The Static Stochastic Ground Holding Problem

Robert L. Hoffman [*]
Ryan M. Rifkin [†]Michael O. Ball [‡]Amedeo R. Odoni [§]

1 Introduction

Over the last two decades, air traffic in the United States has grown more rapidly than the capacity of the national airspace system (NAS). While there are congestion problems in the air space itself, most of the air traffic delays occur at the terminals, particularly in times of bad weather. On any given day, the total number of flights into an airport rarely exceeds the daily capacity but the conincidence of arrival capacity reductions and arrival demand surges induce arrival queues and massive delays. As a result, many airports are currently operating at the fringe of their capacity.

Spikes in arrival demand are generally caused by the clustering of flights created by airline scheduling. The airlines tend to group flights around the times of day that passengers most prefer to travel, namely, in the morning and evening hours. Another scheduling factor is the hubbing operations of some of the major U.S. carriers. A hub-and-spoke operation provides an ecomnomically feasible way for an airline to service its locations without forming direct flights between every city pair. However, this practice tends to concentrate traffic in the major metropolitan areas, where the hub sites are usually located.

Since there would be little support for any long-range plan that asks the airlines to curtail well-established operationl paradigms or that asks the

[*]University of Maryland
[†]Massachusetts Institute of Technology
[‡]University of Maryland
[§]Massachusetts Institute of Technology

public to travel at odd times of the day to unpopular locations, it has proved to be more promising to concentrate efforts on increasing airport capacity rather than reducing demand. Unfortunately, plans for airport constuctions and expansion are often hindered by prohibitively expensive real estate and by the protests of surrounding communities concerned with the potential increase in noise and automotive traffic. As a result, efforts in both research and air traffic flow management are focused on the efficient use of current resources.

The Air Traffic Control Systems Command Center (ATCSCC) facility of the FAA monitors airports throughout the United States for capacity-demand inequities. Whenever it is predicted that the number of flights arriving at an airport within a 15-minute time interval will exceed the number of flights predicted to land, the ATCSCC is requried by FAA regulation to take some form of action. Short-term periods of capacity shortfall are alleveiated by airborne tactics such as rerouting and variations in ariborne speed. Longer-term priods of capacity shortfall are met by the ATCSCC with ground holding strategies in which aircraft are held at their respective departure gates in lieu of airborne holding at the destination airport.

In some capacity-demand instances, the ATCSCC issues a ground stop in which all flights destined for an affliced airport are held on the ground at their departure gates until airport arrival capacity rises above demand. These ground strops are reserved for extreme cases in which arrival capacity was severly underestimated or dropped suddenly without warning.

The primary tool of the ATCSCC for addressing arrival capacity shortfall is a ground delay program (GDP). This is a more strategic action than a ground stop and is generally planned several hours in advance of a predicted capacity-demand inequity. In a GDP, the rate of the arrival stream into an afflicted airport is reduced by holding each inbound flight at its departure gate long enough to ensure that it will be able to land without delay. For instance, a flight with a predicted airborne holding of 30 minutes at the arrival airport would be held for 30 minutes at its departure gate, thereby replacing aniticpated airborne holding with gate holding.

Although there are many practical complications to consider when implementing a GDP, the basic procedure is as follows. The time horizon for the period of the capacity-demand inequity is divided into hourly periods, $t = 1, 2, ..., T$. Usually, there are 4-6 hours in all but GDPs as long as 12 hours are becoming more commonplace. For each hour t, an arrival acceptance rate of the airport in question is forecasted. This is the number of flights that the airport will be able to handle in the time period t. (Strictly speaking, this is a capacity not an arrival rate, however, it is the established terminology.) Each time period t is divided into evenly spaced arrival slots. For instance, if the AAR for the first hour is 30 flights, then the period $t = 1$ is divided into 30 sequential, two-minute arrival slots. Next, the flights are linearly ordered according to their scheduled arrival time (formerly, their estimated time of arrival, more on this later) and assigned to arrival slots, preserving order. Since the AAR for each time period is usually less than the number of flights predicted to arrive, this has the effect of stretching out the list of arrivals over time. Once a controlled time of arrival (CTA) is set for each flight, its controlled time of departure (CTD) is set by subtracting its predicted en route time from its CTA.

Certain categories of flights, such as international flights and general aviation, are exempt from ground holding. In addition, the traffic flow manager has the option to exempt flights from the program based on the proximity of the origination airport to the destination airport. For instance, when formulating a GDP for an airport on the east coast of the United States, the traffic flow manager may choose to exempt all long-distance flights, hence, all flights originating on the west coast.

One must be careful to distinguish between the arrival acceptance rate (AAR) that will actually occur and the planned arrival acceptance rate (PAAR) that is set by the traffic flow manager at the ATCSCC. The former is the number of flights that will actually be able to land at the airport while the latter is the number of flights that will attempt to land, based on the controlled times of arrival assigned during the GDP. In a perfectly planned and executed ground delay program, $AAR_t = PAAR_t$ for each time period t in the time horizon. AARs are a function of runway configurations, which are, in turn, dictated by weather conditions. Since most GDPs are necessitated by bad weather, AARs are generally unpredictable and it is rare that $AAR_t = PAAR_t$ for each time period t.

If weather conditions turn out to be more favorable than forecasted, then AARs are overestimated and the GDP does not send enough flights to match airport resources. It is difficult for the manager of a GDP to adjust for changing weather conditions once a GDP is in effect. For example, consider the scenario in which it is predicted that an airport will be fogged in betwen the hours of 12:00 and 16:00. In response, ground delays have been issued so that the number of flights that will arrive in the first hour is cut back from 60 to 30 flights (PAAR = 30). Now, suppose that the fog lifts just before 12:00.

Then the capacity of the airport during the first hour will be at its normal opeating level of, say, 60 flights per hour (AAR = 60). The flights bound for the airport in the 12:00 hour (12:00-12:59) are most likely airborne and it is too late to retract the ground delay issued to them. These flights will have suffered needless delays and the airport arrival resources will be wasted because the airport will accept only 30 flights when (in hindsight) it could actually have accepted 60 (AAR = 60).

On the other hand, if weather conditions turn out to be worse than forecasted, then AARs are underestimated and the GDP sends more flights than airport resources can accommodate. As a result, incoming aircraft must be put into airborne queues, airport safety conditions are joepardized, and flights incur airborne delay that could have been absorbed on the ground, had a more aggresive GDP been enforced.

The purpose of this paper is to develop a model to help the traffic flow managers at the ATCSCC plan for the appropriate amount of ground holding in each time period (hour) of a GDP, when faced with uncertain capacity.

The problem of assigning ground delay to flights bound for a single airport can be mathematically modeled as a transportation problem known as the *ground holding problem* (GHP). GHP was originally proposed by Odoni in [14]. Odoni, Andreatta, Richetta and Odoni furthered work on this single-airport problem in [3], [16] and [17]. Both GHP and air traffic flow management in general have been treated on a network-wide level (taking multiple airports or flight connectivity into account) in Attwool [4], Sokkapia [21], Andreatta and Romanin-Jacur [2], Wang [26] and by Vranas, et. al., in [24] and [25], and, more recently, by Bertsimas and Stock [5]. In [11], Hoffman and Ball have adapted the single-airport GHP to accommodate the banks of flights associated with the hubbing operations of the major airlines.

In [22], Terrab and Odoni developed a dynamic programming formulation for the stochastic version of GH as well as heuristics to handle the larger cases. Using stochastic linear programming with recourse, Richetta and Odoni expanded this work to include the dynamic case, in which ground holdings are updated as time progresses (see [16]). Although their dynamic solution yielded considerable savings over the static solution, the speed of solution proved to be too slow for realistic cases (see [6]).

There are many approaches to stochastic programming. See [7], [12], or [15] for background. In this paper, we assume the existence of a discrete number S of scenarios, or profiles, of arrival capacity, exactly one of which will come to pass. Each scenario consists of a sequence of T arrival capacities

$A_1, A_2, ..., A_T$, where each t corresponds to a time interval in the horizon. Thus, A_t is the arrival capacity that will occur in the t^{th} time period if that scenario should be realized. This use of discrete scenarios is similar to the technique applied by Richetta and Odoni in [17] and Chapter 3 of [18].

We feel that this view is consistent with our current knowledge of weather patterns and runway configurations. Forecasts are formulated based on the net movements of weather components such as high pressure areas and storm fronts. The intuition for this type of forecasting is based on experience at airports such as San Francisco where fog is the main problem and a set of scenarios can be generated such that each one reflects a time at which the fog might roll in and a time at which it might roll out. Similarly, scenarios can be generated for an inevitable winter storm at an airport in the eastern United States. In this case, the scenarios would reflect the duration and severity of the storm.

In Figure 1, we graphically display a "typical" multiple AAR scenario forecast for a fictitious airport whose normal AAR is 70 flights per hour. The pessimistic forecast, scenario 1, predicts that the AAR will plummet to 30 flights per hour at the third hour and not recover until the eighth hour. Scenario 2 is similar in pattern to scenario 1 but predicts a less severe, shorter storm period that will begin one hour earlier. Scenario 3, the optimistic scenario, predicts heavy impact only in hours 4, 5 and 6, with hour 6 being the worst at an AAR of 55 flights per minute.

Let us consider an intuitive approach to planning AAR's and assume that scenario 2 is to occur with high probability, say, $P(1) = .95$. From an intuitive standpoint, the traffic flow manager would do well to set his $PAAR's$ in close accordance with, if not exactly equal to, this scenario. i.e., $X_1 = 60, X_2 = 50, X_3 = 50, X_4 = 50, X_5 = 60, X_6 = 55$, and so on. However, as the likelihood shifts toward scenario 3, say, $P(1) = .05, P(2) = .55, P(3) = .40$, considerable weight should be given to the optimism of scenario 3 and very little to the pessimism of scenario 1. A GDP based on scenario 2 would probably prove to be overly aggressive and lead to unacceptable levels of unnecessary ground-holding, thus pushing airport demand well below that of capacity. It seems that in the formulation of the GDP, the traffic flow manager should hedge toward scenario 3 and perhaps disregard scenario 1 altogether.

It is not uncommon for storm intensity to peak twice or more in an affected region, thus leading to peaks and valleys in the AAR scenario, instead of the simplified concave patterns of the foregoing example. This greatly obscures the optimal solution and renders almost useless any intuitive approach for formulating a GDP based on simple graphical analysis.

The next section (section 2) of this paper develops an integer programming model for the single-airport, static (non-dynamic) ground-holding problem. Section 2.1 presents a simplified version of the model in which some of the variables have been eliminated through preprocessing. Section 3 addresses some of the practical issues involved with implementing the model. Section 4 explores the and shows that the proposed model allows the integer solution to be obtained directly from its linear programming relaxation. Section 5 gives computational results for several large-scale, realistic test cases. Section 6 summarizes the major results of the paper and discusses areas of future research.

2 Model Formulation

Let $t = 1, 2, ..., T$ be a set of discrete, contiguous time intervals during which the AAR of an airport is in jeopardy. Each t could represent, say, a 15-minute time interval. Let $\overline{D}_t$ be the total number of flights scheduled to arrive during time interval t. Let D_t be the total number of (non-exempt) flights that will be included in the GDP. Then the number of exempt flights is given by $E_t = \overline{D}_t - D_t$ with $E_t \geq 0$.

We assume that the set of flights that are candidates for a given GDP is partitioned into disjoint classes $e = 1, 2, ..., E$, which we will call **tiers**. For each tier, we define a decision variable y_e, such that $y_e = 1$, if tier e is to be included in the GDP and $y_e = 0$, otherwise. Let d_t^e be the number of flights in tier e. Then we have $D_t = \sum_{e=1}^{E} d_t^e y_e$.

Let A_t be the PAAR (planned arrival acceptance rate) for time period t, i.e., the number of flights that will be assigned a controlled time of arrival that falls within time period t. Because of the stochastic nature of weather, A_t flights will not necessarily be able to land during time interval t. For instance, if $A_1 = 10$ and the airport can land only 8 flights at time $t = 1$, then 2 of the 10 flights scheduled to arrive at time $t = 1$ will be held in the air by the controllers at the destination airport. Let us assume that $A_2 = 10$ and that the capacity of the airport is, say, 15 flights for the next period so that the two airborne-held flights are able to land in the next time interval, $t = 2$. Then the airborne delay under this single AAR scenario can be depicted by the flow diagram in Figure 2.

Assuming that ground holding is less desirable than airborne holding (otherwise there is no need for a GDP), it would have been better to hold two flights on the ground for one time unit and have them arrive at the terminal airspace at time $t = 2$. The single-unit ground delay of these two flights can be represented by adding a horizontal arc at the top of Figure 3.

Note that Figure 3 treats flights on an aggregate level, with no mention of which flights will absorb the ground delay. Recall from the introduction that the flights are sequenced by landing order prior to the formulation of a GDP. Thus, it is easily deduced that the first eight flights are allowed to land at time interval $t = 1$ while the ninth and tenth flights will absorb the ground delay and land in time interval $t = 2$.

Figure 3 is easily extended to allow for an arbitrary number of time periods, T. Moreover, arcs can been added to allow for the exemption of flights, as in Figure 4. The exempted flights flow directly to the bottom nodes, indicating that they will be accepted to the terminal airspace under any conditions. These flights are not exempt, however, from possible airborne holding. Figure 4 assumes the following notation.

E_t = number of flights scheduled to arrive at time t that are exempt from ground delay.

b_t = number of flights that can be landed at the airport (the actual AAR) during time period t.

G_t = number of (non-exempt) flights wishing to land at time t but not accepted to the terminal airspace at time t (these flights have been ground delayed).

W_t = number of flights held in the air from time period t to $t + 1$ (these flights have been airborne delayed).

Next, we extend Figure 4 to accommodate multiple AAR scenarios. Let $s = 1, 2, ..., S$ be a set of AAR scenarios, each with an associated probability, p_s. The capacity of the airport (AAR) at time t varies with the scenario, s, hence, so does the amount of airborne holding at time t. We generalize the above notation as follows.

W_t^s = number of flights held in the air from time period t to $t + 1$,under scenario s.

b_t^s = number of flights that can be landed at the airport (the AAR) during time period t, under scenario s.

p_s = the probability of scenario s occurring.

We obtain Figure 5 from Figure 4 by replicating the lower portion of Figure 4 once for each scenario. Our integer programming model will be based on this diagram.

The hypothetical path of a single flight through the flow diagram is depicted in Figure 6. The flight enters the system at time $t = 1$, indicating that it is originally scheduled to arrive at time $t = 1$. Next, the flight absorbs two units of ground delay, then is allowed to enter the terminal airspace and absorbs two unit of airborne delay. Note that the flow pattern gives the illusion that the flight is being held on the ground just prior to arriving at the terminal space (attempting to land). In reality, it will absorb its ground delay at its origin airport at some earlier time, which is easily computed based on the estimated enroute travel time for the flight.

Let c_g and c_a be, respectively, the costs of holding a flight for one time unit on the ground and one time unit in the air. Given a scenario, s, the total delay cost of the GDP is given by $\sum_{t=1}^{T}(c_g G_t + c_a W_t^s)$, and the expected total delay cost of the GDP is given by $\sum_{s=1}^{S}\left[p_s \cdot \sum_{t=1}^{T}(c_g G_t + c_a W_t^s)\right]$.

In formulating a GDP, the traffic flow manager tries to minimize the total expected delay cost subject to the constraints imposed by Figure 5. We now develop those constraints.

First, we model the flow through the upper (round) nodes of Figure 5, using the following relation:

$$G_{t-1} + D_t = A_t + G_t \qquad t = 1, \ldots, T,\ G_0 = 0 \tag{1}$$

To see that these equations are valid, recall that A_t is the number of flights that will be accepted to the terminal (airport) airspace (but not necessarily land) at time t, that D_t is the number of flights scheduled to land, and that W_{t-1} is the number of flights in the airport airspace at time $t-1$ that were unable to land. Then $W_{t-1} + D_t$ is the number of flights wishing to land at time t and (1) simply says that this number is equal to the number accepted to the airspace at time t (A_t) plus the number kept away from the airspace at time t (G_t).

We should point out that, unless $S = 1$, the equations of the form 1 are not based on conservation of flow. For each scenario s, there is one arc flowing out the bottom of any given upper (round) node, and each of these S arcs has flow A_t. This labeling is justified by the fact that once the planned arrival acceptance rate A_t is set, A_t flights are expected to arrive at the destination airport no matter which scenario occurs. (Each of these arcs could have been superscripted with their respective scenarios (A_t^s), in which case we would require coupling constraints of the form $A_t^s = A_t^{s'}$, for $s \neq s'$.

The formulation presented here avoids the need for these constraints.) As a consequence, there could be more flow out of one of these nodes than into it.

For convenience, we define u_t^s to be the number of flights that arrive at time t under scenario s (this variable need not appear in the final formulation). This is the flow out the bottom of Figure 5 and is limited by the arrival capacity at time t under scenario s, b_t^s. Thus, we have that

$$u_t^s \leq b_t^s \qquad s = 1, \ldots, S, \qquad t = 1, \ldots, T \tag{2}$$

By conservation of flow at the lower nodes, variables of the form u_t^s can be defined as follows.

$$u_t^s = W_{t-1}^s + A_t + E_t - W_t^s \qquad s = 1, \ldots, S, \qquad t = 1, \ldots, T, \qquad W_0^s = 0 \tag{3}$$

Substituting (3) into (2), we obtain the following set of constraints.

$$W_{t-1}^s + A_t + E_t - W_t^s \leq b_t^s \qquad s = 1, \ldots, S, \qquad t = 1, \ldots, T, \qquad W_0^s = 0 \tag{4}$$

Since the flow across each arc represents a number of flights, this value must be integer and non-negative. In all, we have the following integer program.

(SGH)

$$Minimize \sum_{t=1}^{T} \left(c_g G_t + \sum_{s=1}^{S} c_a A_t^s \cdot p(s) \right) \tag{5}$$

subject to

$$(1), (4)$$

$$G_t \in \mathcal{Z}_+, A_t \in \mathcal{Z}_+, W_{s,t} \in \mathcal{Z}_+$$

Since the model treats flights on an aggregate level, the total number of constraints is independent of the number of flights and kept quite small: $O(S \cdot T)$. The number of variables is also $O(S \cdot T)$.

For the case $S = 1$, Figure 5 reduces to the diagram in Figure 4. The deterministic case $S = 1$ is easily solved by a greedy algorithm which lands as many flights as possible during each time period, remembering to take exempt flights into account. When $S \geq 2$, the problem is truly stochastic and the greedy algorithm is ill-defined because the phrase "land as many flights as possible" becomes ambiguous.

The model resembles a classic production-inventory model in which an inventory can be held in one of two states such as crude oil (see section 4.5 of [9]). A quantity of materials (the flights) is purchased when it enters the diagram at a time period and can then be held in raw form (ground holding) or finished form (airborne holding). The materials are sold when they exit the diagram at the bottom. But the model presented here is also distinct in three ways. First, it does not seek production levels to be set for each time period; these levels are set in advance by the parameters of the form D_t. Second, and more crucially, while the product can be held over (in two different types of inventory) at two different costs, as in a raw material/finished good production model such as crude oil/refined oil, the amount held over in the second form is stochastic, depending on a random variable. This prevents the model from being solved by standard techniques. (See [15] for a treatment of stochastic network routing.) Third, the model cannot be translated into a network model, even though it is depicted by a flow diagram. (The dual of the model, however, can be, as shown in section 4.)

2.1 Pre-processing

In this section, we show that for each tier e and each time interval t, the variables y_e and E_t can be eliminated from the formulation of SGH by pre-processing. We start with the variables y_e.

Part of the output of model SGH would be values for the variables y_e. From an optimization standpoint, the exemption of tier e from the program can only worsen the objective function because one has lost the ability to ground delay the flights of tier e. Therefore, there will be an optimal solution to SGH satisfying $y_e = 1$ for all e, as the following proposition shows:

Proposition 1 *For every feasible solution, $\mathbf{z}$, to SGH, there is a feasible solution, $\mathbf{z}^*$, with the properties that $F(\mathbf{z}^*) = F(\mathbf{z})$ and $y_e = 1$ for $e = 1, 2, ..., E$.*

PROOF: *Let $\mathbf{z}$ be any feasible solution to SGH with corresponding function value $F(\mathbf{z})$. If $y_e = 1$ for every e, then we set $z = z^*$ and there is nothing to show. Otherwise, we create a new solution, $\mathbf{z}^*$,as follows. For each index*

e such that $y_e = 0$ in solution $\mathbf{z}$, we set $y_e = 1$. By definition D_t, this increases the value of D_t by the amount d_t^e and decreases the value of E_t by the same amount. This jeopardizes equality in equation (1), but equality can be maintained by adding the amount d_t^e to the value of A_t. This also preserves feasibility in constraints (4). Since the values of W_t^s and G_t remain unchanged for every s and every t, we have created a new feasible solution, $\mathbf{z}^*$ with the same function value as $\mathbf{z}$. The result follows from the fact that y_e does not appear in the objective function and that the index e was chosen arbitrarily.$\square$

As a consequence of Proposition 1, we can make the simplifying assumption that $y_e = 1$ for $e = 1, 2, ..., E$ and discard these variables from the formulation of the problem. In practice, this means that if the values of the variables y_e are set a priori by the traffic flow manager (as opposed to being output by the solution), then optimality may be sacrificed.

We comment that proposition 1 would not hold in an alternative model in which the variables y_e appear in the objective function with associated penalties.

Now we show how to eliminate through preprocessing the variables of the form E_t, which reflect the number of flights scheduled to arrive at the airport at time t. For each scenario s, we perform the following procedure, shown here in pseudocode:

$$
\begin{array}{l}
t \leftarrow 1 \\
C_E \leftarrow 0 \\
E^* \leftarrow E \\
while(t < T)\{ \\
\quad E^*_{t+1} \leftarrow E^*_{t+1} + max(0, E^*_t - b^s_t) \\
\quad C_E \leftarrow C_E + c_a \cdot max(0, E^*_t - b^s_t) \\
\quad b^s_t \leftarrow b^s_t - \min(b^s_t, E^*_t) \\
\quad E^*_t \leftarrow 0 \\
\quad t \leftarrow t + 1 \\
\}
\end{array}
\tag{6}
$$

Intuitively, for each time period t, we subtract the number of exempt flights from the capacity for that time period. If there are more exempt flights than capacity, we set the capacity to zero and add any remaining exempt flights to the next time period. Once we have performed the above procedure for every scenario s, we set all the E_t to 0, yielding a new model. This model has the same set of feasible solutions as our original model, and the cost of any feasible solution in the two models differs by C_E, the total costs that must be added to offset those cases where "excess" exempt flights are moved forward from one time period to the next in the transformation — these flights would have experienced airborne delay in our original model. If we are interested only in the optimal solution itself, rather than the cost of the optimal solution, we can ignore the value C_E (see section 3).

Henceforth, we assume that in any formulation of SGH, there are no exempt flights so that $\bar{D}_t = D_t$, and that $y_e = 1$ for every e. We have the following simplification of SGH: (SGH_2)

$$\min \sum_{t=1}^{T} \left(c_g G_t + \sum_{s=1}^{S} c_a W_t^s \cdot p(s) \right)$$

$$\begin{gathered} \textit{subject to} \\ G_{t-1} + D_t = A_t + G_t \qquad t = 1, \ldots, T, G_0 = 0 \\ W_{t-1}^s + A_t - W_t^s \le b_t^s \qquad s = 1, \ldots, S,\ t = 1, \ldots, T, W_0^s = 0 \\ G_t \in \mathcal{Z}_+, A_t \in \mathcal{Z}_+, W_{s,t} \in \mathcal{Z}_+ \end{gathered}$$

3 Practical Implementation

In this section, we discuss some of the issues that might arise in the practical implementation of our model.

3.1 Holding Costs

Our model requires that both a ground holding cost and an airborne holding cost be fixed uniformally for all flights. By nature of the objective function, it would suffice to specify a ratio of these costs, $\frac{c_a}{c_g}$. Moreover, one could set $c_g = 1$. The two issues at hand are exactly what this cost ratio represents and how it would be set.

The most obvious and direct intrepretation of the cost ratio is the relative costs of ground delay and airborne delay that are actually incurred by the airlines. Note that our model assumes a uniform ground holding cost and a uniform airborne holding cost for each flight. This is similar to the approach used by Richetta in [16], in which he proposed that flights be grouped into a small number (e.g., three) of classes and that each class would have a single ground delay cost and a single airborne delay cost representative of the whole class. There are figures to support the implementation of this approach. For instance, the Air Transportation Association [1] estimates that for a typical flight, every minute of ground delay costs \$20.35 while every minute of airborne delay costs \$45.85. But these figures are averages and do not reflect complex factors that vary day to day such as crew costs, fuel consumption, connectivity to other flights, and passenger safety. These types of factors may be difficult to quantify, especially in real time.

The primary mission of the FAA, hence, the ATCSCC (Air Traffic Control Systems Command Center), is to ensure passenger safety while simultaneously managing resources within the NAS (national airspace system) in an efficient, even-handed manner. The traffic flow manager at the ATCSCC must address airline needs and perceptions. In particular, an airline finds it aggravating when a GDP is aborted in mid-operation. This happens whenever the traffic flow manager has clear evidence that the original capacity and weather forecasts were overly pessimistic. For this reason, the airlines are frequently willing to risk small or moderate amounts of airborne delay and tend to favor relatively liberal (i.e., not particularly aggresive) ground holding policies.

These two objectives are inherently somewhat in conflict. A ground delay program which is aggressive enough to ensure that, under any reasonable circumstances, no flight is subjected to airborne holding would, on average, lead to a gross under-utilization of airport arrival capacity and the services provided by the air traffic controllers. On the other hand, running no ground delay programs whatsoever would allow any available airport arrival capacities to be maximally exploited, but this would clearly pose unacceptable risks to passenger safety due to potentially excessive airborne holding. The traffic flow manager must strike a balance between restricted traffic flow and efficient allocation of resources.

The alternative interpretation of the holding cost ratio is that of a numerical expression of the willingness to trade ground holding for airborne holding, in light of the twin objectives of safety and manageability on the one hand and efficient resource use on the other. The higher the cost ratio, the more conservative would be the resulting GDP. In particular, we envision that the traffic flow manager at the ATCSCC might explore several values of the cost ratio before implementing a ground delay program.

3.2 Variable Time Lengths

The structure of the objective function allows the delay costs to vary with t. In particular, there could be a ground holding cost, $c_{g\,t}$, and an airborne holding cost, $c_{a\,t}$, for each time period t. This grants the user of our model the option to establish time intervals of varying length. This flexibility could be used to compensate for uncertainty in arrival capacities of the airport in time periods farther away from the time at which the GDP is initially formed. For instance, one could establish periods of length 15 minutes for $t = 1$, 30 minutes for $t = 2$, 60 minutes for $t = 3$, and so on. Since the final output of the model is the number of flights that should be issued arrival times in each time period, this allows for detailed planning in the nearer time periods and coarser planning in the more distant time periods. Refinements to controlled times of arrival could then be made as the GDP is carried out.

4 Theoretical Results

In this section, we show that the integer programming problem SGH_2 can be solved to optimality in polynomial time. In particular, we will show that SGH_2 is a dual network flow problem. As immediate corollaries, the constraint matrix associated with SGH_2 is totally unimodular, and the LP relaxation of SGH_2 yields integral solutions.

We say that an m row, n column $(0, 1, -1)$ matrix M is a *network matrix* if there exists a directed tree R on $m+1$ nodes and a one-to-one mapping of the rows of our matrix onto the edges of R with the property that each column of the matrix corresponds to the characteristic vector of a path in R. This definition was introduced in [23] and is explored in [13]. The more familiar *node-arc incidence matrices*, with a single 1 and -1 in each column, are a special case of this construction. Network matrices are desireable because they are totally unimodular and give rise to integral polyhedra.

Theorem 1 *Let A^T be the transpose of the constraint matrix associated with SGH_2. A^T is a network matrix.*

Proof Sketch: *We will illustrate the construction of the tree R for the case of two time periods and two scenarios ($T = 2, Q = 2$). It will be clear* *that this construction can be extended to an arbitrary number of time periods and scenarios. The general construction is explicitly formulated as a matrix transformation in [10], providing a more formal proof.*

For a two time period, two scenario problem, with slack variables, the primal problem is:

$$
\begin{array}{llllllllllllll}
\min & & & c_g \cdot G_1 + c_g \cdot G_2 + c_a \cdot W_{11} + c_a \cdot W_{12} + c_a \cdot W_{21} + c_a \cdot W_{22} & & & & & & & & & & \\
A_1 & & +G_1 & & & & & & & & & & = & D_1 \\
& A_2 & -G_1 & +G_2 & & & & & & & & & = & D_2 \\
-A_1 & & & & +W_{11} & & & & -S_{11} & & & & = & -M_{11} \\
& -A_2 & & & -W_{11} & +W_{12} & & & & -S_{12} & & & = & -M_{12} \\
-A_1 & & & & & & +W_{21} & & & & -S_{21} & & = & -M_{21} \\
& -A_2 & & & & & -W_{21} & W_{22} & & & & -S_{22} & = & -M_{22}
\end{array}
$$

We write the matrix associated with this problem as:

$$
A = \begin{bmatrix}
A_1 & & G_1 & & & & & & & & & \\
& A_2 & -G_1 & G_2 & & & & & & & & \\
-A_1 & & & & W_{11} & & & & -S_{11} & & & \\
& -A_2 & & & -W_{11} & W_{12} & & & & -S_{12} & & \\
-A_1 & & & & & & W_{21} & & & & -S_{21} & \\
& -A_2 & & & & & -W_{21} & W_{22} & & & & -S_{22}
\end{bmatrix}
$$

All the entries in A are 0, 1, or -1; the variable names have been added for clarity. To show that A^T is a network matrix, we must find a directed tree R on 13 nodes such that every column of A^T (equivalently, every row *of A) corresponds to the characteristic vector of a path in R. Figure 7 shows the required tree. We have labeled each arc of R with the corresponding primal variable. It can be verified that every row of A corresponds to the characteristic vector of a path in R. For example, the fourth row of A corresponds to the path formed by traversing arc S_{12} backward, then arc W_{12} forward, then arcs A_2 and W_{11} backward. Figure 8 shows the graph G spanned by the tree R. We label each arc with the associated dual cost coefficient (in the primal, the associated RHS element). This construction extends to an arbitrary of number of scenarios and time periods.*

Corollary 2 *The constraint matrix associated with SGH_2 is totally unimodular.*

Corollary 3 *The LP relaxation of SGH_2 yields integral solutions.*

Corollary 4 *The integer program SGH_2 can be solved to optimality in polynomial time.*

The dual problem can be recast in a more familiar form as a min-cost flow problem by negating the objective function and multiplying the dual LP in equality form by N, where N is the node-arc incidence matrix associated with R with a single row deleted. Figure 9 shows the resulting flow problem for the previous example ($T = 2, Q = 2$). The labeled arcs can accomodate either positive or negative flow, as they are associated with equality constraints in the original primal problem; the flow must be nonnegative on the unlabeled arcs, which are associated with dual slack variables. The unlabeled arcs are zero-cost arcs and the unlabeled nodes are neither sources nor a sinks.

We note that the primal problem cannot be recast as a network flow problem. This can be verified by applying standard network recognition algorithms as in [8] or [13]. See [10] for further discussion of this issue. In [20], it is shown that scenario-stochastic min-cost flow problems in outerplanar graphs can always be recast as dual min-cost flow problems.

5 Computational Results and Implementation

We constructed three realistic instances of SGH_2 using the following guidelines. Each test case was solved for two levels of granularity of time periods (60 minutes per time period and 15 minutes per time period, respectively), for a total of six test cases. For each instance, we constructed three AAR scenarios. In test case 1, the AAR's were randomly generated numbers in the range 30 - 60. In test case 2, each AAR scenario drops from 60 flights per hour to 40 flights per hour for several hours, then returns to 60 flights per hour. The times were varied at which each AAR scenario drops and recovers. In test case 3, the first scenario is designed to be a fair weather scenario, the second is a bad weather scenario and the third is randomly generated.

In each instance, the costs used were 2.0 for ground holding and 5.0 for airborne holding. Each instance was solved using CPLEX 3.0 with the default options on a SPARC STATION 10.

The results in Table 1 show that the problem is highly tractable. The integer solution was obtained in zero nodes of the mixed integer program algorithm of CPLEX, meaning that the integer solution was obtained directly from the linear program relaxation. This is to be expected, given Corollary 4 to Theorem 1. The largest number of iterations of the simplex procedure was 289 and the longest run time was barely more than half a second. Note that the run time is almost linear in the coarseness (length of) the time periods.

Test Case	Numb flights	T = Numb periods	Min per period t	Opt Func Value	Time (sec)	Simplex Iteration	Node B&B
1a	624	12	60	642.00	0.07	51	0
1b	624	48	15	3329.00	0.30	267	0
2a	624	12	60	1088.00	0.07	59	0
2b	624	48	15	4209.00	0.37	289	0
3a	624	12	60	485.00	0.05	55	0
3b	624	48	15	1745.50	0.20	223	0

Table 1: SGH model performance

We would expect that an implementation that directly exploited the dual network structure of the problem would result in substantially faster runtimes for larger problems. Because the LP model is solved so quickly on problem sizes currently considered realistic, this is unnecessary. Our goal is to deliver a model to the FAA which can run "at the push of a button" using only modest computing power; the current implementation achieves these goals.

Ideally, we would hope that a small increase in the cost ratio would lead to small changes in the optimal solution. However, under certain circumstances, the model exhibits behavior in which the optimal solution is constant over a fairly wide range of cost ratios, then, when the cost ratio reaches some critical value, the character of the optimal solution changes completely. An extreme example of this effect can be seen when capacity is constant within each scenario. In this case, the optimal solution will always mimic one of the scenarios precisely. For a study of this and several other examples illustrating the behavior of optimal solutions to SGH_2, see either [10] or [19]. We note that this effect should be considered pathological — for a reasonable number of substantively different scenarios, the effect will not be observed.

6 Conclusion

The stochastic ground holding model presented in this paper finds the optimal trade-off between airborne-holding and ground holding in the formulation of a ground delay program. By treating flights on an aggregate level, as a flow, the integer programming model avoids the need to consider the assignment of every flight to each time period in the planning horizon. Additionally, the fact that the model is dual network allows us to relax the integrality constraints, as we are guaranteed integer optimal solutions directly from the LP relaxation. The combination of these two facts result in a highly tractable model which requires only modest computing power. Additionally, the model is designed to be an integral component of the collaborative decision-making process, and can be easily integrated into existing tools.

Our model assumes the production of multiple arrival acceptance rate scenarios with associated probabilities. The generation of the scenarios and their probabilities represents a unique research challenge. This requires fusion of weather forecasts with optimal runway configurations using a combination of statistical techniques and analysis of the impact of meteorological conditions on incoming aircraft. Moreover, the interaction of incoming aircraft with departing aircraft could lead to many interesting problems in optimization.

References

[1] "Aircraft Operating Costs by Stage of Flight and Associated Air Traffic Control Delay Costs Based on a Typical Flight," 1993, *Air Transportation Association*, Washington, 1995.

[2] Andreatta, G., and G. Romanin-Jacur. 1987, "Aircraft Flow Management Under Congestion," *Transportation Science*, **21**, 249-253.

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

[4] Attwool, V.W. 1977, "Some Mathematical Aspects of Air Traffic Systems," J. Inst. Navigation., **30**, 394-411.

[5] Bertsimas, D. and S. Stock. 1996, "Air Traffic Flow Management Problem with Enroute Capacities," to appear in *Operations Research.*

[6] Ball, M., R. Dahl, L. Stone, and T. Thompson. 1993, "OPTIFLOW Build-I Design Document, Version 1.5," *Optiflow Project Report to FAA*, December.

[7] Birge, J. and F. Louveaux. 1997, *Introduction to Stochastic Programming*, Springer-Verlag, New York, New York.

[8] Bixby, R.E. and W.H.Cunningham. 1980, "Converting Linear Programs to Network Flow Problems," *Math. of Operations Research*, August, 483-508.

[9] Glover, F., D. Klingman, and N. Phillips. 1992, *Network Models in Optimization and Their Applications in Practice*, John Wiley and Sons, Inc., New York, New York.

[10] Hoffman, R. 1997, "Integer Programming Models for Ground-Holding in Air Traffic Flow Management," Ph.D. Dissertation, University of Maryland.

[11] Hoffman, R. and M. Ball. 1997, "A Comparison of Formulations for the Single-Airport Ground Holding Problem with Banking Constraints," Submitted for publication.

[12] Infanger, G. 1994, *Planning Under Uncertainty: Solving Large-Scale Stochastic Linear Programs.* Boyd and Fraser publising company, Danvers, Massachusetts.

[13] Nemhauser, G.L. and L.A.Wolsey. 1988, *Integer and Combinatorial Optimization*, John Wiley and Sons, Inc., New york, New York.

[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., 269-288, Springer-Verlag, Berlin.

[15] Powell, Warren B., P. Jaillet and A. Odoni. 1995, "Stochastic and Dynamic Networks and Routing," *Handbook in Operations Research and Management Science*, **8**, M. O. Ball, et. al., editors, Elsevier Science, Berlin.

[16] Richetta, O. and A.R. Odoni 1993. "Solving Optimally the Static Ground-Holding Policy Problem in Air Traffic Control," *Transportation Science,* **27**, 228-238.

[17] Richetta, O. 1995, "Optimal Algorithms and a Remarkably Efficient Heuristic for the Ground-Holding Problem in Air Traffic Control," *Operations Research*, **43**, 758-770.

[18] Richetta, O. 1991, "Ground Holding Strategies for Air Traffic Control under Uncertainty,", Ph.D. dissertation, Massachusetts Institute of Technology.

[19] Rifkin, R. 1998, "The Static Stochastic Ground Holding Problem," Masters Thesis, Massachusetts Institute of Technology.

[20] Rifkin, R. 1998, "Scenario-Stochastic Min-Cost Flows", Manuscript in Preparation.

[21] Sokkapia, B. G. 1985, "Arrival Flow Management as a Feedback Control System," The Mitre Corporation, Washington, D.C.

[22] Terrab, M. and A.R. Odoni. 1993, "Strategic Flow Management for Air Traffic Control," *Operations Research*, **41**, 138-152.

[23] Tutte, W. T. 1965, "Lectures on Matroids", *Journal of Research of the National Bureau of Standards B*, **69B**, 1-47.

[24] Vranas, P., D. Bertsimas, and A. R. Odoni. 1994, "The Multi-Airport Ground-Holding Problem in Air Traffic Control," *Operations Research,* **42**, 249-261.

[25] Vranas, P., D. Bertsimas, and A. R. Odoni. 1994, "Dynamic Ground-Holding Policies for a Network of Airports," *Transportation Science,* **28**, 275-291.

[26] Wang, H. 1991, "A Dynamic Programming Framework for the Global Flow Control Problem in Air Traffic Management," *Transportation Science,* **25,** 308-313.

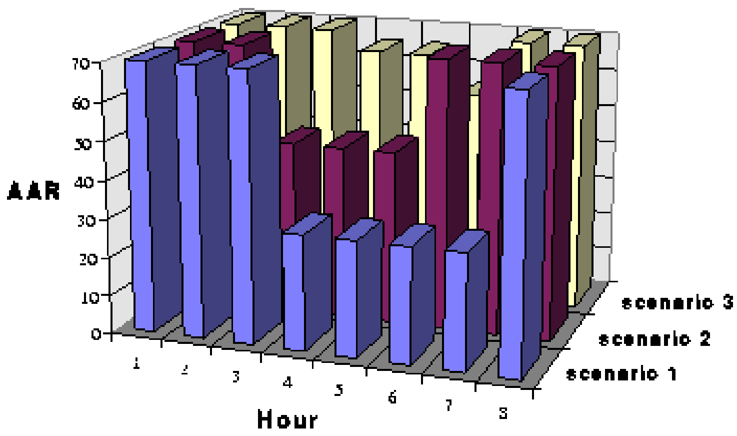

Figure 1: Multiple AAR scenarios

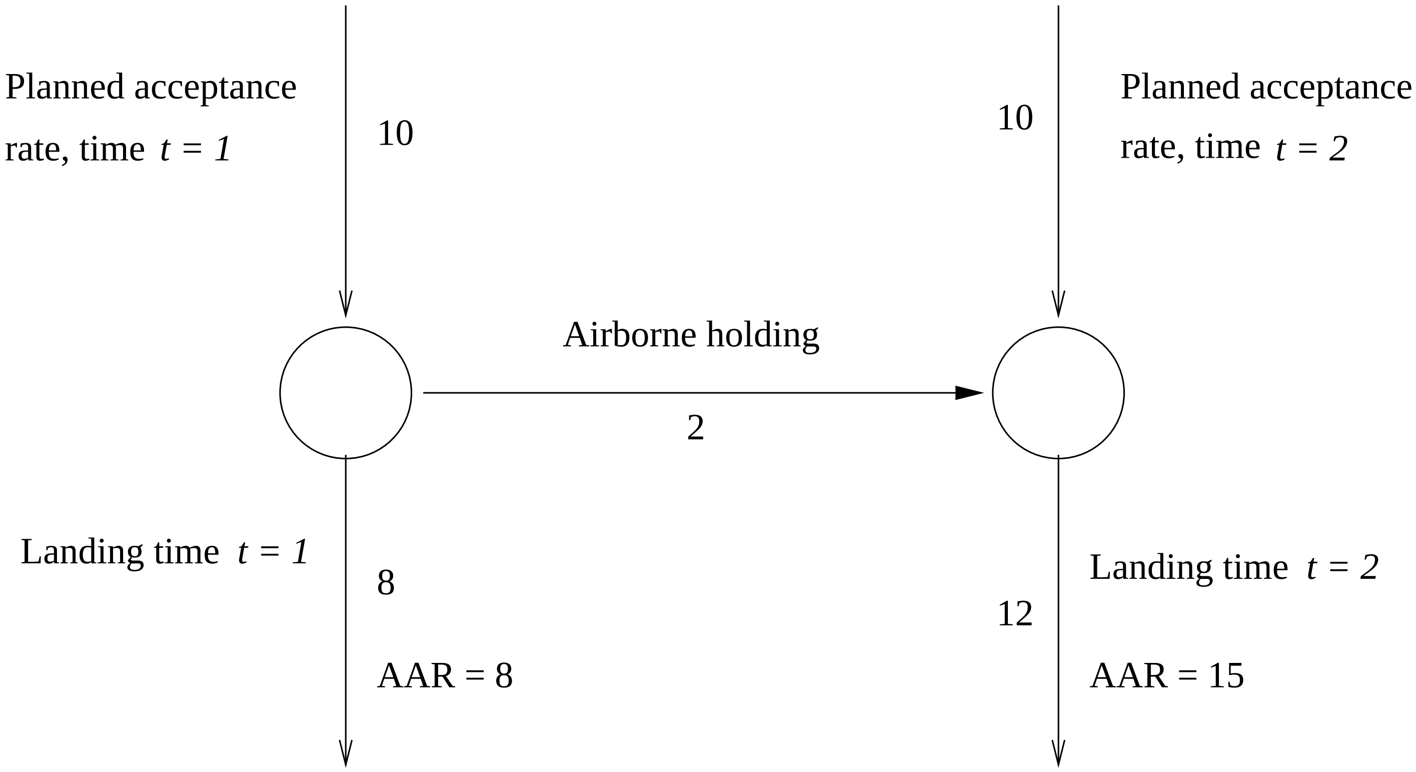

Figure 2: Airborne Holding of Two flights

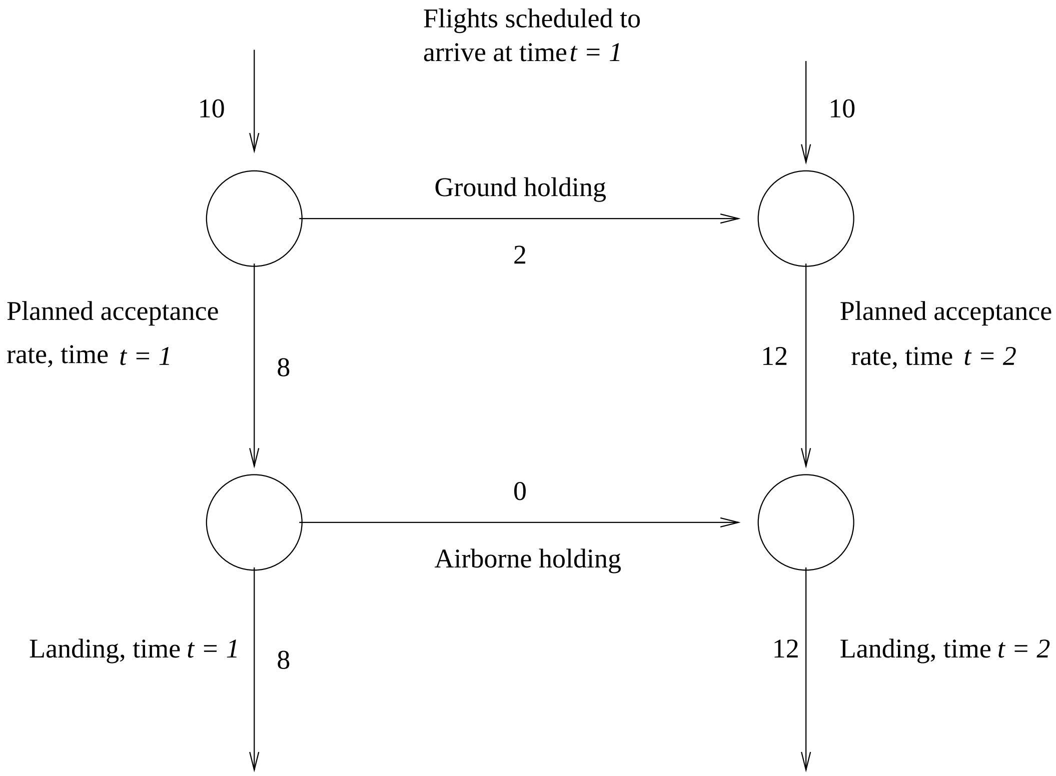

Figure 3: Ground Holding of Two Flights

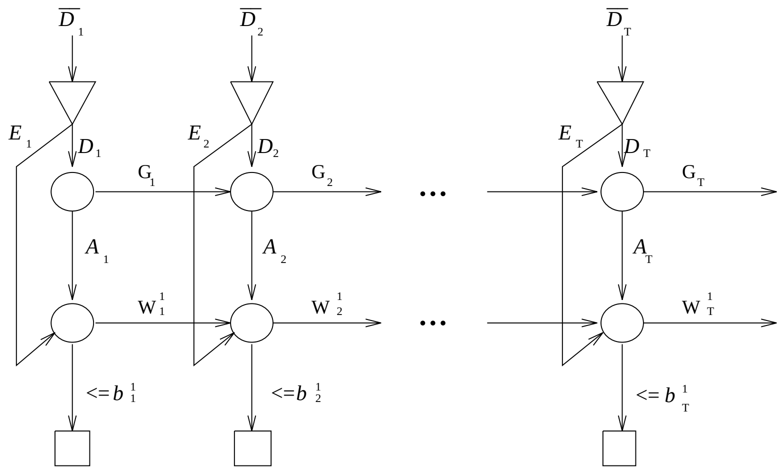

Figure 4: Deterministic Flow Model (S = 1)

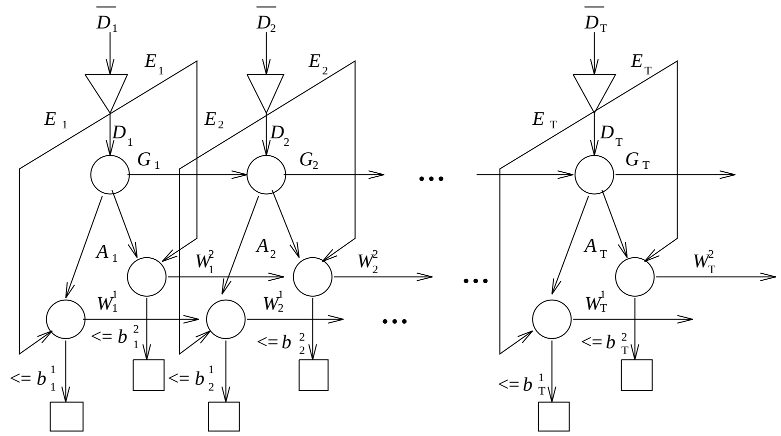

Figure 5: Stochastic Flow Model (S = 2)

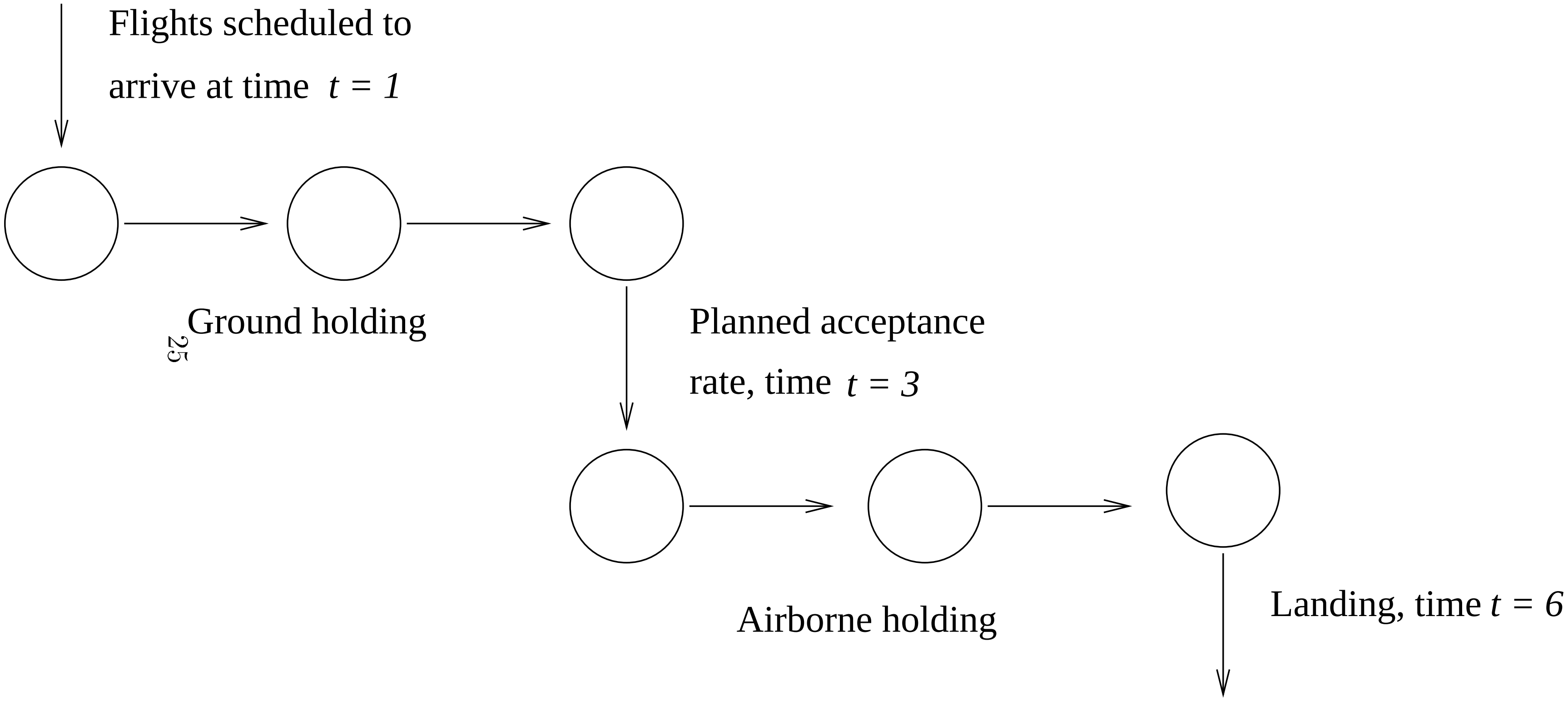

Figure 6: The Path of a Single Flight

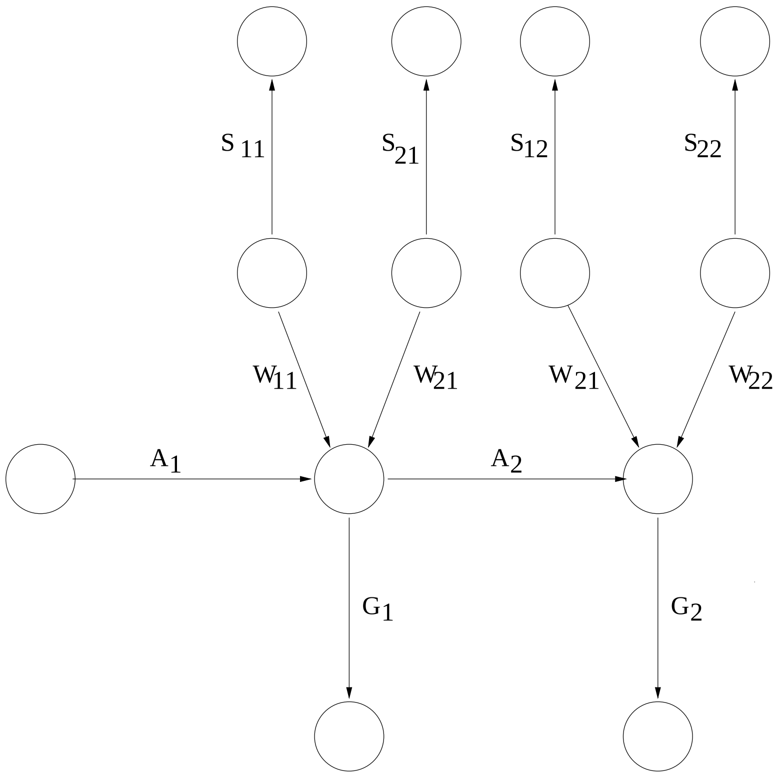

Figure 7: The Dual Tree

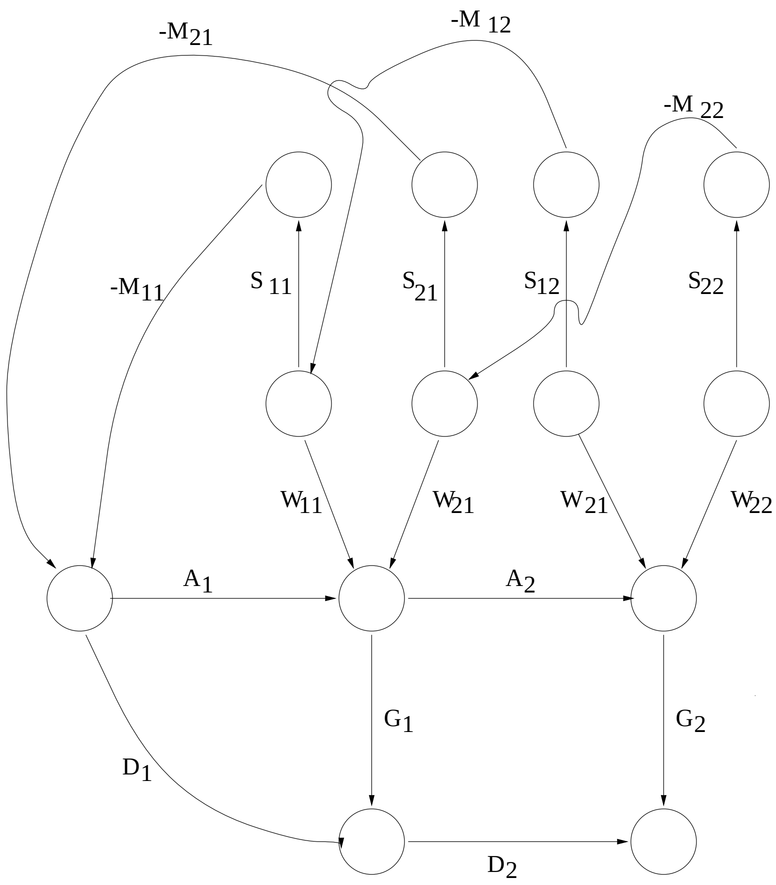

Figure 8: The Dual Tree With Columns

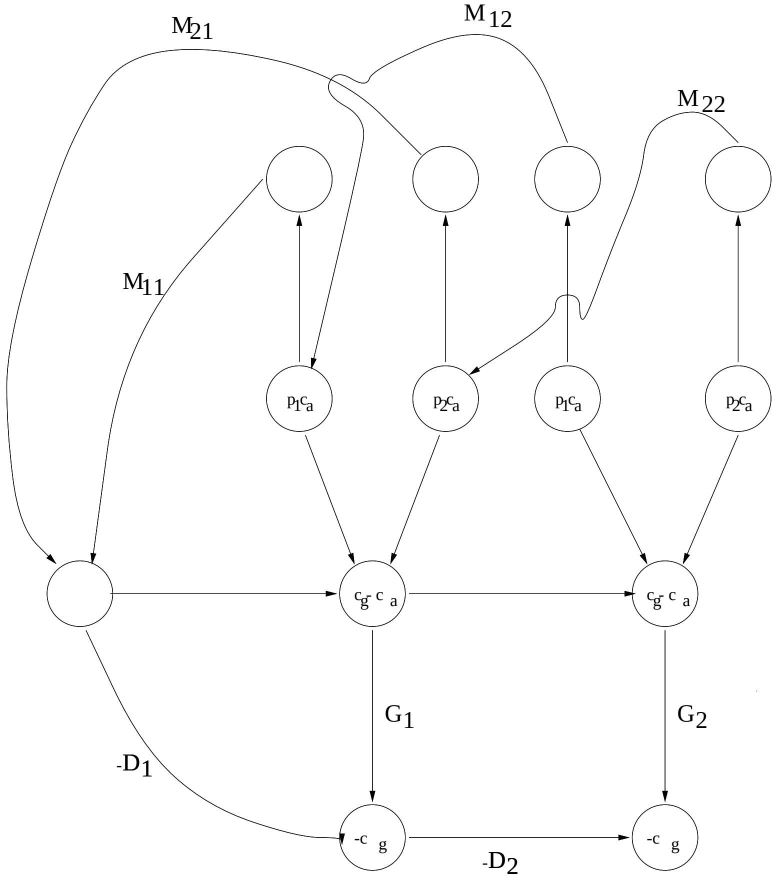

Figure 9: Min Cost Flow