does not explicitly account for the objectives of the users and their customers. In fact, the FAA does not have accurate projections of demand for arrival slots, despite the fact that such projections could be produced from information known to the users; as will be discussed further in Section 2.2, the present system of slot-to-flight assignment penalizes a user for providing the necessary information [Wam96].

Partly as a result of this lack of information sharing, potential exists for great improvement in the performance of the Air Transportation System. According to the Investment Analysis and Operations Research group at the FAA (ASD-400), as reported on the CDM in Aviation Transportation web page [Wam97b], "...the CDM elements comprise a potential of $2.6 billion reduction in costs to the airline industry through the year 2004." The web page further asserts that, "based on an established FAA formula, this number becomes $8.9 billion if passenger value of time is taken into account."

This paper is organized as follows. Section 2 describes the arrival slot assignment problem. Section 3 discusses some of the optimization methods that treat the arrival slot assignment problem and discusses the fact that determining valid objective function information remains an obstacle to their use. Section 4 introduces a method of assigning arrival slots to flights that explicitly treats this problem. Section 5 discusses various aspects of the method of Section 4 that could benefit from further study, and discusses the generalized problem of dynamically distributing a publicly-owned resource among multiple interests.

2 Assignment of Slots to Flights

The Free Flight objectives concerning the assignment of slots to flights are to increase the economic value of the air transportation system while maintaining or improving its safety.

For safety to be maintained, an *arrival slot* must be assigned to every aircraft before it is allowed to land. An arrival slot consists of the resources necessary for an aircraft to land: the sole use of a specific runway and approach airspace for a period of time and the attention of the air traffic control officials required to ensure a safe landing. The chief identifying characteristics of an arrival slot are the airport and the time at which the aircraft is to land.

The amount of time required between aircraft landings determines a limit on the number of arrival slots used in a given amount of time (the *arrival rate*). The time required between landings depends on a number of factors, including weather conditions at the time of landing, the type of aircraft, the type of the previously-landing aircraft, and the experience and number of the air traffic control staff on duty.

The weather characteristics having the greatest influence on the arrival rate limit are wind speed, wind direction, precipitation, and visibility. Wind speed and direction determine the set of runways that may be used, often referred to as the *configuration* of the airport. If the configuration of the airport includes runways for which the cross-runway component of the wind is large, landings become more difficult for the pilots to execute, resulting in lower landing rates. If the cross-runway component of the wind reaches a threshold (that can depend on various other conditions such as visibility and precipitation) the runway may not be used. Low visibility conditions reduce the rate at which aircraft can land as well; among other effects they cause, low visibility conditions raise the distance between parallel runways on which aircraft may legally and safely land simultaneously, often reducing the number of runways available for landings. A more thorough treatment of the effects of weather on an airport's capacity is given in [Yu96].

Weather forecasts of visibility, wind speed, and direction often are not accurate to the level of detail required for deterministic prediction of the arrival rate limit far in advance. As a result, arrival slots cannot be assigned to flights far in advance without either running the risk of wasting considerable amounts of capacity or risking large airborne queues of aircraft waiting to land. Wasting capacity is economically undesirable, and large airborne queues of aircraft make it more difficult for the air traffic controllers to ensure safety. Therefore, any practical slot assignment method must be dynamic, that is, its design and implementation must take into account the fact that frequent replanning will be necessary.

The economic factors of the flight-to-arrival slot assignment problem are best understood by the users. Users generally do not allow early departures; doing so would anger many of their customers. The costs to the users of late arrival are complicated functions of many factors, including the amount of delay, the number of passengers delayed, the amount and type of cargo and mail delayed, the number of passengers missing their connections as a result of the delay, the expense of lodging for passengers that must wait overnight because of missing a connection, the amount of additional crew time required, the delay caused to future flights using the same aircraft or crew, and so on [Byr97]. Some of these costs may be direct costs, such as penalties for delivering the mail late, but many costs, such as the cost of damage to the user's reputation, are costs that vary depending on the way the user markets itself and on the demeanor of its passengers. Many of the costs, such as the cost of future delays to outbound flights caused by the delay to equipment and crew of an inbound flight, require logistical analysis to compute.

In short, the objective functions of the users are extremely complicated. Any method that ignores the user objective functions will perform poorly, from the users' perspectives. Each user stands to profit by taking actions that minimize its cost. Whether the actions a user is motivated to take in the slot assignment process tend also to maximize the value to the entire community depends on the manner in which the slot assignment is made.

2.1 Assignment System Used Today

Currently, the assignment of arrival slots to flights is made by the FAA. Before taking off, a flight is assigned an arrival slot. When capacity is adequate to accommodate the scheduled traffic with little or no delay, the flights are assigned slots on a first-scheduled, first-served basis. That is, flights are assigned slots so that they would land in the order they are scheduled to arrive at the destination airport. In practice, the flights land in a different order because the FAA does not enforce the slot ordering; it uses the pre-takeoff slot assignments only to meter the flow of aircraft into the destination airport's airspace.

When landing capacity is inadequate to accommodate the demand at an airport, the FAA temporarily lowers the airport arrival rate by implementing a *ground delay program.* In a ground delay program, the FAA specifies that some of the aircraft bound for the capacitated airport that have not yet taken off must delay their departure by a *ground hold time.* Those aircraft delayed on the ground are then treated as if they had been scheduled to arrive at their original scheduled time plus the ground hold time in the first-scheduled, first-served slot assignment. For several reasons, international flights are not issued ground delays.

Once the assignments have been made, each user is free to cancel flights and to swap slots among its flights, with the restriction that no swap may assign a flight to a slot that would require it to land more than twenty minutes ahead of schedule. Through swapping, the user objectives are accommodated in the solution. This step is important to the users, since the costs to the users of flight delay are very complicated functions of many factors, as mentioned in Section 2.

This assignment system was implemented in February 1994 [Ois97]. Prior to its implementation, the rearrangement of slot assignments by users was severely restricted: swaps were not allowed at all; in order to rearrange its assignment, a user had to cancel a flight first, then it could fill the emptied slot with another flight (a *substitution*), and fill the substituted flight's slot with another flight, and so forth. Each substitution was only allowed if it satisfied a number of criteria, one of which was that it had to advance the arrival time of the substituting flight by a specified number of minutes (the specific number has changed a number of times in the last few years) [Wam97a].

It has been observed that users often do not inform the FAA of flight cancellations [Wam96]. Without flight cancellation information, the FAA estimate of demand for landing slots, upon which the decision to implement ground delay programs is based, is worse than it could be. The FAA-Airline Data Exchange Project (FADE) was begun to address this issue.

2.2 FAA-Airline Data Exchange Project

In the FAA's current assignment method, a user is penalized for reporting flight cancellations in advance. This is a result of the fact that when a user cancels a flight, the flight is removed from the FAA's schedule of arrivals for the destination airport. Subsequent ground delay programs assign landing slots among the users in proportion to the number of flights the user has in the FAA's schedule of arrivals. Thus a user reporting a cancellation before implementation of a ground delay program receives a smaller proportion of the slots than a user with an identical schedule that did not report any cancellations.

The FAA-Airline Data Exchange project (FADE) devised a revision to the FAA's slot assignment method, and called the new method the FADE Compression Algorithm, or simply Compression. Compression is designed to encourage users to reveal cancellations in advance (an airline doing so is said to be *releasing* the slot), and works according to the following steps (slightly modified from [Ois96]).

1. Construct the list of slots released by the airlines, listed in the order in which the slots were released.

2. Construct two lists of controlled flights, one for the releasing airline and one for all other airlines.

3. Try to move flights into the slots on the list of released slots, according to the following steps.

   (a) If there is another slot to be filled on the list of released slots, go to step 3b. If not, go to step 4.

   (b) Go down the list of flights of the releasing airline; if a flight is found that can be moved into the released slot, then do so. If such a flight is found, then this might create a new released slot. If so, repeat step 3b with the new released slot.

   (c) Go down the list of flights of all other airlines; if a flight is found that can be moved into the released slot, then do so. If such a flight is found, and it creates a new released slot, go to step 3b with the new released slot. Otherwise go to step 3a.

4. Return slots that cannot be filled to the releasing airlines.

5. Send out the new slot assignments.

In effect, the compression algorithm keeps track of the airline that reported a slot as being canceled, and gives that airline as much of the benefit of the slot's release as the airline can use. Compression especially encourages users with few flights at a busy airport to report cancellations. For instance, imagine that a small carrier has only two flights landing at SFO on a certain day, one in the morning and one in the afternoon. If the landing rate is reduced at SFO that day, and the delay extends to the afternoon flight, a cancellation of the morning flight will allow a string of substitutions of flights of other carriers until the time at which the afternoon flight is scheduled to arrive; the afternoon flight then receives a slot which allows it to land on time.

The benefit to a user with a large number of flights of releasing its slots is questionable, however. If the user has enough flights to fill any slots it might release, then it has no incentive to release those slots. Of course, it has no disincentive either, since the compression algorithm would return the released slots to the user, unless the airline has determined that by holding the slots it can cause a further ground delay program that would hurt its competitors.

The released slots are processed in the order in which they are released. Therefore, the first slots to be released are most likely to be returned to the airline in the form of slots that the airline can use. As more slots are released, the number of flights that can potentially be moved into the released slot in step 3 is reduced, and it is less likely that the improvement to the releasing airline will be as great.

2.3 Ghost Flight Cancellation Problem

Both the FAA's current assignment method and the FADE Compression method rely on the schedule to determine demand information. If a user schedules flights it never intends to fly, it gains slots in a ground delay program because it is over-represented in the schedule. Under previous, more restrictive FAA assignment methods, the advantage to a user of having slots available to be canceled was not as great. An example illustrates the problem.

Airline	FLTNO	ETA	CTA	Delay
A	101	0700	0700	0
B	102	0705	0710	5
B	103	0705	0720	15
B	104	0710	0730	20
B	105	0710	0740	30
A	106	0710	0750	40
C	107	0720	0800	40
B	108	0740	0810	30
C	109	0740	0820	40
A	110	0830	0830	0
Total				220

Table 1: FADE Delay Example.

2.3.1 Cancellation Example

Consider the example given in [Wam96], shown in Tables 1 and 2. This example was used in [Wam96] to demonstrate how FADE Compression removes the double penalty associated with a user reporting cancellations. It works equally well, however, to demonstrate the incentive a user has to add flights to the schedule that it does not intend to use.

The columns in Table 1 are airline, flight number, scheduled slot time (provided by airlines), first-scheduled first-served (FSFS) slot time, and amount of delay. The FSFS slot time is the slot that would be assigned by the FAA to the flight in the absence of cancellations and substitutions. Notice that airline A incurs 40 minutes of delay, all on flight 7. However, by adding a nonexistent flight to the reported schedule, airline A can effect a change in the allocation to that shown in Table 2.

Table 2 shows the revised allocation of slots to flights after airline A has introduced imaginary flight 100 and then canceled it. The columns, from left to right, are the airline, flight number, scheduled arrival time, initial allocated arrival time, delay incurred, flight canceled / substituted, arrival time allocated after running the FADE algorithm, and resultant delay. Notice that airline A has succeeded in shifting its portion of the delay to its competitors.

This example demonstrates the motivation for a user to schedule and then cancel flights. Much of the focus of the CDM working group is on enforcement means to ensure that the users report accurate schedules. Schedule fabrication can be difficult to detect, however. An

Airline	FLTNO	ETA	CTA	Delay	Can/Sub	New CTA	Rdly
A	100	0700	0700	0	C	-	-
A	101	0700	0710	10	S	0700	0
B	102	0705	0720	15		0720	15
B	103	0705	0730	25		0730	25
B	104	0710	0740	30		0740	30
B	105	0710	0750	40		0750	40
A	106	0710	0800	40	S	0710	0
C	107	0720	0810	50		0800	40
B	108	0740	0820	40		0810	30
C	109	0740	0830	50		0820	40
A	110	0830	0840	10		0830	0
Total			310				220

Table 2: Imaginary Flight Cancellation under FADE.

airline might schedule two flights between the same pair of cities within a relatively short period of time and fly them both on days when demand warranted, canceling one or the other on days when passenger demand was light enough. If the airline seldom flew both flights on the same day, it would clearly be gaining an advantage of the sort enjoyed by airline A in the example above. Yet, it might be argued that the practice of adjusting the set of flights to be used at the last minute to accommodate uncertainties such as passenger demand or to deal with equipment or crew scheduling problems is in everyone's best interest. The airline can serve more passengers that way when demand warrants even if the airline cannot predict the demand, and when demand does not warrant two flights the system is saved the expense of serving the extra flight. For instance, imagine that Logan Airport in Boston has a low arrival rate due to foggy conditions, and that an airline with two lightly-loaded flights destined for Boston during the morning rush hour could consolidate them. Doing so would help to alleviate the capacity problem at Logan.

The task of determining whether this sort of operation is legitimate could keep an army of regulators employed, and would inevitably involve a somewhat arbitrary decision that would inflict a cost on the overall system. The users would be compelled by their bottom lines to exploit any loopholes in the regulation, and in doing so could only hurt the airline industry as a whole.

3 Optimization-Based Approaches

An attractive approach to solving complex logistical problems such as the slot assignment problem is to use optimization-based methods. Optimization-based methods have the potential to make as good an assignment as is possible, where quality of the assignment is measured by an objective function. Optimization-based methods frequently find solutions that are both intuitive and far better than any other solution found by traditional methods for problems in a wide range of application areas, from telecommunications to airline fleet scheduling.

Optimization-based approaches to a wide variety of problems faced by the airlines have had a huge impact on the 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 yield management software [Sab]. Often simply by changing a few decisions, the profitability of a logistically complicated business can be improved dramatically.

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 others, each solve different aspects of the ground holding problem. The model presented in [Ber94] addresses the ground holding problem combined with the routing problem.

None of these solutions has been used in practice, however. As with all optimization methods, these methods search for solutions that are optimal with respect to the objective functions used in the methods and feasible with respect to the methods' constraints. Such methods may improve operation of a system only to the extent that the models and objective functions used by the methods accurately reflect the system characteristics. When such 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. An adequate way of gathering this data and ensuring that it is accurate has yet to be discovered; until one is discovered, the optimization methods mentioned above will be hampered by practical considerations.

In short, the potential value of optimization-based techniques will only be realized if the objective functions being optimized are accurate reflections of the real objectives of the parties involved. The objective functions used in the optimization-based techniques often fail to capture aspects of the real costs that are important to someone involved with the real problem. For the FAA, the job of finding realistic objective functions is complicated by the fact that 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. The challenge for the FAA is to find a way to gather objective function information from the users of the NAS and to combine them to find solutions that all users can agree are globally optimal. The information must be used in a way that motivates the users to provide accurate information; in the literature of game theory, the FAA's solutions must exhibit *incentive compatibility* (see, for example, [FT91]). Joseph Milner proposed a method of assigning arrival slots to flights that treated the incentive compatibility issue [Mil95]. His method required certain unrealistic assumptions to hold, as discussed in Section 4.5. Section 4 presents a new method that exhibits incentive compatibility under more reasonable conditions.

4 Objective-Based Assignment Method

A feature of the existing methods for distributing slots among the users is that they ignore the relative need for those slots. For example, if on a certain day one user has a large number of flights that are very sensitive to delay, and another user's flights are relatively insensitive to delay, it might make sense to schedule more of the first user's flights during peak demand periods, but existing methods would not do so.

This section introduces the Objective-Based Assignment Method (OBAM) for assigning arrival slots to flights based on the user objective functions. In assigning slots, OBAM attempts to maximize the value produced collectively by the users. Section 4.1 provides a precise definition of arrival slots required by OBAM. Section 4.2 describes the method, and Sections 4.3 and 4.4 present a proof that the method would maximize the value achieved collectively by the users of the system under some conditions.

4.1 Arrival Slot Definition

Recall that an arrival slot consists of the resources necessary for an aircraft to land: the sole use of a specific runway for a period of time and the attention of the air traffic control officials required to ensure a safe landing. Let arrival slot with label i be the i^{th} slot in sequence at a given airport. Define m_i to be the time at which slot i would begin were the airport to operate with the highest possible arrival capacity under perfect weather conditions. For an airport with a high maximum acceptance rate, the m_i are closely spaced in time, whereas at an airport with a low maximum acceptance rate, the m_i are more widely spaced.

Consider a realization in which fewer aircraft land, due either to reduced capacity constraints or to lack of demand. Arrival slot i retains its place in slot order, but is constrained to begin no earlier than time m_i. In other words, slot i begins as early as possible, but not before its time, m_i, and not before its place in slot order (after slots $1, \ldots, i-1$ are complete.) Slot i is complete when the aircraft reserving slot i has landed. If no aircraft has reserved slot i, it takes no time to complete, that is, its completion time is equal to its start time.

Generally, realized aircraft landings will be spaced farther apart in time than the m_i. On days of reduced landing capacity, the time required between landings is greater than in the maximal acceptance rate case. Let w_i denote the amount of time required between slots i and $i+1$. The w_i could be modeled as the realization of random variables W_i, which take on value $w_i = 0$ with probability corresponding to the probability that slot i will be unused, and value $w_i \geq s$ distributed according to the probabilistic characteristics of the weather and airport under consideration, where s is the spacing between the m_i. The time t_i at which an assigned slot i occurs in a given realization is given by

$$t_i = \max\{m_i, t_{i-1} + w_{i-1}\}$$

Note that the m_i are a deterministic feature of a given airport, independent of any realization, whereas the w_i take on values particular to the realization.

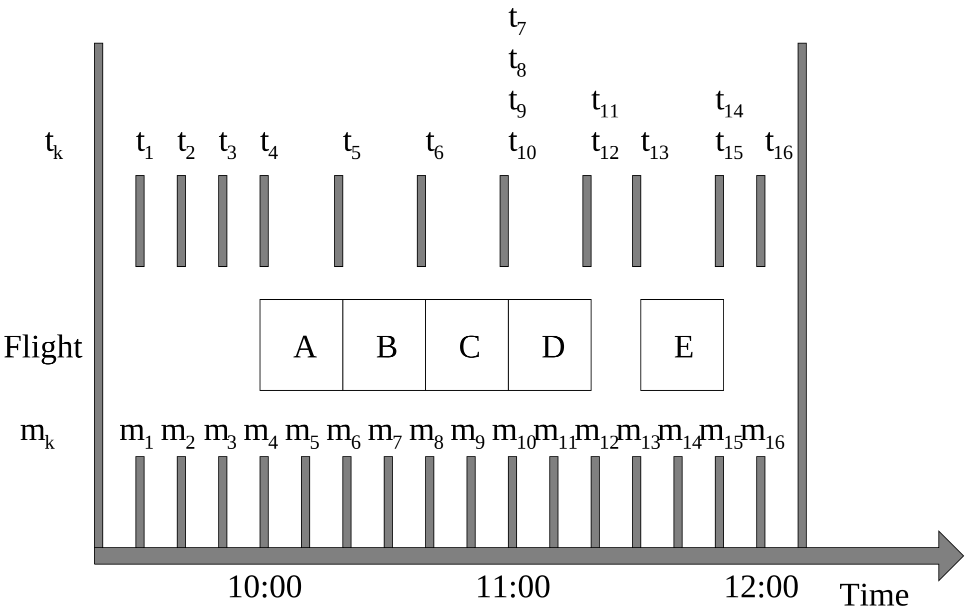

Figure 1: Fictional Assignment of Slots to Times at GAK.

For example, consider the situation today at the Gunk Hole, AK international airport (GAK) (see Figure 1). Arrival capacity is three aircraft per hour today, whereas maximal arrival capacity is six aircraft per hour. In other words, $s = 10$ minutes, $w_i = 20$ minutes for all flights. There are six slots per hour under maximal capacity conditions, so the m_i are spaced one every ten minutes. Five aircraft are landing at GAK today. Aircraft A has reserved slot 4, and $m_4 =$ 10:00 AM. Since aircraft A is the first arrival of the day, $t_4 = m_4 =$ 10:00 AM. Aircraft B has reserved slot 5, and $m_5 =$ 10:10 AM. Thus, $t_5 = \max\{m_5, t_4 + w_4\} =$ 10:20 AM, so aircraft B does not land until 10:20. Aircraft C, D, and E have reserved slots 6, 10, and 13 respectively, with times $m_6 =$ 10:20 AM, $m_{10} =$ 11:00 AM, and $m_{13} =$ 11:30 AM. Due to the reduced capacity, aircraft C's slot does not occur until 10:40 ($t_6 = t_5 + 20$) although $m_6 =$ 10:20 AM. Since slots 7,8, and 9 are unfilled, they each occur and are completed instantaneously at 11:00 AM. Slot 10 occurs at 11:00 AM as well ($t_{10} = \max\{m_{10}, t_9 + 0\} = \max\{11{:}00, 11{:}00\} = 11{:}00$, at which time aircraft D lands. Slots 11 and 12, also unfilled, occur instantaneously at 11:20, and finally aircraft E lands in slot 13 at 11:30.

A flight awarded use of a slot is able to produce a transportation service for its customers using that slot. The value of the slot to the flight may not be known exactly a priori, since the precise time at which the slot will occur is not known ahead of time, and the value of a flight depends to a large extent upon its time of arrival. Slot timing is uncertain largely because airport capacity cannot be forecast accurately.

4.2 The Objective-Based Assignment Method

OBAM uses a fixed *planning horizon*, the amount of time it plans into the future. Let T represent the current time, and let H represent the length of the planning horizon. At time T, the method considers the assignment of unassigned slots $\{i : T \leq m_i \leq T + H\}$. For instance, if H is 48 hours, all slots that would nominally occur within two days will be considered by the method.

The method represents the user preferences by a set of values, v_{ij}, each of which is interpreted as the value that can be produced by an assignment of slot i to flight j. The method also uses a *commitment time* for each possible slot-to-flight assignment. The commitment time, q_{ij}, is the time by which flight j must be notified of an assignment to slot i if such an assignment is made. This representation of user preferences is an approximation in two ways.

- The time at which the assignment is made may affect the value the user can achieve from the flight. For instance, the user may be better able to allocate resources the further in advance it knows the aircraft arrival times.

- The value of assigning slot i to flight j is not independent of other assignments, especially for users operating hub airports at which a bank of flights arrives, passengers change planes, and then the bank departs; in these instances, the value of arrival depends on time relative to the rest of the bank as well as on the time of day.

A characteristic typical of airport demand patterns is that the demand decreases during the night hours. Overnight landing capacity is almost always adequate to handle the demand and to allow any aircraft that were unable to land during the day to do so. OBAM is based on the assumption that such excess capacity exists at regular intervals. To be precise, let the term *slack point* refer to a point in time around which there is capacity sufficient so that any assignment of slots to flights that maximizes the sum of values achieved by the flights results in each flight receiving a slot with higher value to it than the value to it of any slot on the other side of the slack point. That is, if an assignment of slots to flights that maximizes collective value awards flight j a slot earlier than the slack point, the value to flight j of that slot is higher than the value of any slot after the slack point. In practice, it is likely that slack points meeting this definition would occur almost every night, since user value functions are roughly unimodal and of duration less than a day. Quantification of this statement is beyond the scope of this paper. Of course, there are plenty of examples at busy airports during badly congested days, such as Boston's Logan airport during a blizzard, for which it is likely that no slack point exists.

Denote the earliest slack point that is after the latest slot time that has a commitment time before or at time T as the *horizon slack point*. If the horizon slack point is within the planning horizon, and every flight with nonzero values v_{ij} for slots i before the horizon slack point has no nonzero values later than $T + H$, then OBAM will behave exactly as it would were there an infinite planning horizon. Figure 2 depicts this situation. In the figure, each horizontal bar represents the temporal characteristics of a potential slot-flight assignment: the left end of the bar represents the commitment time (q_{ij}), and the right end of the bar represents the slot time (t_i). Each flight would have one of these bars for each slot to which it assigned positive value v_{ij}. The definition of the horizon slack point requires that every bar with its left end to the left of T must have its right end to the left of the horizon slack point. If the time horizon is long enough so that every flight with a bar ending to the left of the horizon slack point does not have a bar ending to the right of the horizon end (time $T + H$), then the time horizon is long enough to take into account all values of every flight that might be assigned a slot before the horizon slack point. By the definition of a slack point, each flight can be constrained to land before the slack point if its maximal slot value occurs before the slack point, and it can be constrained to land after the slack point otherwise, without sacrificing any value. The infinite-horizon problem is thus separated into a series of independent problems covering the finite time between slack points. This is the motivation behind step 6 of OBAM below. In practice, when the existence of slack points is not known accurately in advance, it might be better to omit step 6.

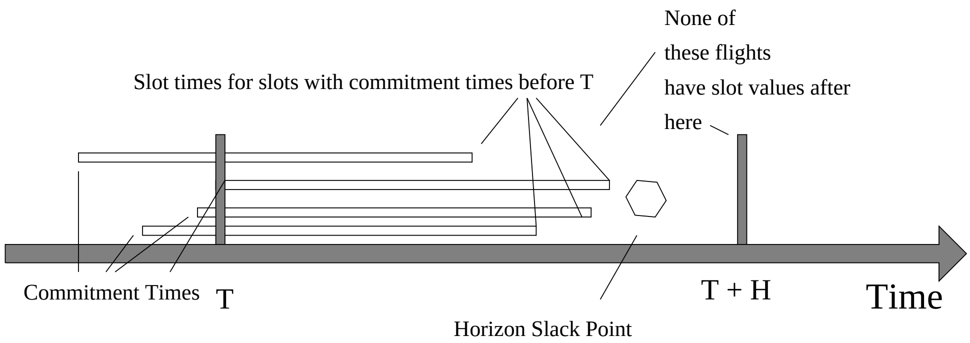

Figure 2: Time Horizon Constraint

Let $A = \{J_1, J_2, \ldots, J_a\}$ represent the set of *coalitions*. A coalition for the purposes of this paper is a maximal cardinality group of flights that cooperate to optimize their collective results from the assignment process. In other words, the flights of a coalition act on their collective interests rather than individual ones. For example, J_1 might be the flights belonging to American Airlines at the airport, J_2 might be flights belonging to another airline, and so on. Let $J = \cup_{i=1}^{a} J_i$, be the set of all flights. Note that by definition, $J_i \cap J_j = \emptyset, \ \forall i \neq j$.

OBAM follows the steps outlined below.

1. Initialize I^f, the set of indices of slots that have been assigned by the method, to $\emptyset$.

2. Set I to the indices of all slots to be assigned over the planning horizon, that is,

$$I = \{i : T \leq m_i \leq T + H\} \backslash I^f$$

3. Report to the users the estimates of times $t_i, \ \forall i \in I$

4. Accept input from flights. Each flight reports the following data:

   (a) For each slot $i \in I$, a value v_{ij} representing the economic value the slot represents to flight j relative to the value of receiving no slot.

   (b) Required commitment time, q_{ij}, for each slot i and flight j. The commitment time is the time by which the flight must be notified of an assignment to slot i.

There are at most $2|I|$ quantities that need to be reported for each flight, a value v_{ij} and a value q_{ij} for each slot in I. There could be fewer quantities if unreported values defaulted to zero.

5. Continue accepting input, and wait for the next time q_{ij} to occur. If while waiting, a new slot enters the time horizon ($\exists i \notin \{I \cup I^f\} : T \leq m_i \leq T + H$), go to step 2.

6. For each flight j whose values v_{ij} are nonzero on both sides of the horizon slack point, find the slot i which maximizes v_{ij}. Keep all values for that flight on the same side of the horizon slack point as slot i, and set all other values for that flight to zero.

7. Solve the assignment problem:

$$IZ(T) = \max_{x \in AP(I,J)} \sum_{j \in J} \sum_{i \in I} v_{ij} x_{ij}$$

where the set $AP(I, J)$ is the set of all x satisfying

$$\sum_{j \in J} x_{ij} \leq 1 \ \forall i \in I$$

$$\sum_{i \in I} x_{ij} \leq 1 \ \forall j \in J$$

$$x_{ij} \in \{0, \ 1\}$$

x is a vector of binary decision variables representing the assignment of slot i to flight j when x_{ij} is equal to 1. Let x^* represent the optimal solution to this problem.

8. Remove slot i and flight j from I and J respectively, and add i to I^f, for all i and j such that

   (a) $x^*_{ij} = 1$

   (b) $q_{ij} \leq T$

   Such assignments will be referred to as *final assignments.* Final assignments are made at their commitment times, and are assigned to the corresponding flight for use. If for some reason flight j determines that it is unable to use slot i after it has been assigned by a final assignment (as might happen if a mechanical problem with the aircraft is discovered after the assignment has been made), slot i is removed from the set I^f and returned to the set I, and may be used by a flight with a shorter commitment time. Assignments that are not final will be referred to as *reservations*.

9. Record all v_{ij} for all flights that have received final assignments, and record all v_{ij} for all slots that have been assigned as final assignments. These values will be referred to as *final values* and denoted v_{ij}^f. Note that a flight cannot change its value for a slot after it has received a final assignment without relinquishing its assignment, and it cannot change its value for a slot that has already been assigned in a final assignment, because values are only accepted for flights and slots in the sets I and J in step 4.

10. If T is not at a slack point, go to step 2.

11. For each coalition $r \in \{1, 2, \ldots, a\}$ calculate the final assignments to slots with times between T and the previous slack point that would have occurred had coalition r not been present. Let x_{ij}^{fr} represent these final assignments, and let $x_{ij}^{f\emptyset}$ represent the final assignments achieved with all coalitions present (i.e., the assignments realized in steps 2 through 10).

12. Define p_{I_r} to be the value of x^{fr} according to the final values submitted by the coalitions, minus the value to the coalitions other than r achieved by $x^{f\emptyset}$. That is,

$$p_{I_r} = \sum_{j \in J \backslash J_r} \sum_{i \in I} v_{ij}^f x_{ij}^{fr} - v_{ij}^f x_{ij}^{f\emptyset} \tag{1}$$

13. Assess each coalition $r \in \{1, 2, \ldots, a\}$ a fee p_{I_r}, and go to step 2.

4.3 Simplifying Assumptions

To simplify the analysis of the method, it is assumed that the objective values of the flights are solely a function of landing time. This is a strong assumption that ignores the relationship between flights from and to which passengers connect, and that relationship is a large component of the value of flights in banks at hub airports. As a practical matter, if this method were to be implemented, a user would be able to iteratively submit value functions to achieve results satisfying its connection objectives. That is, a user could in each iteration submit value functions for each flight j representing the value of landing flight j in slot i conditional on the slot assignments of flights with connections to and from flight j received in the previous iteration. Another approach would be to allow the users to specify contributions to the cost function of each flight j due to the time of flight j's slot relative to the slot times of flights with connections to and from flight j. Analysis of either approach is beyond the scope of this paper.

It is also assumed that each flight j must be notified of its assignment to slot i by a fixed commitment time, q_{ij}, and that assignments made by the commitment time have value v_{ij}, whereas assignments made after the commitment time have no value. In general, it is possible that a user might be able to realize greater value by an earlier commitment time, but still find later commitment times feasible (see Section 5.1). This information is not used explicitly in this formulation; however, it could be used by having the user submit q_{ij} with corresponding value dynamically.

Furthermore, it is assumed that the assignment of slots to flights does not affect the slot times. In reality, the order in which aircraft land affects the amount of time required for the landing. Large aircraft create more turbulence in their wake than do small aircraft, and large aircraft can fly through stronger turbulence than can small aircraft. Thus, a large aircraft may follow another large aircraft fairly closely along its landing approach, whereas a small aircraft must wait for the turbulence from a large aircraft to dissipate somewhat before following its path.

In practice, the FAA does not account for the dependence of slot times on assignment until the aircraft are near the airport. Once the aircraft are near the airport, the air traffic controllers may attempt to sequence the aircraft to meet a variety of constraints and to maximize the landing rate. The Center TRACON Automation System (CTAS), a tool developed at NASA, automates much of this task [Tom93]. The Final Approach Spacing Tool (FAST) is a tool under development at NASA that works with CTAS to optimize the landing rate [NPMD96]. Prior to aircraft entering the terminal airspace, the FAA uses an estimate of landing rate based on the expected performance of the tools in use at the destination airport. Since precise times of arrival at the destination airport are uncertain before takeoff, the assumption of independence of landing rate and assigned arrival sequence is reasonable; the weather causes more uncertainty in landing rate.

Finally, the analysis in the next section assumes that each coalition desires to maximize the value achieved in the assignment minus the fee assessed for the slots assigned to its flights.

4.4 Observations Concerning OBAM

Under the assumption that the values v_{ij} are an accurate representation of the user preferences and that they are fixed, it is clear that the method maximizes value realized by the community as a whole.

What remains to be shown is that the users are motivated to provide accurate information. The proof restricts attention to the times at which assignments are made. Since the flight data reported by each user are not shared with other users, the times at which assignments are made are unknown to the users. There is no motivation to report inaccurate information between assignments, since the values reported have no effect between assignment times. Thus, if there is motivation to report accurate information at assignment times, the users will be motivated to report accurate information at all times. The term *bid history* will be used to refer to the values v_{ij} in effect at the times at which assignments are made.

Proposition 1 *If each coalition knows its value functions in advance, then the solution in which each coalition accurately reports its value functions forms a Nash equilibrium*[1].

Proof: Assume coalition r knows its value functions in advance and that all coalitions other than coalition r report fixed values v_{ij}. Consider the result to coalition r of making an accurate bid. Under fixed values v_{ij}, OBAM results in an assignment x^*:

$$x^* = \arg \max_{x \in AP(I,J)} \sum_{i \in I} \sum_{j \in J} x_{ij} v_{ij} \tag{2}$$

[1] *A Nash equilibrium is defined in [FT91] as "a profile of strategies such that each player's strategy is an optimal response to the other players' strategies."*

Since all values v_{ij} for all coalitions other than r are fixed in this case, the fee assessed for the set of slots I_r, calculated by (1), simplifies

$$p_{I_r} = \max_{y \in AP(I,J)} \sum_{j \in J \setminus J_r} \sum_{i \in I} \left(v_{ij} y_{ij} - v_{ij} x^*_{ij} \right) \tag{3}$$

Notice that the fee charged to coalition r is a function only of the assignment x^*_{ij} and of the value functions v_{ij} of other coalitions. It is not a function of the values submitted by the coalition, except indirectly through the assignment. Therefore, the net value to coalition r, the value received through the assignment minus the fee assessed for it, is a function only of the assignment.

Assume that there exists an assignment $\hat{x} \in AP(I,J)$ that assigns slots $\hat{I}_r$ to coalition r and is preferable to x^* for coalition r, that is,

$$\sum_{i \in I} \sum_{j \in J_r} \hat{x}_{ij} v_{ij} - p_{\hat{I}_r} > \sum_{i \in I} \sum_{j \in J_r} x^*_{ij} v_{ij} - p_{I_r}$$

Using equation (1) to substitute for $p_{\hat{I}_r}$ and p_{I_r} gives:

$$\begin{aligned} &\sum_{i \in I} \sum_{j \in J_r} \hat{x}_{ij} v_{ij} - \max_{\hat{y} \in AP(I,J)} \sum_{j \in J \setminus J_r} \sum_{i \in I} \left(v_{ij} \hat{y}_{ij} - v_{ij} \hat{x}_{ij} \right) \\ > &\sum_{i \in I} \sum_{j \in J_r} x^*_{ij} v_{ij} - \max_{y \in AP(I,J)} \sum_{j \in J \setminus J_r} \sum_{i \in I} \left(v_{ij} y_{ij} - v_{ij} x^*_{ij} \right) \end{aligned}$$

Since the maximization over y and over $\hat{y}$ are on the same cost function and over the same feasible region, the objective values realized are equal and their sums cancel. The equation simplifies to:

$$\sum_{i \in I} \sum_{j \in J_r} \hat{x}_{ij} v_{ij} + \sum_{j \in J \setminus J_r} \sum_{i \in I} v_{ij} \hat{x}_{ij} > \sum_{i \in I} \sum_{j \in J_r} x^*_{ij} v_{ij} + \sum_{j \in J \setminus J_r} \sum_{i \in I} v_{ij} x^*_{ij}$$

Or, more simply,

$$\sum_{i \in I} \sum_{j \in J} \hat{x}_{ij} v_{ij} > \sum_{i \in I} \sum_{j \in J} x^*_{ij} v_{ij}$$

But this can only be true if x^* is not the optimal solution, violating (2). The contradiction proves that the optimal bid history for the coalition is the accurate one. Therefore, the proposition holds. □

Proposition 2 *If all coalitions other than r hold their bids constant, it is optimal for coalition r to correct any errors in its bid immediately.*

Proof: Consider the situation in which all coalitions other than r submit fixed bids, and in which coalition r has made arbitrary bids until time T. For purposes of this proof, denote the set of slots that have not been assigned in final assignments by time T as $I(T)$. Let $I^f(T)$ denote the set of slots that have been assigned in final assignments by time T, and let $I = I^f(T) \cup I(T)$. A fixed, accurate bid from time T onward by coalition r results in an assignment $x^* \in AP(I, J)$:

$$x^* = \arg\max_x \sum_{i \in I(T)} \sum_{j \in J} x_{ij} v_{ij} \tag{4}$$

Let $I_r(T)$ represent the set of slots assigned to coalition r after time T. Assume there exists a bid strategy that coalition r could employ at time T that would have a result preferable in the view of the coalition to the result achieved through an accurate bid. That is, assume that there exists $\hat{x} \in AP(I, J)$ which awards slots $\hat{I}_r(T)$ after time T to the coalition such that

$$\sum_{i \in I} \sum_{j \in J_r} \hat{x}_{ij} v_{ij} - p_{\hat{I}_r} > \sum_{i \in I} \sum_{j \in J_r} x^*_{ij} v_{ij} - p_{I_r}$$

and

$$\hat{x}_{ij} = x^*_{ij}, \ \forall i \in I^f(T) \tag{5}$$

Equation (5) constrains $\hat{x}$ to be an assignment that respects the assignments made final before time T. Separating the parts of the assignment made final before and after T, and applying equation (3) for $p_{\hat{I}_r}$ and p_{I_r} gives

$$\sum_{i \in I^f(T)} \sum_{j \in J_r} \hat{x}_{ij} v_{ij} + \sum_{i \in I(T)} \sum_{j \in J_r} \hat{x}_{ij} v_{ij}$$

$$- \max_{\hat{y} \in AP(I,J)} \sum_{j \in J \setminus J_r} \left(\sum_{i \in I^f(T)} (v_{ij}\hat{y}_{ij} - v_{ij}\hat{x}_{ij}) + \sum_{i \in I(T)} (v_{ij}\hat{y}_{ij} - v_{ij}\hat{x}_{ij}) \right)$$

$$> \sum_{i \in I^f(T)} \sum_{j \in J_r} x^*_{ij} v_{ij} + \sum_{i \in I(T)} \sum_{j \in J_r} x^*_{ij} v_{ij}$$

$$- \max_{y \in AP(I,J)} \sum_{j \in J \setminus J_r} \left(\sum_{i \in I^f(T)} \left(v_{ij} y_{ij} - v_{ij} x^*_{ij}\right) + \sum_{i \in I(T)} \left(v_{ij} y_{ij} - v_{ij} x^*_{ij}\right) \right)$$

By equation (5), the sums involving $\hat{x}_{ij}$ and x^*_{ij} over the set $I^f(T)$ cancel, leaving

$$\sum_{i \in I(T)} \sum_{j \in J_r} \hat{x}_{ij} v_{ij} - \max_{\hat{y} \in AP(I,J)} \sum_{j \in J \setminus J_r} \left(\sum_{i \in I^f(T)} v_{ij}\hat{y}_{ij} + \sum_{i \in I(T)} (v_{ij}\hat{y}_{ij} - v_{ij}\hat{x}_{ij}) \right)$$

$$> \sum_{i \in I(T)} \sum_{j \in J_r} x^*_{ij} v_{ij} - \max_{y \in AP(I,J)} \sum_{j \in J \setminus J_r} \left(\sum_{i \in I^f(T)} v_{ij} y_{ij} + \sum_{i \in I(T)} \left(v_{ij} y_{ij} - v_{ij} x^*_{ij}\right) \right)$$

or,

$$\sum_{i \in I(T)} \sum_{j \in J_r} \hat{x}_{ij} v_{ij} + \max_{\hat{y} \in AP(I,J)} \sum_{j \in J \setminus J_r} \left(\sum_{i \in I(T)} v_{ij}\hat{x}_{ij} - \sum_{i \in I} v_{ij}\hat{y}_{ij} \right)$$

$$> \sum_{i \in I(T)} \sum_{j \in J_r} x^*_{ij} v_{ij} + \max_{y \in AP(I,J)} \sum_{j \in J \setminus J_r} \left(\sum_{i \in I(T)} v_{ij} x^*_{ij} - \sum_{i \in I} v_{ij} y_{ij} \right)$$

Again, the maximizations over y and $\hat{y}$ use the same cost functions over the same feasible sets, and cancel, leaving

$$\sum_{i \in I(T)} \sum_{j \in J} \hat{x}_{ij} v_{ij} > \sum_{i \in I(T)} \sum_{j \in J} x^*_{ij} v_{ij}$$

But this contradicts equation (4). Therefore, coalition r can do no better than to make an accurate bid immediately, and the proposition holds. □

Of course, these results do not mean that the Nash equilibrium solution would occur in practice. For instance, if a coalition knew that future slots would be scarcer than everyone else thought, that coalition could improve its solution by lowering the value it placed on the scarce slots. Conversely, if it knew that future slots would be more abundant than generally thought, it could lower the value placed on the slots in the near future in order to get a cheaper slot later. In effect, OBAM allows coalitions to use proprietary information to improve the results achieved. If something like OBAM were implemented, it would encourage the users of the NAS to improve their forecasting capabilities. Notice that by using its proprietary information, if that information is correct, the coalition improves not only its results, but it improves the overall results as well by lowering demand for the congested slots.

4.5 Static Specialization

Joseph Milner proposed a simpler method of assigning slots to flights in [Mil95]. That method applied results from [RS90], requiring two assumptions to be made. First, it was assumed that slots for an entire day could be assigned to flights at the beginning of the day, and that the resulting assignment would remain feasible throughout the day. Second, it was assumed that each flight sought to maximize its own profit. In reality, it is each airline, rather than each flight, that attempts to maximize its profit; this fact has the consequence that the airlines would not be motivated to report the v_{ij} accurately. Under Milner's approach, each flight pays a fee for the lost productivity of all other flights. As a result, a coalition of flights pays a fee for the lost productivity that it causes itself. In many situations, the coalition can report artificially low values for some of its flights to reduce the fees it pays without significantly affecting the slot assignments; it is simple to construct examples in which a coalition's profit is maximized when it reports values that achieve a globally suboptimal solution.

OBAM can be specialized to exactly the method used by Milner and [RS90] if the coalitions are limited to contain at most one flight each and the commitment times for all flights to occur during the day are set equal to a time in the beginning of the day. If OBAM is further restricted so that each flight desires only a single slot, the famous second-price auction [Vic61] is realized.

4.6 Potential Benefits and Drawbacks of OBAM

There are several benefits to the objective-based approach. Both the current system of assigning slots to flights and the CDM group's proposed revisions to it share the problem that they reward a user for scheduling and then canceling flights they never intended to fly; the CDM group has proposed regulation and penalties for such acts, however there is no clear way to differentiate intentional abuse of the system from legitimate operations. Regulations currently in place to prevent abuse of the system have the disadvantage that they prevent airlines from making legitimate cancellations, wasting airline and airport resources. None of these problems occurs under the objective-based assignment method: a user that schedules and cancels a ghost flight pays for the lost productivity that act causes other users and achieves no overall improvement in profit from it. Furthermore, a user that realizes that its schedule is not profitable can cancel flights; Proposition 2, if its result can be extended to practice, indicates that it is in the user's best interest to inform the FAA of the true situation as soon as possible.

Another benefit of the objective-based assignment method is that it takes into account user preference information. Current operations distribute slots among users with the goal of minimizing aircraft-minutes of delay. Collaborative decision-making methods such as OBAM allow users to specify their own criteria to be used in part or all of the slot distribution. For instance, an airline that prefers to minimize passenger-minutes of delay, late baggage costs, missed passenger connections, fuel and crew costs, and any other important criteria can do so. Measured in these terms, the throughput of the NAS could be higher than it is currently, especially during periods of reduced acceptance rate. Improving system throughput is becoming increasingly important now with growth of demand for use of the NAS far outpacing increases in its capacity.

OBAM also motivates the users to provide preference information, which may have value in and of itself. The preference information could be used to improve the objective functions of other tools such as CTAS and FAST, further improving the throughput of the NAS. It could also be used to provide an accurate quantification of the costs of delay and of the economic benefit to capacity-increasing technology. This quantification could be used as a basis for policy and investment decisions, improving the benefit-to-cost ratio of infrastructure and other improvement projects undertaken.

A drawback of OBAM is that it requires the users to pay fees for the slots they get. Under the current system, users are taxed through other means for the use of landing facilities; they are accustomed to thinking of landing slots as free, and would likely resist any method that charged for them.

One attempt to address this concern, treated by Milner in [Mil95], was to issue chits to the users based on an assignment done the old way. That is, the FAA would run their present assignment method and issue to the users a predetermined number of chits for each slot assigned. The users then used these chits as money with which to pay the slot fees required by this method. One problem with this approach is that a user could run out of chits during the day and be unable to land the rest of its aircraft.

A similar approach might be to run both the FAA's existing assignment method and OBAM as follows. First, the FAA's existing assignment method would be run, and the slots assigned in the process would be given to the users. Then, OBAM could be run, charging real money for the slots assigned, and giving the money collected for each slot to the owner of the slot. This approach has the advantage that a user uncomfortable with OBAM could specify a very high value on each of its FAA-assigned slots and zero value elsewhere, and would as a result receive its FAA-assigned slots at no cost (since OBAM would refund any costs to the user). Users desiring to participate in OBAM would be able to do so by submitting bids reflective of their true slot values, and would benefit by doing so, according to the analysis of Section 4.4; the resulting assignment might cost the user some money, but the cost would be more than offset by the improvement in the value of the slots received.

The benefit of both of these approaches is that they rely on the FAA's existing slot assignment method, which has been accepted by the users. Additionally, the latter approach has the benefit that each user would be guaranteed to do at least as well as it would in the FAA's assignment method. The drawback of this is that the existing assignment method assigns slots to a user proportional to the user's representation in the schedule. It would be in a user's interest to find ways to schedule ghost flights.

5 Future Work

A number of efforts have the potential to advance this work. One effort that may prove fruitful is to extend the work of Section 4.4 to consider the behavior of the coalitions when they possess imperfect information about their future values. It is not immediately obvious what the optimal bidding strategy for a coalition is under imperfect information, since it is possible for the bids submitted by other coalitions to change as the other coalitions gain new information. Perhaps the largest source of uncertainty in the coalitions' values is the uncertainty of slot times; the practicality of the method might be improved by consideration of the coalitions' optimal responses given distributions for the w_i of Section 4.1 rather than the actual values t_i reported by the FAA in OBAM. Further investigation in this direction might consider situations in which each coalition has access to proprietary information about the w_i and t_i.

Furthermore, it might make sense to combine the slot assignment problem discussed in this paper with the stochastic ground holding problem discussed in [Ric91] and [Rif97]. Optimization approaches to the ground holding problem have relied on cost functions meant to model the airlines' cost structure, but have not provided any means for the airlines to provide their cost functions directly. Indeed, addition of mechanisms to get airline input into the cost functions used in the ground holding problem may prove useful in and of itself.

Another topic that merits work is the addition of a way for the users to specify cost functions that capture the costs inherent in banks of flights and passenger connections. There is value to a user in having a flight arrive early enough for its passengers to make their connections on other outgoing flights. Banks are sets of flights that are scheduled to be at an airport at about the same time and among which many passengers have connections. The cost function form used by OBAM, wherein the value of an assignment is assumed to equal the sums of the independent values of each slot-to-flight assignment, cannot represent the costs of banks and passenger connections. One approach might be to add a component to each flight's cost function representing the value of slot time relative to other flights' slot times. Another approach might be to analyze the solutions that would be achieved by users that had bank-type costs but who were limited to the cost function form used by OBAM.

It is possible that these users could achieve good results by adjusting their submitted v_{ij} to achieve slots that worked well for their banks as well.

Another direction work might take is the analysis of dynamic distribution of a public resource in general. In general, it may not be possible or desirable to require players to specify commitment times before which an assignment has fixed value and after which the assignment has no value. In general, an earlier assignment gives the player more options and limits the central authority's options. The next section discusses a generalization of the problem.

5.1 Generalization

There is a constrained resource, discretized into slots, that must be divided among several players. Each player can use some slots in his business to produce a product. The value to the various players of the use of any given slot is different, and is not known perfectly a priori. It may be necessary due to the logistics of his business for a player to commit to using a resource ahead of time, such that he will incur a *commitment cost* if he changes his mind.

Assume the objective is to maximize the expected value of the outcome, taking into account the fact that the quality of the available information improves with time. The problem can be posed as a dynamic program. The state of the dynamic program in stage i consists of the set of players (p_i) and of slots (s_i) that have yet to be assigned and the states y_i of the players relevant to their ability to use the slots:

$$x_i = \begin{pmatrix} p_i \\ s_i \\ y_i \end{pmatrix}$$

The controls of the dynamic program are the assignments that can be made at each stage and the controls that the players have. The state equations are

$$x_{i+1} = f_i(x_i, u_i, w_i)$$

Where x_i is the state at stage i, u_i is the control made at stage i, and w_i is the outcome of a random variable at stage i. f_i is a general state transition function involving the states of the players' businesses as well as the assignments made. The controls available to the central authority are to make assignments of slots to players, thus the state transition function for central authority controls simplifies to

$$x_{i+1} = \begin{pmatrix} p_i \backslash p_i^f \\ s_i \backslash s_i^f \\ y_i \end{pmatrix}$$

where p_i^f and s_i^f are the sets of players and slots, respectively, involved in final assignments in stage i.

The controls available to the players are both the operational decisions required to run their businesses and the information to give the central authority. The dynamic programming algorithm is specified by Bellman's equation for the optimal expected cost to go, $J^*(x_i)$, following the notation of [Ber95].

$$J^*(x_i) = \max_{u_i \in U(x_i)} E[V_i(x_i, u_i) + J^*(f_i(x_i, u_i, w_i))]$$

where $V_i(x_i, u_i)$ is the stage i value of control u_i from state x_i.

The assumption is that the decisions in the system are being made by multiple decision-makers, one of whom is the central authority with a social welfare objective and the rest of whom can be classified as players with private objectives. Player k's decision can be summarized as

$$J^*(x_i) = \max_{u_i^k \in U^k(x_i)} E[V_i^k(x_i, u_i^k) + J^*(f_i(x_i, u_i^k, w_i))]$$

where $V_i^k(x_i, u_i^k)$ is the stage i value to user k of user k's control u_i^k from state x_i.

While the central authority's decision of which assignments to make at stage i can be represented by

$$J^*(x_i) = \max_{(p_i^f, s_i^f) \in x_i,} V_i(p_i^f, s_i^f) + E[J^* \begin{pmatrix} p_i \backslash p_i^f \\ s_i \backslash s_i^f \\ y_i \end{pmatrix}]$$

OBAM made vast simplifications to the users' state transition functions in order to reduce the state space and simplify the central authority's decision problem. In particular, OBAM reduces the dynamics of the players' businesses by assuming that each player has a fixed commitment time before which the player can feasibly use a slot with no additional cost and after which it is infeasible for the player to use the slot. It is possible that different simplifications of the generalized dynamic program could lead to methods that model the real situation more accurately than does OBAM.

References

[Ber94] Dimitris J. Bertsimas. The air traffic flow management problem with enroute capacities. In *Working Papers*. MIT Sloan School of Management, 1994.

[Ber95] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 1995.

[Byr97] Tim Byrne. Email correspondence, February 1997. Information on which TWA's substitution software is based.

[CLP+95] H. Stewart Cobb, David G. Lawrence, Boris S. Pervan, Clark E. Cohen, J. David Powell, and Bradford W. Parkinson. Precision landing tests with improved integrity beacon pseudolites. In *ION GPS-95 Proceedings*, Palm Springs, California, September 1995. Paper available as of May 7, 1997 at http://stugyro.stanford.edu/GPS/ABS/improved_IBs_stu95.html.

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

[Mil95] Joseph M. Milner. *Dynamic Slot Allocation with Airline Participation*. PhD thesis, Massachusetts Institute of Technology, 1995.

[NPMD96] Don Nolan-Proxmire, Michael Mewhinney, and Les Dorr, Jr. Nasa/faa testing new air traffic management tool. Nasa Press Release 96-31, February 14 1996.

[Ois96] Rick Oisen. Issues concerning the compression algorithm. Technical report, Volpe National Transportation Systems Center, June 1996.

[Ois97] Rick Oisen. Email correspondence, April 1997. Data found in Volpe National Transportation Systems Center configuration management logs.

[PCL+95] Boris S. Pervan, Clark E. Cohen, David G. Lawrence, H. Stewart Cobb, J. David Powell, and Bradford W. Parkinson. High integrity gps-based precision landing using integrity beacon pseudolites. In *ISPA*, Braunschweig Germany, February 1995. Paper available as of May 7, 1997 at http:/stugyro.stanford.edu/GPS/ABS/prec_landing_w_ibls_bsp95.html.

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

[Rif97] Ryan Rifkin. The single-airport static stochastic ground holding problem. Research-Oriented Paper for Operations Research Center General Examination, Massachusetts Institute of Technology, May 1997.

[RS90] Alvin E. Roth and Marilda A. Oliveira Sotomayor. *Two-Sided Matching; A Study in Game-Theoretic Modeling and Analysis.* Cambridge University Press, 1990.

[Sab] Yield management solutions. Sabre Decision Technologies web page. Available as of May 7, 1997 at http://www.sabre.com/sabr_grp/sdt/solution/yield.htm.

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

[Tom93] Pat Tomasetti. Ctas on the fast track. FAA World, Employee Publication of the FAA, November 1993. Article available as of May 7, 1997 at http://www.tc.faa.gov/ZDV/faaworld/ctas.html.

[Vic61] William Vickrey. Counterspeculation, auctions, and competetive sealed tenders. *Journal of Finance*, 16, 1961.

[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 May 7, 1997 at Web Site http://www.metsci.com/faa/gen_doc.html, July 1996.

[Wam97a] Michael Wambsganss. Email correspondence, February 1997.

[Wam97b] Michael Wambsganss. What's cdm in aviation transportation? *Collaborative Decision Making in Aviation Transportation*, 1997. Web Page available as of May 7, 1997 at http://www.metsci.com/faa/bsum.html.

[Yu96] Jung S. Yu. Airport capacity and regional weather modelling. Master's thesis, Massachusetts Institute of Technology, 1996.