THE MULTI-AIRPORT GROUND-HOLDING PROBLEM IN AIR TRAFFIC CONTROL

PETER B. VRANAS, DIMITRIS J. BERTSIMAS and AMEDEO R. ODONI

Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received February 1992; revision received October 1992; accepted February 1993)

Motivated by the important problem of congestion costs (they were estimated to be \$2 billion in 1991) in air transportation and observing that ground delays are more preferable than airborne delays, we have formulated and studied several integer programming models to assign ground-holding delays optimally in a general network of airports, so that the total (ground plus airborne) delay cost of all flights is minimized. All previous research on this problem has been restricted to the single-airport case, which neglects "down-the-road" effects due to transmission of delays between successive flights performed by the same aircraft. We formulate several models, and then propose a heuristic algorithm which finds a feasible solution to the integer program by rounding the optimal solution of the LP relaxation. Finally, we present extensive computational results with the goal of obtaining qualitative insights on the behavior of the problem under various combinations of the input parameters. We demonstrate that the problem can be solved in reasonable computation times for networks with at least as many as 6 airports and 3,000 flights.

Congestion problems are becoming increasingly acute in many major European and American airports. For European airlines, the total yearly delay cost due to congestion (including cost to passengers) was estimated to be \$5 billion in 1989 (Terrab 1990). For U.S. airlines, the direct delay cost due to congestion is claimed to amount to approximately \$2 billion per year. Given the fact that the total profits of the U.S. airline industry rarely exceed \$1 billion, congestion problems are a phenomenon of undeniable significance.

Limited capacity is the major cause of congestion. The problem with airport capacity is that it is highly variable, because it is heavily influenced by, among other factors, weather conditions (visibility, wind, precipitation). It is not unusual to encounter 2:1 and even 3:1 ratios between the highest and the lowest capacity of an airport.

Solution approaches to this problem vary according to the contemplated time horizon. Long-term approaches include construction of additional airports, construction of additional runways at existing airports, improved air traffic control technologies and procedures and use of larger aircraft. Medium-term approaches include modification of the temporal pattern of aircraft flow to eliminate periods of "peak" demand. Short-term approaches have a planning horizon of 6–12 hours and include, most importantly, *ground-holding* policies. These policies are motivated by the fundamental fact that airborne delays are much costlier than ground delays, because the former include fuel, maintenance, depreciation, and safety costs. Thus, the premise underlying ground-holding policies is that one may hold an aircraft on the ground before take-off so that, when the aircraft arrives at its destination, it will not have to wait in the air before landing.

Ground-holding has been in use for several years. The Federal Aviation Administration operates an Air Traffic Control System Command Center (ATCSCC, formerly called the Central Flow Control Facility) in Washington, D.C., equipped with outstanding information-gathering capabilities. ATCSCC, however, relies primarily on the judgment of its expert air traffic controllers rather than on any decision-support or optimization models to develop flow management and ground-holding strategies.

The problem of determining how much (if at all) each aircraft must be held on the ground before take-off (and also, possibly, in the air during the flight, e.g., by means of a speed reduction *en route*) to minimize the total (ground plus airborne) delay cost will be referred to as the ground-holding problem (**GHP**). *Static* and *dynamic* versions of the **GHP** can be distinguished. In the static versions, the ground (and airborne) holds are decided once at the beginning of

Subject classifications: Programming: integer, applications. Transportation: air traffic.
Area of review: DISTRIBUTION, TRANSPORTATION AND LOGISTICS.

0030-364X/94/4202-0249\$01.25

the day, whereas in the dynamic versions they are updated during the course of the day as better weather (and, hence, capacity) forecasts become available. *Deterministic* and *probabilistic* versions of the **GHP** can also be distinguished, according to whether airport capacities are considered deterministic or probabilistic.

Because each of a large number of aircraft performs more than one flight on any given day, "network" (or "down-the-road") effects may be important: When a specific aircraft is delayed, in many cases the next flight performed by the same aircraft will also be delayed. Moreover, at a hub airport, a late-arriving aircraft may delay the departure of several flights, given current airline scheduling practices which emphasize passenger transfers. To the best of our knowledge, previous research on the **GHP** has neglected network effects, and has been restricted to the single-airport problem. Odoni (1987) seems to be the first to have given a systematic description of the problem. Andreatta and Romanin-Jacur (1987) proposed a dynamic programming algorithm for the single-airport static probabilistic **GHP** with one time period. Terrab proposed an efficient algorithm to solve the single-airport static deterministic **GHP**, as well as several heuristics for the single-airport probabilistic **GHP**. He also suggested a two-airport formulation and a closed three-airport formulation for the static deterministic **GHP**. Finally, Richetta (1991) dealt with the single-airport dynamic probabilistic **GHP**. It seems that no significant research has been done to date concerning the effects of ground-holding policies on an entire network of airports.

In this paper, the multi-airport **GHP** is addressed for the first time. By using a mathematical programming approach, we solve the *deterministic* network **GHP** in a general setting. We propose several integer programming formulations which have the important advantages of being remarkably simple, while capturing the essential apects of the problem, and sufficiently flexible to accommodate various degrees of modeling detail. We present several structural insights on the parameters that influence the problem, based on extensive computational experience. Most importantly, our approach enables one to solve realistic-size problems involving, e.g., 6 airports and 3,000 flights in reasonable computation times. Our approach can thus be used to assign ground holds for at least a major part of the network of the most important U.S. or European airports. Although we focus on the *static* multi-airport **GHP**, our algorithms could also be used dynamically by solving the problem, say, every two hours, as better capacity estimates become available.

The outline of this paper is as follows. Section 1 defines the problem and gives integer programming formulations of three versions of it. Section 2 proposes a heuristic based on the solution of a linear programming relaxation. Section 3 gives insights on the parameters influencing the behavior of the problem, based on an extensive series of actual runs. Finally, Section 4 summarizes the results of the paper and points out directions for future research.

1. PROBLEM DEFINITION AND FORMULATIONS

1.1. Notation

Consider a set of airports $\mathcal{K} = \{1, \ldots, K\}$ and an ordered set of time periods $\mathcal{T} = \{1, \ldots, T\}$. For instance, $\mathcal{K}$ might be the set of the 20 or so busiest U.S. airports, and $\mathcal{T}$ might be a set of 64 time periods of 15 minutes each, amounting to a time horizon of 16 hours, i.e., the portion of a day from 7 a.m. to 11 p.m. (when most flights take place). Consider, finally, a set of flights $\mathcal{F} = \{1, \ldots, F\}$. (Note that a single aircraft may perform several of these flights.) Here $\mathcal{F}$ is the set of all flights of interest, e.g., all flights departing from an airport in $\mathcal{K}$ and arriving to another airport in $\mathcal{K}$. This interpretation of $\mathcal{F}$ corresponds to a *closed* network of airports, for which departures from and arrivals to the external world are not considered important. If an *open* network of airports is to be considered, then $\mathcal{F}$ will be the set of all flights departing from an airport in $\mathcal{K}$ *or* arriving to an airport in $\mathcal{K}$ (or both).

For each flight $f \in \mathcal{F}$, the following data are assumed to be known: $k_f^d \in \mathcal{K}$, the airport from which f is scheduled to depart; $k_f^a \in \mathcal{K}$, the airport to which f is scheduled to arrive; $d_f \in \mathcal{T}$, the scheduled departure time of f; $r_f \in \mathcal{T}$, the scheduled arrival time of f; $c_f^g(\cdot)$, the ground delay cost function of f (whose argument is the ground delay of f in time periods); and $c_f^a(\cdot)$, the airborne delay cost function of f (whose argument is the airborne delay of f in time periods). For each $(k, t) \in \mathcal{K} \times \mathcal{T}$, the departure capacity $D_k(t)$ and the arrival capacity $R_k(t)$ (in number of aircraft) are also given. Since this paper deals with *deterministic* versions of the **GHP**, these capacities are considered fixed numbers rather than random variables.

Consider the set $\mathcal{F}' \subset \mathcal{F}$ of those flights that are *continued*. A flight is continued if the aircraft which is scheduled to perform it is also scheduled to perform at least one more flight later in the day. For each flight $f' \in \mathcal{F}'$, we assume that we know the next flight f scheduled to be performed by the same aircraft, and the "slack" or "absorption" time $s_{f'}$ such that, if f'

arrives at its destination at most $s_{f'}$ time periods late, the departure of the next flight f will not be affected. Then $s_{f'}$ is obviously equal to the difference between the time interval between the scheduled departure time of f and the scheduled arrival time of f'; and the minimum "turnaround" time of the aircraft performing both flights.

1.2. Preliminary Remarks

We define the decision variables g_f, $f \in \mathscr{F}$, equal to the number of time periods that flight f is held on the ground before being allowed to take-off, and the decision variables a_f, $f \in \mathscr{F}$, equal to the number of time periods that flight f is further held in the air (e.g., by means of an *en route* speed reduction) before being allowed to land. Since this paper deals with *static* versions of the **GHP**, we assume that these ground and airborne holds are decided once at the beginning of the day for all flights.

Consider the following description of the real-world situation. If a flight f is scheduled to depart at period d_f and is delayed on the ground for g_f periods, then it will be *available to depart* at period $d_f + g_f$. Will it actually depart at that period? This will depend on whether the total number of aircraft *available to depart* from airport k_f^d at that time period will exceed (or not) the available departure capacity. If it does exceed it, then the aircraft performing flight f will have to wait q_f^d time periods in the *departure queue*. Here q_f^d will depend on the particular *service discipline* adopted for the departure queue. So flight f will actually take-off at period $d_f + g_f + q_f^d$. Since flight f will be further delayed in the air for a_f time periods, it will arrive at its destination, airport k_f^a, and will be *available to land* at period $r_f + g_f + q_f^d + a_f$. Will it actually land at that period? This will depend on whether the total number of aircraft *available to land* at airport k_f^a at that period will exceed (or not) the available landing capacity. If it does exceed it, then the aircraft performing flight f will have to wait q_f^a time periods in the *arrival queue*, and will actually land at period $r_f + g_f + q_f^d + a_f + q_f^a$. The total cost corresponding to flight f will be the sum of $c_f^g(g_f + q_f^d)$ (the ground delay cost) and $c_f^a(a_f + q_f^a)$ (the airborne delay cost).

Because we are examining the deterministic case, the above description can be considerably simplified. It makes little sense to assign to a flight f a ground hold of g_f time periods such that f will have to further wait q_f^d time periods in the departure queue: One might as well assign to f a total ground hold of $g_f + q_f^d$ time periods such that f will not have to wait in the departure queue. Similar remarks hold for airborne delays. Given this simplification, the total ground delay of flight f will be g_f, and its total airborne delay will be a_f, resulting in a cost of $c_f^g(g_f) + c_f^a(a_f)$.

1.3. A Pure 0-1 Integer Programming Formulation of the Multi-Airport GHP

The *delay decision variables* g_f and a_f were introduced before. Now we introduce the *assignment decision variables* u_{ft}, defined to be 1 if flight f is assigned to take-off at period t (i.e., if $r_f + g_f = t$) and 0 otherwise, and v_{ft}, defined to be 1 if flight f finally is assigned to land at period t (i.e., if $r_f + g_f + a_f = t$) and 0 otherwise. These new decision variables are introduced because the capacity constraints cannot be expressed in a simple linear way in terms of the more natural delay decision variables.

Moreover, since we do not want to have excessive ground or airborne delays, we introduce upper bounds on those delays. Here G_f is the maximum number of time periods that flight f may be held on the ground, and A_f is the maximum number of time periods that flight f may be held in the air. Introduction of these bounds results in no loss of generality, because they can be arbitrarily large. In practice, however, typical values are G_f = 4-5 and A_f = 2-3, corresponding to maximum ground and airborne delays of about one hour and half an hour, respectively.

Given this setup, the set $\mathscr{T}_f^d$ of time periods to which flight f may be assigned to take-off is given by:

$$\mathscr{T}_f^d = \{t \in \mathscr{T}: d_f \leqslant t \leqslant \min(d_f + G_f, T)\}. \tag{1}$$

Similarly, the set $\mathscr{T}_f^a$ of time periods to which flight f may be assigned to land is given by:

$$\mathscr{T}_f^a = \{t \in \mathscr{T}: r_f \leqslant t \leqslant \min(r_f + G_f + A_f, T)\}. \tag{2}$$

For every flight f, exactly one of the variables u_{ft} must be equal to 1 and the others must be equal to zero, and similarly for the variables v_{ft}. Given this fact, the delay variables g_f and a_f can be expressed in terms of the assignment variables u_{ft} and v_{ft}:

$$g_f = \sum_{t \in \mathscr{T}_f^d} t u_{ft} - d_f, \quad f \in \mathscr{F}; \tag{3}$$

$$a_f = \sum_{t \in \mathscr{T}_f^a} t v_{ft} - r_f - g_f, \quad f \in \mathscr{F}. \tag{4}$$

We are now ready to give a first pure 0-1 integer programming formulation of the static deterministic multi-airport **GHP**.

Problem P_1

$$\text{Minimize} \sum_{f=1}^{F} (c_f^g g_f + c_f^a a_f)$$

subject to

$$\sum_{f:k_f^d=k} u_{ft} \leq D_k(t), \quad (k, t) \in \mathcal{K} \times \mathcal{T}; \tag{5}$$

$$\sum_{f:k_f^a=k} v_{ft} \leq R_k(t), \quad (k, t) \in \mathcal{K} \times \mathcal{T}; \tag{6}$$

$$\sum_{t \in \mathcal{T}_f^d} u_{ft} = 1, \qquad f \in \mathcal{F}; \tag{7}$$

$$\sum_{t \in \mathcal{T}_f^d} v_{ft} = 1, \qquad f \in \mathcal{F}; \tag{8}$$

$$g_{f'} + a_{f'} - s_{f'} \leq g_f, \qquad f' \in \mathcal{F}'; \tag{9}$$

$$a_f \geq 0, \qquad f \in \mathcal{F}; \tag{10}$$

$$u_{ft}, v_{ft} \in \{0, 1\}.$$

In the objective function of $\mathbf{P}_1$, the cost functions $c_f^g(t)$, $c_f^a(t)$ were replaced by their linear counterparts $c_f^g t$, $c_f^a t$ (c_f^g, c_f^a being the constant marginal costs). (The assumption of linear cost functions is an approximation which, however, is widely used by the FAA and throughout the airline industry, for lack of a better alternative.) Constraints (5) and (6) are the departure and arrival capacity constraints, respectively. Recall that these have to be satisfied because we choose g_f and a_f such that the queueing delays q_f^d, q_f^a are 0 (we can do this because the problem is deterministic). (Strictly speaking, we also need the condition that G_f and A_f be sufficiently large.) Constraints (7) (together with 3) ensure that, for a given f, exactly one u_{ft} will be 1 and the rest will be 0. Similarly for (8).

Constraints (9) are the *coupling* constraints: They "transfer" any excessive delay of flight f' to its next flight f. The coupling constraints say that, if flight f' arrives at its destination with a total delay $g_{f'} + a_{f'}$ which is greater than $s_{f'}$ (the "slack" defined above), then the next flight f will have to be delayed on the ground at least $g_{f'} + a_{f'} - s_{f'}$ time periods; otherwise, the departure of the next flight f will not be affected. Note that the existence of these coupling constraints allows us to have a separable objective function: The cost of delaying flight f because of an excessive delay of its previous flight f' is taken into account via the term of the objective function corresponding to f (i.e., $c_f^g g_f$), and so need not be included in the term corresponding to f'. Also, if the coupling constraints did not exist the problem would be decomposable into K subproblems concerning one airport each, so that one could use the already existing techniques to solve for each of the K airports separately. A final interesting remark concerning the coupling constraints is that they can be interpreted in a more general way than the linking of successive flights scheduled to be performed by the same aircraft; i.e., they can be used to link any pair of flights f' and f such that f cannot be allowed to depart before f' has arrived (possibly because passengers in f' connect to f). In this interpretation, a flight f' may have more than one "next" flights f. This interpretation will *not* be pursued in the sequel.

Note that nonnegativity of g_f is guaranteed by (3), whereas nonnegativity of a_f is not guaranteed. This is why constraints (10) are needed.

For simplicity of exposition, variables g_f and a_f were kept in formulation $\mathbf{P}_1$, but it should be clear that they can be eliminated by mere substitution through (3) and (4), so that u_{ft} and v_{ft} are the only decision variables. The result of this substitution is given in Appendix A as $\mathbf{P}_1'$, where only u_{ft} and v_{ft} appear.

1.4. A Simpler Case: Infinite Departure Capacities and Zero Airborne Delays

Formulation $\mathbf{P}_1$ is sufficiently general for the static deterministic case, but it can be simplified considerably without significant loss of applicability. First, it is usually undesirable to delay aircraft in the air. In fact, the fundamental goal of ground holding policies is to avoid this kind of delay. Therefore, we may eliminate airborne delays *as decision variables*. We will be left with airborne delays resulting only from arrival queueing (denoted earlier by q_f^a), and our only decision variables will be g_f. (Note that because the problem is deterministic, q_f^a are determined if g_f and service disciplines for the arrival queues are given.)

Departure capacities are typically higher than landing capacities. This is due to the fact that the minimum separation between landings is greater than the minimum separation between take-offs. Motivated by this fact, we examined what happens if departure capacities are very large and theoretically infinite.

We will show that if departure capacities are infinite, ground and airborne delay cost functions are linear, and $c_f^a > c_f^g$, then, if $\mathbf{P}_1$ without airborne delays as decision variables has an optimal solution, then it also has an optimal solution in which no flight incurs an airborne delay.

Consider a feasible solution $\{g_f, f \in \mathcal{F}\}$ and the associated arrival queueing delays $\{q_f^a, f \in \mathcal{F}\}$, and compare its cost with the cost of the new solution $\{g_f + q_f^a, f \in \mathcal{F}\}$, in which all airborne delays are incorporated into ground holds. Given that the cost functions are linear, and that airborne delays are costlier than ground delays (i.e., for any positive t, and for all f, $c_f^a(t) > c_f^g(t)$), it is easy to show that the new solution will have a lower cost than the previous solution. In fact, $c_f^g(g_f + q_f^a) = c_f^g g_f + c_f^g q_f^a < c_f^g g_f + c_f^a q_f^a$. Moreover, the new solution $\{g_f + q_f^a, f \in \mathcal{F}\}$ is feasible (assuming sufficiently large G_f), because there are no departure capacity constraints.

Are we entitled to assume that departure capacities are infinite? For practical purposes, this assumption may often be a good approximation, because congestion problems are mostly due to limited landing rather than departure capacities. Moreover, computational experience reported in Section 3 shows that the impact of finite departure capacities is negligible (when departure capacities are higher than arrival capacities by realistic amounts). This *a posteriori* argument justifies the assumption of infinite departure capacities. Note that in the single-airport case, which is the only case considered so far in the literature, no departure capacities are involved, so that one is rigorously justified to consider only feasible solutions with zero airborne delays (provided that the problem is deterministic, the cost functions are linear, and airborne delays are costlier than ground delays).

Assuming infinite departure capacities eliminate airborne delays we give a second pure 0–1 integer programming formulation of the static deterministic multi-airport **GHP**. The second formulation is, in some sense, a special case of $\mathbf{P_1}$ but requires some manipulations to be derived from $\mathbf{P_1}$. Given (4), by setting $a_f = 0$, one gets for g_f:

$$g_f = \sum_{t \in \mathcal{T}_f^a} t v_{ft} - r_f, \quad f \in \mathcal{F}. \tag{11}$$

By comparing (1) and (2), one can see that A_f must be equal to 0 in the case of infinite departure capacities without airborne delays as decision variables: If flight f takes off at $d_f + t$, it will land at $r_f + t$. By comparing (11) with (3), we see that:

$$\sum_{t \in \mathcal{T}_f^a} t v_{ft} - \sum_{t \in \mathcal{T}_f^d} t u_{ft} = r_f - d_f, \quad f \in \mathcal{F}, \tag{12}$$

so (given (7) and (8)) one of the two sets of variables is now redundant. We choose to discard u_{ft} and to keep v_{ft}, because v_{ft} appears in the arrival capacity constraints (6), which must be kept. The departure capacity constraints (5) are discarded, as are the assignment constraints (7). We are left with the following formulation.

Problem $\mathbf{P_2}$

$$\text{Minimize} \sum_{f=1}^{F} c_f^g g_f$$

subject to

$$\begin{aligned}
&\sum_{f: k_f^a = k} v_{ft} \le R_k(t), \quad (k, t) \in \mathcal{K} \times \mathcal{T};\\
&\sum_{t \in \mathcal{T}_f^a} v_{ft} = 1, \qquad f \in \mathcal{F};\\
&g_{f'} - s_{f'} \le g_f, \qquad f' \in \mathcal{F}';\\
&\qquad v_{ft} \in \{0, 1\}, \quad f \in \mathcal{F}, t \in \mathcal{T}_f^a.
\end{aligned} \tag{13}$$

The result of substituting (11) into $\mathbf{P_2}$ is given in Appendix A as $\mathbf{P_2'}$, in which only the decision variables v_{ft} appear.

Note the simplicity of $\mathbf{P_2}$. The number of constraints is $F + F' + KT$, and the number of variables is at most $\sum_{f \in \mathcal{F}} (G_f + 1)$ which, if all G_f are equal to 4 (corresponding to a maximum ground hold of one hour), becomes $5F$. Therefore, the total number of flights F is the major determinant of the size of the problem. The number of time periods T has almost no influence on the size of the problem, and the same holds for the number of airports K. Of course, the number of airports has an indirect influence on the size of the problem, because it influences the number of flights to be considered. Typically, a major U.S. airport has 600–2,000 operations (landings plus take-offs) each day, corresponding to 300–1,000 flights per day. But still, the fact that the problem is insensitive as to how the total number of flights is distributed among airports and time periods is welcome. This becomes clear in dynamic versions of the ground-holding problem (not treated in this paper), where the time horizon is limited to a portion of a day, so that fewer flights per airport have to be considered, and it becomes possible to solve the problem for a large number of airports.

Note, finally, that if the coupling constraints (13) are omitted from the formulation, what is left is essentially one of the single-airport formulations given in Terrab for the static deterministic case. The coupling constraints are the gist of the model. It is indeed surprising that the network effects can be taken into account in such a simple way without loss of generality.

1.5. How to Handle Infeasibility: Cancelling Flights

In situations where delays become excessive, it is common airline practice to cancel some flights, especially at hub airports. Motivated by this fact, we developed formulations which take into account the possibility of cancelling flights. These formulations have the additional advantage that they escape infeasibility problems which might arise with $\mathbf{P}_1$ and $\mathbf{P}_2$. Infeasibility occurs when airport capacities are low: Even though the total daily capacity of an airport may be sufficient to accommodate the total number of flights scheduled to depart from or arrive at that airport, the problem may still be infeasible if excessive congestion appears during some portion of the day. This is mainly due to the requirement that there be upper bounds, G_f and A_f, to the delays of flight f. To grasp this point with respect, e.g., to $\mathbf{P}_2$, take the extreme case where the landing capacity of an airport is reduced to zero for $G_f + 1$ successive time periods. Then, if a flight was scheduled to arrive exactly before the zero capacity interval, it will be impossible to reassign this flight and the problem will become infeasible. Similar remarks hold for $\mathbf{P}_1$.

We will give a new formulation, $\mathbf{P}_3$, that generalizes $\mathbf{P}_2$. Another formulation, generalizing $\mathbf{P}_1$, can be derived similarly and is given in Vranas (1992a).

Keep the old decision variables v_{ft} and define the decision variables z_f, $f \in \mathcal{F}$, to be 1 if flight f is cancelled and 0 otherwise. Denote by M_f the *cancellation cost* of flight f. When a flight in $\mathcal{F}'$ (i.e., a flight that is "continued") is cancelled, there are two possibilities concerning the next flight initially scheduled to be performed by the same aircraft: Either it is performed by a replacement (or a "spare") aircraft, or it is also cancelled. The first case is more common in practice, especially in hub airports where most cancellations take place, but the formulation is general enough to incorporate a combination of both cases. Partition $\mathcal{F}'$ into $\mathcal{F}'_1$, the set of those flights in $\mathcal{F}'$ whose cancellation will not affect their next flight, and $\mathcal{F}'_2$, the set of those flights in $\mathcal{F}'$ whose cancellation will entail the cancellation of their next flight. We will now give the new formulation and then comment on it.

Problem $\mathbf{P}_3$

$$\text{Minimize} \sum_{f=1}^{F} (c_f^g g_f + (M_f + c_f^g r_f) z_f) \tag{14}$$

subject to

$$\sum_{f: k_f^a = k} v_{ft} \leq R_k(t), \quad (k, t) \in \mathcal{K} \times \mathcal{T}; \tag{15}$$

$$z_f + \sum_{t \in \mathcal{T}_f^a} v_{ft} = 1, \quad f \in \mathcal{F}; \tag{16}$$

$$g_{f'} - s_{f'} + (s_{f'} + r_{f'} - r_f) z_{f'} \leq g_f, \quad f' \in \mathcal{F}'_1; \tag{17}$$

$$g_{f'} - s_{f'} + (s_{f'} + r_{f'} + G_f + 1) z_{f'} \leq g_f + (r_f + G_f + 1) z_f, \quad f' \in \mathcal{F}'_2; \tag{18}$$

$$v_{ft}, z_f \in \{0, 1\}. \tag{19}$$

The above formulation incorporates some technical tricks which are necessitated by the fact that, when a flight f is cancelled (i.e., $z_f = 1$), then all v_{ft} corresponding to f are 0 (by 16), so that (11) gives $g_f = -r_f$. Keeping this fact in mind, we can see immediately that, when $z_f = 1$, the objective function term corresponding to f is M_f. It is also clear that, when $z_{f'} = 1$, (17) becomes $-r_f \leq g_f$, which holds even if flight f is cancelled (so that cancellation of f' leaves f unaffected). Finally, if $z_{f'} = 1$, (18) becomes $G_f + 1 \leq g_f + (r_f + G_f + 1) z_f$, entailing $z_f = 1$ (because $g_f \leq G_f$ always), which is precisely what we wanted: If f' is cancelled, then f is also cancelled.

The variables g_f were again left in the formulation, but it should be clear that they can be eliminated by mere substitution through (11). It is important to notice that the variables z_f can also be eliminated through (16), provided that (16) is replaced by $\sum_{t \in \mathcal{T}_f^a} v_{ft} \leq 1$. The outcome of effecting all these substitutions is $\mathbf{P}'_3$, given in Appendix A.

The fact that the new formulation $\mathbf{P}_3$ has exactly the same number of variables and constraints as the previous corresponding formulation $\mathbf{P}_2$ is particularly interesting, because $\mathbf{P}_3$ enjoys considerable advantages both in terms of generality (the real-world problem is better approximated) and flexibility (infeasibility problems are eliminated).

2. A HEURISTIC

This section presents a heuristic which finds a feasible solution of the integer program $\mathbf{P}_3$ starting from a feasible solution of the linear programming (LP) relaxation of $\mathbf{P}_3$. The next section will show, on the basis of computational experience, that it is easy to optimally solve the LP relaxation of $\mathbf{P}_3$, and when one applies the heuristic to this optimal solution, one gets a "good" feasible solution of the integer program $\mathbf{P}_3$.

The heuristic will be presented in rough outline here. An algorithmic presentation is given in Appendix B.

Consider a feasible solution

$$\{v_{ft}: f \in \mathcal{F}, t \in \mathcal{T}_f^a\} \cup \{z_f: f \in \mathcal{F}\}$$

of the LP relaxation of $\mathbf{P}_3$ and denote by Φ the set of "problematic" flights $f \in \mathcal{F}$, i.e., the set of flights for which some integrality constraint is violated. The heuristic gives a "rounding" scheme for flights in Φ which leaves undisturbed, as far as possible, the remaining flights (which already satisfy integrality). The basic idea of the heuristic is to treat each flight in Φ once.

The heuristic starts by partitioning Φ into classes, each class corresponding to an aircraft and containing all and only the flights of Φ scheduled to be performed by that aircraft. The heuristic treats each class separately; the order in which the classes are treated is arbitrary.

Each class is treated in the following way. The flights in the class are examined one at a time, in the order in which they are scheduled to be performed by the aircraft defining the class. For each specific flight ϕ, the heuristic takes the following actions. (It will help the reader at this point to refer to $\mathbf{P}_3$.) For each time period t at which ϕ can be allowed to land, it computes the available "capacity slacks" $R_{k_\phi^a}(t) - \sum_{f:k_f^a=k_\phi^a} v_{ft}$ (i.e., the slacks of 15), which will be denoted by $S_\phi(t)$. (If some v_{ft} have already been updated by new values, then the new values are used in the computation of the capacity slacks.) It can be seen that if $S_\phi(t) \geq 1 - v_{\phi t}$, then it is possible to assign flight ϕ to period t without violating the corresponding capacity constraint. If this is possible for no t, then flight ϕ is cancelled and we are done with it. Otherwise, when there are time periods to which it is possible to assign flight ϕ without violating the corresponding capacity constraint, flight ϕ is assigned to the earliest such period τ. (Recall that this assignment is made once.) After this assignment, all constraints involving flight ϕ are satisfied, with the possible exception of the coupling constraints.

To deal with the coupling constraint linking flight ϕ with its next flight $\hat{\phi}$ (if such a next flight exists), the heuristic removes certain time periods from the set of time periods at which $\hat{\phi}$ can be allowed to land, and proceeds to examine $\hat{\phi}$. The removed time periods are those that would violate the coupling constraint in question if $\hat{\phi}$ were assigned to them (given that ϕ has already been assigned to τ). We can see that if flight ϕ has a *previous* flight ϕ', the coupling constraint linking ϕ' and ϕ need not be dealt with while examining flight ϕ, because it has been dealt with when examining flight ϕ' (because ϕ is the *next* flight to ϕ').

As pointed out, this is only a rough outline; a more rigorous and detailed description is given in Appendix B.

3. STRUCTURAL INSIGHTS

This section investigates the behavior of the **GHP** on the basis of extensive computational experience. The investigation is conducted in three parts; each part deals with one of the formulations, $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$. For each formulation, we examine the variation, as a function of the input parameters, of the optimal objective function values of the following three mathematical optimization problems: the integer program (denoted by I), the corresponding linear programming relaxation (denoted by L), and the "decomposed" program (denoted by D), defined as the integer program without the coupling constraints.

It is important to understand the role of D in the comparison. The *decomposed* **GHP** corresponding to $\mathbf{P}_2$ is simply $\mathbf{P}_2$ without the coupling constraints (13). Solving the decomposed **GHP** is equivalent to solving the **GHP** for each airport separately, and then adding the optimal objective function values corresponding to the various airports. Note that the optimal objective function value of the decomposed **GHP** is equal to the optimal objective function value of the LP relaxation of the decomposed **GHP**, because the constraint matrix of any single-airport **GHP** is totally unimodular (Terrab). Therefore, D can be defined as a linear rather than an integer program.

Denote the optimal values of I, L, and D by v_I, v_L, and v_D, respectively. Now the greater the gap between v_D and v_I (and, *a fortiori*, the greater the gap between v_D and v_L), the greater the impact of the network effects. A large gap between v_D and v_I presumably justifies pursuing the application of algorithms pertaining to the multi-airport (coupled) **GHP**, rather than solving for each airport separately by means of the existing methods for the single-airport **GHP**. This much is clear. What is less clear is how a *small* gap between v_D and v_I should be interpreted. A small gap would not necessarily mean that the multi-airport **GHP** is valueless. Consider the extreme case where $v_D = v_I$. The zero gap means that we could ignore the coupling constraints without any change in the optimal value of I. But if D has multiple optimal solutions, then solving it will not necessarily give a solution satisfying the coupling constraints, i.e., a solution *feasible* for I.

Note that the objective of this section is to investigate the behavior of the problem under various combinations of the input parameters, *not* to demonstrate the efficiency of any particular algorithm. We solved the various instances of the problem by using the well-known commercial package MPSX, rather than any custom-tailored algorithm. We give CPU times simply to indicate whether the problem can be solved in reasonable time, rather than to provide any "good" bounds on computation times.

This section is divided into three subsections. The basic conclusions are reached in the first subsection, which deals with $\mathbf{P}_2$. The second subsection, which deals with $\mathbf{P}_1$, verifies that the impact of finite departure capacities would be negligible in many practical cases. Finally, the third subsection deals with $\mathbf{P}_3$

(with flight cancellations) and the performance of the heuristic.

3.1. The Model Without Flight Cancellations

This subsection deals with $\mathbf{P}_2$ and shows that network effects, defined as the difference between v_I and v_D, are small when all flights have the same cost function but can be large otherwise. The case of identical cost functions is of practical interest, because it reflects the current FAA practice of avoiding any kind of discrimination among classes of users. We also show, however, that even when all cost functions are identical, network formulations are needed, because the optimal solution of the decomposed problem is, typically, infeasible for the coupled problem.

3.1.1. Network Effects Are Insignificant When Cost Functions Are Identical

We consider first a test case with $K = 3$ airports, $T = 100$ time periods, $F = 1{,}800$ flights (600 flights per airport), and $F' = 600$ flights. With the exception of capacities, all parameters are kept fixed in this test case: The cost function slopes are 50, the slacks are 0, and the upper bounds on the delays are 4 time periods. The scheduled arrival times were arbitrarily chosen.

As mentioned in Section 1, if arrival capacities are very low, the problem becomes infeasible. Let us consider only cases in which the arrival capacity of any given airport is constant over the whole time horizon: $R_k(t) = R_k$. Then we find that, for the particular test case under consideration, for $(R_1, R_2, R_3) = (10, 10, 10)$ the problem is feasible, while for (9, 9, 9) the problem is infeasible. Furthermore, for (9, 10, 10), (10, 10, 9), (9, 10, 9), and (10, 10, 8) the problem is feasible, while for (10, 9, 10), (8, 10, 10), and (10, 10, 7) the problem is infeasible. These results give us a fairly good picture of the border between capacity regions that correspond to feasibility and to infeasibility for the test case under consideration. Delimitation of this border is important because it is there that the greatest delays are expected to occur: If capacities are very high, then there is little need to delay aircraft.

Table I gives the optimal objective function values of L, D, and I for the various capacity cases; these values always turn out to be very close. An examination of the optimal solution of D, however, reveals that usually about 180–200 of the 600 coupling constraints are violated. It follows that solving the decomposed problem is probably of little use as far as getting a feasible solution to the coupled problem is concerned. Nevertheless, solving the decomposed problem provides a good indicator of what the optimal value of the coupled problem will be.

The proximity of v_D and v_I needs an explanation, but we must first ascertain that it is a common phenomenon rather than a peculiar feature of the particular test case under consideration. To this end, we examined a systematic series of test cases. In all these cases, T is kept fixed and equal to 64 (corresponding to a 16-hour time horizon with 15-minute periods), and K is determined by F via the assumption that 500 flights are scheduled to land at each airport during the time horizon. Three cases for F are examined: 1,000, 2,000, and 3,000 flights (corresponding, respectively, to 2, 4, and 6 airports). For each particular F, four values of F' are examined, corresponding to a ratio F'/F equal to 0.20, 0.40, 0.60, and 0.80. The results are summarized in Table II. The capacities appearing in the table for any particular case are at the infeasibility borders (and were found by trial and error). The cost function slopes are always 50, all slacks are 1, and all upper bounds on delays are 4.

These results lead to the following conclusions. First, the gap between v_D and v_I is always small. Second, the computation times (given in CPU seconds) t_D and t_L are quite reasonable, but t_I can become excessive. Third, as one would expect, the computation times increase as F increases, because the number of constraints and variables increases. Fourth, for any given F, t_D does not vary significantly with F', while t_L and t_I increase as F' increases. This is due to the fact that an increase in F' increases the number of constraints of L and I (which have $KT + F + F'$ constraints), while it leaves unaffected the number

Table I
Behavior of the Test Case Around the Capacity Border Between Feasibility and Infeasibility

Capacities	v_D	v_L	v_I	No. of Coupling Constraints That D Violates	Percent of $f \in \mathcal{F}'$ Delayed in I	Percent of $f \in \mathcal{F}/\mathcal{F}'$ Delayed in I
(10, 10, 10)	43,550	43,550	43,550	179	12	30
(9, 10, 10)	51,900	52,800	52,900	204	18	36
(10, 10, 9)	48,500	49,000	50,600	183	17	34
(9, 10, 9)	56,850	57,450	57,950	238	20	40
(10, 10, 8)	55,650	56,700	58,000	235	19	37

Table II
Results for Various Cases at the Infeasibility Border

$\mathcal{F}$	$\mathcal{F}'/\mathcal{F}$	Capacity	v_D	t_D	v_L	t_L	v_I	t_I	
1,000	0.20	(12, 14)	71,000	218	71,000	258	71,000	371	63
1,000	0.40	(10, 10)	56,000	235	56,000	327	56,000	894	84
1,000	0.60	(11, 11)	84,200	242	84,300	377	84,700	6,958	128
1,000	0.80	(10, 10)	65,000	235	65,000	453	65,500	9,512	168
2,000	0.20	all 14	96,300	664	96,300	731	99,000	5,126	117
2,000	0.40	all 14	88,400	652	89,933	973	93,200	9,522	195
2,000	0.60	all 12	71,600	644	71,600	1,148	71,800	13,607	252
2,000	0.80	all 17	53,250	617	57,387	1,603	65,500	18,093	355
3,000	0.20	all 12	128,000	1,188	129,200	1,453	129,400	11,360	110
3,000	0.40	all 18	55,800	1208	55,800	1,808	57,300	13,291	119
3,000	0.60	all 17	90,200	1,166	96,550	2,547	99,687	17,980	414
3,000	0.80	all 18	80,500	1,180	84,250	3,072	87,012	25,021	232

of constraints of D $(KT + F)$. Finally, the last column in Table II gives the number of flights for which the optimal solution of L had noninteger values. It can be seen that this number is usually small, around 10% of F. This observation provided the motivation for the development of the heuristic given in Section 2.

3.1.2. Network Effects Significant When Cost Functions Differ

Now we must explain the fact that v_D and v_I are typically very close. Our conclusion will be that this is because all cost functions were identical. Before we argue for this conclusion, let us examine two other possible explanations that might be adduced. A first explanation might be that the capacities at the border between feasibility and infeasibility, although they cannot be lowered in the context of the present model, are still too high for network effects to have a severe impact. This explanation, if true, would undermine the utility of $\mathbf{P}_2$ (though not of $\mathbf{P}_1$) as a representation of the real-world situation. This explanation, however, is not true. First, v_D and v_I are very close even with low capacities (see the second and the fourth rows of Table II). Second, in subsection 3.3, where $\mathbf{P}_3$, which is immune to infeasibility, is examined, it will be seen (cf. fifth row of Table IV) that v_L and v_D are very close even with capacities as low as 256 aircraft per airport per day (4 per period) (with 500 aircraft scheduled to land, so that the remaining flights are cancelled).

A second possible explanation is that arrival capacities were taken to be uniform (i.e., constant over the whole time horizon). Ground-holding policies make sense when one delays aircraft on the ground because one expects less congestion later on at the destination airports of the delayed aircraft. But when airport capacities are uniform throughout the day, how can one expect less congestion later on? The answer is that less congestion can be expected when fewer aircraft are scheduled to arrive later on, even if arrival capacities are uniform. Nevertheless, this second possible explanation has some validity, as shown by the computational results reported in subsection 3.3 (cf. Table V), where nonuniform capacities can give somewhat significant network effects.

The main explanation, however, is the identity of cost functions. If there is a choice (in I) between delaying a continued flight and a noncontinued flight, it will usually be preferable to delay the latter, because delaying the former would probably result in a greater total cost (because the next flight might also have to be delayed). If this is the case, then, in the optimal solution of I, few flights in $\mathcal{F}'$ will be delayed. This effect would be particularly noticeable for small slacks. A look at the last two columns of Table I corroborates this hypothesis. A second way to confirm this hypothesis is by varying the cost function slopes to disadvantage continued flights. If continued flights have much lower marginal costs than noncontinued flights, then it may often be preferable to delay a continued rather than a noncontinued flight when a choice is available, with the consequence that network effects may be significant. The test case with 1,800 flights was run with capacities equal to 10 and with cost function slopes equal to 10 for the continued flights and equal to 100 for the noncontinued flights; the results were $v_D = 13{,}950$ and $v_L = 22{,}811$, a significant gap. Other results with different cost functions, reported in subsection 3.3 (Table VI), also show significant network effects.

3.2. The Negligible Impact of Finite Departure Capacities

To check the impact of finite departure capacities and to demonstrate that $\mathbf{P}_1$, which has more than twice as many variables and constraints as $\mathbf{P}_2$, can be also solved in reasonable computation times, we examined the problems of the first two rows of Table II with various departure capacities. To make meaningful comparisons, the scheduled arrival times were kept unchanged. The new data, besides the departure capacities, were the scheduled departure times or, equivalently, the flight times. Table III gives results for various combinations of departure capacities and flight times. Airborne marginal delay costs were taken to be 75 versus ground marginal delay costs of 50.

Table III shows that when flight times are uniform (e.g., equal to 2 time periods) or slightly nonuniform, the differences between finite and infinite departure capacities are negligible. It is only with strongly nonuniform flight times that some minor differences appear. (The nonuniform flight times of Table III were 1 or 2 time periods for $F'/F = 0.20$ and varied from 1 to 30 time periods for $F'/F = 0.40$.) These results justify pursuing the investigation with the more manageable formulation $\mathbf{P}_2$. In any event, however, $\mathbf{P}_1$ is also manageable (running times for the cases of Table III were about 2,000 CPU seconds).

It is important to note that departure capacities were implicitly assumed to be independent from arrival capacities. Often the departure and arrival capacities of a given airport are interdependent, because they are determined by the way in which runway use is assigned to departing or arriving aircraft. Our formulations can easily be modified to take this interdependence into account (Vranas 1992, 1994). Computational results reported in Vranas (1992) show that, by optimally varying the mix between departure and arrival capacities as time progresses, one can achieve significant cost savings (35–40%) with respect to $\mathbf{P}_2$.

3.3. The Model With Flight Cancellations

Table IV gives results for selected cases from Table II, but for $\mathbf{P}_3$ and for various capacities and cancellation costs M. The rows with "infinite" cancellation costs correspond to $\mathbf{P}_2$ and are taken from Table II. All marginal delay costs are equal to 50.

These results strongly support the conclusion that, for cancellation costs greater than 100 times the marginal delay cost (i.e., here, $M > 5{,}000$), no flight is ever cancelled, so that models $\mathbf{P}_2$ and $\mathbf{P}_3$ give the same results. For cancellation costs greater than 20 times the marginal delay costs ($M > 1{,}000$), few flights are cancelled, so that the optimal values of $\mathbf{P}_2$ and $\mathbf{P}_3$ are very close. Finally, for cancellation costs less than 10 times the marginal delay cost ($M < 500$), more flights are cancelled and significant differences between $\mathbf{P}_2$ and $\mathbf{P}_3$ emerge. Note also that, in that last region of cancellation costs, the slope of the optimal value as a function of the cancellation cost becomes quite abrupt.

The last column of Table IV shows the value v_H of the objective function corresponding to the feasible solution found by the heuristic. It can be seen that v_H is quite close to v_L (hence, to v_I) for small cancellation costs. For large cancellation costs, however, the heuristic performs poorly. This was to be expected, because the heuristic will inevitably cancel some flights, and these will inflate the objective function value if the cancellation cost is excessive. This is not worrisome, however, because, as pointed out before, for cancellation costs above 1,000 few flights are cancelled, so that for such high cancellation costs the heuristic has little practical use, because one should solve $\mathbf{P}_2$ rather than $\mathbf{P}_3$.

Table V gives results concerning cases with nonuniform arrival capacities. It can be seen that gaps between v_D and v_I are somewhat significant.

As explained in subsection 3.1, the main reason why network effects were found to be insignificant was the assumption that all cost functions are identical. To check this, we ran some cases with three classes of costs: 40% of all flights had cost 100, 40% had cost 50, and 20% had cost 20, corresponding to the relative direct operating costs of large, medium-sized, and small aircraft, respectively. Aircraft performing continued flights were generally assigned to the large- or medium-cost category. The results are shown in Table VI; the differences are quite significant (22–27%).

Table III
Results for Various Cases With Finite Departure Capacities

$\mathcal{F}$	$\mathcal{F}'/\mathcal{F}$	Arrival Capacities	Departure Capacities	Flight Times	v_L
1,000	0.20	(12, 14)	∞	—	71,000
1,000	0.20	(12, 14)	(12, 14)	Uniform: 2	71,000
1,000	0.20	(12, 14)	(15, 17)	Nonuniform: 1 or 2	71,500
1,000	0.40	(10, 10)	∞	—	56,000
1,000	0.40	(10, 10)	(10, 10)	Uniform: 2	56,000
1,000	0.40	(10, 10)	(15, 15)	Nonuniform: 1 to 30	62,083
1,000	0.40	(10, 10)	(16, 16)	Nonuniform: 1 to 30	57,250

Table IV
Results for Various Cases With Flight Cancellations

$\mathcal{F}$	$\mathcal{F}'/\mathcal{F}$	Capacities	M	v_D	t_D	v_L	t_L	v_H
1,000	0.60	11	1,000	70,300	297	70,300	479	78,500
1,000	0.60	10	1,000	117,000	286	117,000	475	125,450
1,000	0.60	08	1,000	240,700	280	241,805	524	253,750
1,000	0.60	06	1,000	402,600	274	403,476	513	411,500
1,000	0.60	04	1,000	582,300	272	583,417	484	586,700
1,000	0.60	11	100	28,700	283	28,700	498	30,250
1,000	0.60	11	1,000	70,300	297	70,300	473	78,500
1,000	0.60	11	10,000	84,200	276	84,300	444	240,700
1,000	0.60	11	∞	84,200	242	84,300	377	—
2,000	0.20	14	500	77,500	652	77,500	803	82,000
2,000	0.20	14	1,000	94,000	691	94,000	922	103,300
2,000	0.20	14	5,000	96,300	717	96,300	931	165,200
2,000	0.20	14	∞	96,300	664	96,300	731	—
2,000	0.40	14	500	73,100	815	74,983	1,020	75,800
2,000	0.40	14	1,000	86,100	690	86,372	1,102	93,650
2,000	0.40	14	5,000	88,400	675	89,933	1,176	168,900
2,000	0.40	14	∞	88,400	652	89,933	973	—
3,000	0.60	17	100	38,250	1,119	38,693	1,911	42,350
3,000	0.60	17	500	71,800	1,128	72,240	1,708	84,600
3,000	0.60	17	750	81,000	1,148	81,338	1,931	95,000
3,000	0.60	17	1,000	87,000	1,187	87,156	2,114	130,300
3,000	0.60	17	10,000	90,200	1,248	96,550	3,767	667,750
3,000	0.60	17	∞	90,200	1,166	96,550	2,547	—
3,000	0.80	18	100	36,600	1,114	38,042	1,846	58,900
3,000	0.80	18	500	71,500	1,140	71,559	2,320	83,350
3,000	0.80	18	750	78,700	1,128	78,707	2,693	106,800
3,000	0.80	18	1,000	80,500	1,235	82,214	2,900	111,350
3,000	0.80	18	10,000	80,500	1,230	84,250	3,227	509,900
3,000	0.80	18	∞	80,500	1,180	84,250	3,072	—

4. CONCLUSIONS

The multi-airport **GHP** was shown to be tractable. Our formulations capture the essential aspects of the problem, for the static deterministic case at least, and do so in a very simple way. It is this simplicity, reflected in the small numbers of constraints and variables, that is responsible for the tractability of large-scale **GHPs**.

The main insights derived from the investigation of Section 3 were:

1. In the general case (when cost functions differ), network effects, defined as the difference between the optimal objective function values of the integer and the decomposed problems, can be large. Network effects can also be large when airport capacities are not uniform.

Table V
Results for Various Cases With Flight Cancellations and Nonuniform Capacities

$\mathcal{F}$	$\mathcal{F}'/\mathcal{F}$	Capacities	M	v_D	t_D	v_L	t_L
3,000	0.80	Nonun.	500	232,800	1,142	252,045	1,973
3,000	0.80	Nonun.	750	302,700	1,200	330,040	2,217
3,000	0.80	Nonun.	1,000	366,200	1,215	403,127	2,228

Table VI
Results for Various Cases With Flight Cancellations and Three Cost Classes

$\mathcal{F}$	$\mathcal{F}'/\mathcal{F}$	Capacities	M	v_D	t_D	v_L	t_L
3,000	0.60	Nonun.	500	305,690	1,236	373,271	2,099
3,000	0.60	Nonun.	750	385,830	1,253	491,791	2,219
3,000	0.60	Nonun.	1,000	460,230	1,305	601,331	2,332

2. In the special case where all cost functions are identical, network effects are of small magnitude. The assumption of identical cost functions is incorrect, because the delay of large aircraft is more costly than the delay of small aircraft. However, the practice of implicitly considering all cost functions identical seems to be a well-entrenched practice of the FAA, which avoids "discriminating" in any way among users.
3. Even when all cost functions are identical, the optimal solution of the decomposed problem typically violates a large number of coupling constraints and is thus useless for practical purposes. This means that network formulations are needed to assign feasible ground holds to a network of airports.
4. Finite departure capacities have negligible impact if they are assumed not to influence arrival capacities. On the other hand, the possibility of having interdependent departure and arrival capacities offers the potential for significant cost savings.
5. As far as the model with flight cancellations is concerned, high cancellation costs are impractical because they result in no flights ever being cancelled.
6. The heuristic which finds a feasible solution of the IP with cancellations on the basis of the optimal solution of the LP relaxation performs well for low cancellation costs.

It is not yet clear how large a network one can deal with by means of our formulations. We went up to 6 airports and 3,000 flights, but one could probably go far beyond this if one were willing to use supercomputers. This would not be unrealistic, given the importance of the practical problem. One could also look for special purpose algorithms or for heuristics providing good feasible solutions.

A direction for future research is to extend our formulations to the dynamic deterministic case. We have already performed this extension (Vranas 1992a, b), which is relatively straightforward, although it needs to incorporate some subtleties due to the fact that airborne delays cannot be totally avoided in the dynamic case.

Another interesting direction of research for the dynamic case would be to run our formulations for limited time horizons of, say, two hours. This would dramatically decrease the size of the problem for a given number of airports, and would enable one to tackle much larger networks of airports.

The most challenging direction for future research is the case of probabilistic airport capacities (see Richetta for the single-airport problem). This case may require a totally new approach.

APPENDIX A

Final Forms of the Formulations

Formulation $\mathbf{P}_1'$ is derived from $\mathbf{P}_1$ by eliminating g_f and a_f in terms of u_{ft} and v_{ft} through (3) and (4):

Minimize

$$\sum_{f=1}^{F} c_f^a\Bigl(\sum_{t\in\mathcal{T}_f^a} tv_{ft} - r_f\Bigr) - (c_f^a - c_f^g)\Bigl(\sum_{t\in\mathcal{T}_f^d} tu_{ft} - d_f\Bigr)$$

subject to

$$\sum_{f:k_f^d=k} u_{ft} \leqslant D_k(t), \qquad (k,t)\in\mathcal{K}\times\mathcal{T};$$

$$\sum_{f:k_f^a=k} v_{ft} \leqslant R_k(t), \qquad (k,t)\in\mathcal{K}\times\mathcal{T};$$

$$\sum_{t\in\mathcal{T}_f^a} u_{ft} = 1, \qquad f\in\mathcal{F};$$

$$\sum_{t\in\mathcal{T}_f^a} v_{ft} = 1, \qquad f\in\mathcal{F};$$

$$\sum_{t\in\mathcal{T}_{f'}^a} tv_{f't} - r_{f'} - s_{f'} \leqslant \sum_{t\in\mathcal{T}_f^d} tu_{ft} - d_f, \quad f'\in\mathcal{F}';$$

$$\sum_{t\in\mathcal{T}_f^a} tv_{ft} - \sum_{t\in\mathcal{T}_f^d} tu_{ft} \geqslant r_f - d_f, \qquad f\in\mathcal{F};$$

$$u_{ft}, v_{ft} \in \{0,1\}.$$

Formulation $\mathbf{P}_2'$ is derived from $\mathbf{P}_2$ by eliminating g_f in terms of v_{ft} through (11):

$$\text{Minimize} \sum_{f=1}^{F} c_f^g\Bigl(\sum_{t\in\mathcal{T}_f^a} tv_{ft} - r_f\Bigr)$$

subject to

$$\sum_{f:k_f^a=k} v_{ft} \leqslant R_k(t), \qquad (k,t)\in\mathcal{K}\times\mathcal{T};$$

$$\sum_{t\in\mathcal{T}_f^a} v_{ft} = 1, \qquad f\in\mathcal{F};$$

$$\sum_{t\in\mathcal{T}_{f'}^a} tv_{f't} - r_{f'} - s_{f'} \leqslant \sum_{t\in\mathcal{T}_f^a} tv_{ft} - r_f, \quad f'\in\mathcal{F}';$$

$$v_{ft}\in\{0,1\}, \qquad f\in\mathcal{F}, t\in\mathcal{T}_f^a.$$

Formulation $\mathbf{P}_3'$ is derived from $\mathbf{P}_3$ by eliminating g_f in terms of v_{ft} through (11) and by eliminating z_f:

$$\text{Minimize} \sum_{f=1}^{F} \Bigl[M_f + \sum_{t\in\mathcal{T}_f^a} v_{ft}(c_f^g(t - r_f) - M_f)\Bigr]$$

subject to

$$\sum_{f:k_f^a=k} v_{ft} \leqslant R_k(t), \quad (k,t)\in\mathcal{K}\times\mathcal{T};$$

$$\sum_{t \in \mathcal{T}_f^a} v_{ft} \leq 1 \quad f \in \mathcal{F};$$

$$\sum_{t \in \mathcal{T}_{f'}^a} v_{f't}(t - s_{f'} - r_{f'} + r_f) \leq \sum_{t \in \mathcal{T}_f^a} t v_{ft}, \quad f' \in \mathcal{F}_1';$$

$$\sum_{t \in \mathcal{T}_{f'}^a} v_{f't}(t - s_{f'} - r_{f'} - G_f - 1)$$

$$\leq \sum_{t \in \mathcal{T}_f^a} v_{ft}(t - r_f - G_f - 1), \quad f' \in \mathcal{F}_2';$$

$$v_{ft}, z_f \in \{0, 1\}.$$

APPENDIX B

Algorithmic Description of the Heuristic

The heuristic takes an input a solution $\{v_{ft}: f \in \mathcal{F}, t \in \mathcal{T}_f^a\} \cup \{z_f: f \in \mathcal{F}\}$ which is feasible for the LP relaxation of $\mathbf{P}_3$, and gives as output a solution which is feasible for $\mathbf{P}_3$. The heuristic is presented here for the case in which the next flight scheduled to be performed by the same aircraft is not affected when a flight is cancelled. The other case, in which the next flight is also cancelled, can be treated *mutatis mutandis*.

BEGIN

Define $\Phi := \{\phi \in \mathcal{F}: (z_\phi \notin \{0, 1\}) \vee (\exists t)(v_{\phi t} \notin \{0, 1\})\}$.

Partition Φ into its equivalence classes corresponding to the equivalence relation "is performed by the same aircraft as": $\Phi = \bigcup_{\psi=1}^{\Psi} \Phi_\psi$.

Order each class according to the order in which the flights in the class are scheduled to be performed by the aircraft defining the class: $\Phi_\psi = \{\phi_{\psi 1}, \ldots, \phi_{\psi, \Xi(\psi)}\}$.

Order the classes, e.g., in decreasing order of the cost of their first flight, and break ties, e.g., according to the increasing order of scheduled arrival times for first flights.

FOR $\psi = 1$ **TO** Ψ **DO**:
  FOR $\xi = 1$ **TO** $\Xi(\psi)$ **DO**:
    Set $\phi = \phi_{\psi\xi}$.
    IF $\xi = 1$ **THEN**:
      Define $\mathcal{T}_\phi := \mathcal{T}_\phi^a$.
      IF ϕ has a previous noncancelled flight ϕ' **THEN**:
        Remove from $\mathcal{T}_\phi$ those t that are smaller than $r_\phi + g_{\phi'} - s_{\phi'}$ (because, if ϕ were assigned to such a t, then the coupling constraint linking ϕ and ϕ' would be violated).
      END IF
    END IF
    Define the capacity slacks $S_\phi(t) := R_{k_\phi^a(t)} - \sum_{f: k_f^a = k_\phi^a} v_{ft}$, $t \in \mathcal{T}_\phi$
    Define $\mathcal{T}_\phi := \{t \in \mathcal{T}_\phi: S_\phi(t) \geq 1 - v_{\phi t}\}$.
    IF $\mathcal{T}_\phi = \emptyset$, **THEN**
      Cancel ϕ: Put $z_\phi = 1$, $v_{\phi t} = 0$, $t \in \mathcal{T}_\phi^a$.
      CONTINUE ξ
    END IF
    Assign current flight to τ, the smallest element of $\mathcal{T}_\phi$: set $z_\phi = 0$, $v_{\phi\tau} = 1$, $v_{\phi t} = 0$, $t \in \mathcal{T}_\phi^a \backslash \{\tau\}$.
    IF ϕ has a next flight $\hat{\phi}$ **THEN**:
      IF $\tau - r_\phi - s_\phi > g_{\hat{\phi}}$ **AND** $\hat{\phi} \notin \Phi$ **THEN**
        Include $\hat{\phi}$ in Φ_ψ as $\phi_{\psi,\xi+1}$ and modify subsequent indices ξ accordingly.
      END IF
      Define $\mathcal{T}_{\hat{\phi}} = \{t \in \mathcal{T}_{\hat{\phi}}^a: t - r_{\hat{\phi}} \geq \tau - r_\phi - s_\phi\}$.
    END IF
  CONTINUE ξ
CONTINUE ψ
END.

ACKNOWLEDGMENT

We thank the Charles S. Draper Laboratory for supporting work on this project. The research of D. Bertsimas was partially supported by the National Science Foundation with a Presidential Young Investigator award DDM-9158118. We also thank undergraduate students, Mercury Schroeppel and Beryl Castello, for performing some of the computational work reported.

REFERENCES

ANDREATTA, G., AND G. ROMANIN-JACUR. 1987. Aircraft Flow Management Under Congestion. *Trans. Sci.* **21**, 249–253.

ODONI, A. R. 1987. The Flow Management Problem in Air Traffic Control. In *Flow Control of Congested Networks*, A. R. Odoni, L. Bianco, and G. Szego (eds.). Springer-Verlag, Berlin, 269–288.

RICHETTA, O. 1991. Ground Holding Strategies for Air Traffic Control Under Uncertainty. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Mass.

TERRAB, M. 1990. Ground Holding Strategies for Air Traffic Control. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Mass.

VRANAS, P. B. 1992. The Multi-Airport Ground-Holding Problem in Air Traffic Control. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Mass.

VRANAS, P. B., D. J. BERTSIMAS AND A. R. ODONI. 1994. Dynamic Ground-Holding Policies for a Network of Airports. *Trans. Sci.* (to appear).