Allocating Banks of Flights to Arrival Slots in Reduced-Capacity Situations

by

Paul M. Carlson

B.S., Computer Science
University of Minnesota, 1989

M.S., Computer Science
Oregon State University, 1995

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1997

Author ....
Department of Civil and Environmental Engineering
May 9, 1997

Certified by ....
Amedeo Odoni
Professor of Aeronautical and Astronautical Engineering
Thesis Supervisor

Accepted by ....
Joseph M. Sussman
Chairman, Departmental Committee on Graduate Students

Allocating Banks of Flights to Arrival Slots in Reduced-Capacity Situations

by

Paul M. Carlson

Submitted to the Department of Civil and Environmental Engineering
on May 9, 1997, in partial fulfillment of the
requirements for the degree of
Master of Science in Transportation

Abstract

To date, the decision support models developed to assist an airline that is facing disruption of its normal operating schedule have, with very little exception, ignored the special consideration that operations at hub airports require. Instead of considering the dependencies induced by flights full of connecting passengers, models have incorrectly tended to view the passengers of these flights as either terminating at the hub, or continuing on the same flight (through passengers). In addition, the objective function of many models is based solely on customer service metrics, a situation at odds with the airline as a profit-maximizing organization.

Due to the two limitations just described, we believe that the existing models are of limited use to airlines who seek to maximize profit by operating a schedule of flights over a hub-and-spoke network. Unfortunately, this describes the majority of the large U.S. airlines.

In this research we present a series of three mixed integer models that are free from the above limitations. We then test and compare the models using a real-world scenario involving over 300 flights spanning 14 hours. One model stands out and is able to solve the real-world scenario in real time. In addition, we present an extensive literature review and classification of the decision support models developed to assist an airline facing schedule disruption.

Thesis Supervisor: Amedeo Odoni
Title: Professor of Aeronautical and Astronautical Engineering

Acknowledgments

I am grateful to my advisor, Professor Amedeo Odoni, for the guidance he provided me and for the freedom he gave me to define the scope of this research.

Professor Cindy Barnhart deserves credit for stimulating my interest in aviation irregular operations decision support models, and I thank her for having the confidence to select me as a referee for *Transportation Science*.

Steve Kolitz of Draper Laboratory has always gone out of his way, often on very short notice, to ensure I had the technology needed to implement and test the models described in this thesis.

Thanks to my colleague and friend Brian Rexing for the camaraderie we've shared and for challenging me to match his high standard of excellence.

And of course, tons of thanks to my best friend and future wife Randi Mosler for all the support and meaning she brings to my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter introduces the context of the research described in this thesis. First we describe the current state of Air Traffic Flow Management (ATFM) in the United States and how ATFM is expected to evolve in the next 5 to 10 years. Then we introduce and define irregular operations, the airlines' response to a disruption in their planned schedule. Next we describe hub-and-spoke networks and the corresponding banking operations which are particularly sensitive to schedule disruption. Then we give some advantages of using profit maximization as the objective function for decision support models. Finally we describe the problem scenario considered in this thesis, followed by the contributions of this research and an outline of the rest of the thesis.

1.1 The Coming Changes in Air Traffic Flow Management

Air Traffic Flow Management (ATFM) is the management and control of aircraft operating through airports and airspace sectors in a manner that achieves safe, orderly and efficient movement of traffic. In the United States ATFM is poised to undergo some evolutionary, yet dramatic, changes. The primary elements of the change are captured in the phrase *collaborative decision making* (CDM). CDM is an evolutionary step from an environment of centralized decision-making by the FAA to a partially-decentralized one, where each airline is directly involved in the decision-making process and thus has the flexibility to manage its own operations according to its own priorities and objectives. CDM completely reverses the current system: today, the airlines provide information to the FAA, who as centralized decision maker makes unilateral decisions (with limited adjustments available to the airlines); under CDM, the FAA will inform the airlines about the state of the system and allocate constrained resources among the airlines, then the airlines will make their own decisions based on this knowledge.

There are a number of factors pushing this evolutionary change. First, as system demand grows without a corresponding increase in supply, ATFM is becoming increasingly important to the smooth operation of the air traffic system and to the efficient utilization of today's resources. Unfortunately, the rules, policies and procedures currently in place are inflexible and too restrictive, and are thus preventing ATFM from reaching its full potential [20]. Second, new technologies are emerging, ranging from GPS navigation systems to automated decision support tools, that enable the participants in the air traffic system to efficiently deal with the dynamic and uncertain environment of ATFM—uncertainties and dynamism that were heretofore inefficiently dealt with through unnecessary and overly-restrictive rules. And third, the airlines have expressed their preference for a decentralized decision-making environment [13].

One area that will undoubtedly change—in fact, hopefully to the point of disappearing—is the FAA ground-holding program. Under the present system, when the number of anticipated arrivals at an airport exceeds its capacity, the FAA assigns landing times to these arrivals on a first-come, first-served basis. (In practice each flight is given a window of time to land at the destination airport.) The FAA then calculates the delay times resulting from these assignments, and holds these flights at their origin airport for the length of the delay. This process is called a Ground Delay Program (GDP), and its motivation is that when delay is inevitable it is cheaper and safer to take this delay on the ground rather than in the air (air delay is typically taken either by circling a destination airport or traveling at slower speeds en route).

The current system does allow the airlines to make limited adjustments to the landing assignments, or arrival slots, they are given under a GDP. The adjustment process is often initiated by a cancellation, which opens up an arrival slot. The airline can then move up, or substitute, a later flight into the open slot, thereby reducing the delay of that flight. And since this step opens up another slot—that one originally assigned to the moved-up flight—the airline can thus engage in a string of substitutions that in the end allows one flight cancellation to reduce the delays of several flights. Today, this substitution process is the primary rescheduling tool available to airlines during GDPs. Unfortunately, the GDP and slot substitution process has a number of restrictions and inefficiencies. For instance, the FAA assigns landing times on the assumption that all flights have equal priority. Obviously the airlines do not view all flights as equal, but there is presently no means for their priorities to influence the assignment process.

Under the proposed partially-decentralized scenario, when demand exceeds supply at an airport the FAA will execute an arrival slot rationing program (typically but not necessarily based on the original schedule, or 'ration-by-schedule'), and will then inform each airline of the time window and number of its arrival slots. The airlines will then themselves decide how to assign their flights to their arrival slots, based on their unique priorities and objectives. (The airlines will also be free to decide when these flights depart from their origin airports, in effect deciding the balance between ground-based and airborne delay.) The airlines then submit their assignments to the FAA, perhaps for approval. It should be noted that this process is dynamic: as conditions change, the airlines may see their arrival slot allocation either increase or decrease. This new process involves collaboration, cooperation, coordination, and information sharing between the FAA and the airlines, characteristics that will be typical of ATFM in the future.

Most of the existing ATFM decision support models were formulated—and appropriately so—under the restrictive FAA slot swapping regulations. However, collaborative decision making has been labeled a high-priority near-term effort and is gaining increased momentum in the aviation community [20]. Models formulated with this new environment in mind will both display today the benefits of the upcoming decentralized ATFM and will be used tomorrow as first-generation tools when the evolution is complete.

1.2 Irregular Operations

In response to demand-capacity imbalances caused by congestion or weather or both, airlines enter into irregular operations, or the management of and recovery from disruptions to scheduled service. Irregular operations can be initiated to handle relatively modest changes in schedule, for example due to congestion or mechanical breakdowns, or to handle heavy delays and cancellations mainly caused by severe weather. Dealing with such major disruptions to scheduled operations is sometimes called high-volume irregular operations.

Weather can cause an airport's capacity to accept arriving flights to be reduced either incrementally, due to the increased separation requirements of IFR flight, or drastically, due to the inability to safely land aircraft at all. For example, cloud cover might reduce capacity by 10-15%, whereas a severe thunder squall with high winds can reduce capacity to near zero. Reductions of the second type are typical of the problem scenarios that the models described in this thesis are intended to help solve; although the models can solve scenarios involving any capacity reduction, those situations with a severe reduction often have the greatest need for automated support. Solutions to these scenarios are not trivial: aircraft rescheduling and rerouting decisions produce flight cancellations and delays that affect maintenance scheduling, crew scheduling, and of course passengers.

As we will see in the Literature Review and Classification chapter, irregular operations often involves two highly dependent steps: schedule reduction and schedule recovery. Although the models in this thesis are intended to assist primarily in schedule reduction, the importance of schedule recovery and its dependence on schedule reduction are realized, and mechanisms are available to ensure that the models' recommendations facilitate the recovery process.

1.3 Hub-and-Spoke Networks

In the aftermath of airline deregulation in 1978 most U.S. airlines adopted hub-and-spoke networks for their domestic operations. Hubbing, the process of routing origin-destination traffic through a connecting airport rather than serving it non-stop, offers many advantages for airlines as well as their passengers, all based on the following principle: under hubbing, a flight from a given origin to the hub can be used by passengers having different destinations, and a flight from the hub to a given destination can be used by passengers coming from multiple origins.

For airlines, consolidating traffic flows to meet at a common point for redistribution maximizes the number of markets served by their network of flights. Hubbing also allows the airlines to take advantage of economies of aircraft size, achieve high levels of aircraft utilization, and in general optimize the efficiency and productivity of flight operations.

For passengers, hubbing results in an increased frequency of service in a given origin-destination market, and in markets not large enough to support direct service, it is responsible for the very existance of service.

The successful use of hubbing requires careful scheduling to balance the conflicting goals of linking the greatest number of city pairs, while minimizing the time that passengers spend at the hub and maximizing the utilization of aircraft. To achieve this balance, many flights—called an arrival bank—are scheduled to arrive at a hub airport within a short interval of time. After a minimum connecting time needed for redistribution of passengers and their baggage and for aircraft servicing, many flights—called the departure bank—are scheduled to depart from the hub.

We assume in this thesis that in order to maximize the number of markets served, arrival and departure banks do not overlap, an assumption typically borne out in practice. There are, however, some scheduling scenarios (beyond the scope of this thesis) in which these banks will legitimately overlap (see [7]). Also note that in this thesis we may refer to arrival banks as inbound banks, and departure banks as outbound banks.

The careful scheduling required for efficient hub operations results in a critical problem: schedules are now acutely sensitive to even the tiniest disruption, due to the existance of *bank-induced dependencies*. For example, the effect of delay is no longer confined to a single flight, as a single inbound delay (FAA-imposed or otherwise) may cause the airline to impose multiple outbound delays in an attempt to maintain the integrity of the bank, to accommodate connecting passengers. But if the outbound delays are too excessive, an airline may choose to 'separate' the delayed inbound flight from its bank, in which case the outbound bank of flights will not wait for its arrival. The cost of separation is that the connecting passengers on the separated inbound flight will now miss their connections to flights in the outbound bank, and will be delayed at the hub airport for some period of time. This trade-off between outbound flight delay and inbound flight separation is the main innovation of the models presented in this thesis.

As we will see in the literature review of Chapter 2, very little work in aviation operations research related to irregular operations has considered the bank-induced dependencies due to hub operations, which is unfortunate since hub-and-spoke networks appear here to stay. Ghobrial and Kanafani [8] expect network hubbing to continue, as projected by a network equilibrium model, and Dennis [7] agrees, saying that "the laws of mathematics and geography mean that the advantages of hubs are here to stay."

1.4 Airlines are Profit-Maximizing Organizations

That the U.S. airlines are profit-maximizing organizations seems to be common knowledge, yet many aviation decision support models seem to forget this. Instead, we see statements like "an airline's first priority is to keep the number of cancellations to a minimum" and objective functions that minimize passenger delay minutes, minimize the number of passengers missing connections, minimize the maximum delay, or maximize on-time performance, as if the airline were a public transit organization. Of course, these metrics are enormously important and by no means are we suggesting that they be ignored. But a single-minded focus on passenger service, especially when taken to an extreme (i.e., optimized for), can be very detrimental to the bottom line.

We suspect that operation-based objective functions are chosen over profit-based ones because, while the former are easy to quantify, the latter require estimation of costs that are often unknown. But ignoring the issue does not make it go away. *It is our strong opinion that any decision support tool intended for use by an airline should have profit maximization as its objective function; if not, the reasons for choosing a different objective function over profit maximization should be clearly stated.* Since scheduled operations were presumably constructed with profit maximization in mind, for irregular operations decision support models it is sufficient to *minimize the total cost of changes to scheduled operations.*

Using maximization of profit as the objective function has the further advantage of using dollars as a universal common denominator. At least three advantages arise from this:

- While some models are touted for their ability to either minimize customer dissatisfaction or maximize schedule execution smoothness, they cannot examine the trade-off between these often-conflicting goals. But if the costs associated with these metrics can be estimated and applied, the trade-off can be captured and solved.

- Most models cannot account for all operating constraints due to the limitations put on the size and complexity of the models by real-time solution requirements. But these missing constraints can be indirectly represented in the cost variables of a profit-maximizing objective function. For instance, a flight using an aircraft with impending maintenance requirements may have high cancellation costs if the destination of the flight has a maintenance depot.

- The airlines often have differing philosophies concerning irregular operations and rather than construct a new model for each philosophy, a model that maximizes profit should be of use to all airlines, since their differing philosophies can be expressed through assigning different values to the input cost variables. For example, one operating philosophy is "one bank on time is better than two banks delayed," and with the proper cost structure (high bank spread costs and high flight separation costs) this philosophy will be represented in the model's solutions.

1.5 The Problem Scenario

The scenario addressed by the models in this thesis is the following: severe weather strikes an airline's hub airport and arrival capacity is reduced. In accordance with the partially-decentralized ATFM environment described in Section 1.1, the FAA notifies the airline of its arrival slot allocation and the airline must thus make a real-time tactical decision about how to reschedule its banks of flights into the hub, through assignment of flights to the limited arrival slots. Specifically, the airline must decide whether to delay the completion time of a bank and by how much (called 'spreading the bank'), and it must choose a subset of flights to stay in the bank, separating or canceling the rest.

It is important to realize that since the constrained resource in this scenario is arrival slots, delays and cancellations are inevitable. This point is made because many model formulations in the literature treat cancellation and delay as a somewhat optional strategy to achieve other ends (for instance, delaying flights to avoid canceling too many [21], or delaying and canceling flights to maintain schedule balance [12]); these formulations do not include constrained arrival slots.

1.6 Contributions and Outline

This research makes the following two primary contributions:

1) We explain the importance of including bank-induced dependencies in decision support models, and we present, as an improvement to a previous model, a series of decision support models that an airline can use to tactically schedule arriving banks of flights into a hub airport in reduced-capacity situations. We test the models using a real-world scenario, and discover that one of them finds optimal solutions in real-time.

2) We present an extensive literature review and classification of the decision support models developed to date for an airline to use during irregular operations.

Chapter 2 contains a literature review that examines other models that have been formulated for irregular operations and classifies these models according to the problem scenario they are meant to solve. Chapter 3 introduces and critiques Milner's Cancellation/Delay Model, the evolutionary forefather to the new models of this thesis; these new models are then presented and explained in Chapter 4. In Chapter 5 we test and compare the new models using a real-world scenario, and Chapter 6 offers a conclusion and a look at future research.

Chapter 2

Review and Classification of Decision Support Models for Irregular Operations

2.1 Introduction

Irregular operations in aviation is the subject of a fair amount of previous and ongoing research, and the decision support models that have been developed can be categorized as those intended for airline use, and those for use by the FAA. In this chapter we first review those models in the first category—intended for airline use—and then we classify these models according to the different problem scenarios they are intended to solve. Finally, we close this chapter by reviewing the few models developed for the FAA that explicitly include banks of flights and their special dependencies.

2.2 Review of Models for Airline Use

Vasquez-Marquez [19] describes a network-based heuristic algorithm implemented at American Airlines in 1989 that reduces passenger delays due to ground hold programs. Called the Arrival Slot Allocation System (ASAS), the heuristic takes the flight schedule, the FAA-assigned times of arrival, and one or more flight cancellations as input and returns a slot substitution sequence that approximately minimizes overall passenger-delay minutes. A disadvantage of the approach is that all passengers are considered to be terminating;

that is, the additional delay that passengers incur while waiting for their connections is not considered. ASAS is an interactive decision support tool, and the human controllers are encouraged to use any additional knowledge they have to influence the heuristic's solution.

In 1990, Teodorovic and Stojkovic [17] published a model and a heuristic solution algorithm that produces a new daily schedule and aircraft rotation when an airline is faced with a shortage of aircraft. The model (and algorithm) has a two-part objective function: the primary objective is to minimize flight cancellations; any ties are broken by choosing the solution with minimum overall passenger delay on the flights not canceled. The usefulness of the model is called to question by Cao and Kanafani [3], who show that the model has a trivial optimal solution—the original schedule.

Teodorovic and Stojkovic [18] have extended their earlier work by accounting for additional operational constraints (crew requirements receive the greatest emphasis). Essentially, after any change to an airline's operational schedule, they produce entirely new aircraft and crew rotations for the affected fleet type. Since this is a difficult problem, they use a heuristic approach that in some cases produces a worse solution than the naive approach of canceling the sequence of flights directly affected by the disturbance [18]. The first step of their algorithm is finding a new crew rotation for the flight schedule according to the same two-part objective function as in their previous work, bearing in mind the appropriate crew operating constraints. Presumably, if the schedule disturbance is due to aircraft problems the existing crew rotation could be used, but they don't mention this. Each crew rotation is now considered to be a single 'big leg', and the second step is finding an aircraft rotation over the resulting schedule made up of these 'big legs', again according to the objective function and bearing in mind the appropriate operating constraints. The final step is checking the aircraft rotation schedule against maintenance requirements. If there are no maintenance conflicts the algorithm is finished; otherwise, the human dispatcher must modify the model inputs (for example, by canceling a flight or changing a flight's departure time) and then re-run the algorithm. One effect of the 'big leg' approach is that a crew will be assigned to a single aircraft for their entire rotation; it seems that this restriction could greatly limit both the number of feasible solutions and the quality of the algorithm solution. In addition, since the objective function minimizes flight cancellations with no regard to delay, the new aircraft rotation could produce new departure times involving significant amounts of delay. Teodorovic and Stojkovic present two heuristic approaches for finding the new aircraft and crew rotations. The first uses the first-in, first-out principle, where every arriving aircraft or crew is assigned to the first available departing flight. The second uses a sequential approach based on dynamic programming. That neither of these approaches involves any optimization techniques, together with the fact that the overall algorithm is also sequential accounts for the potentially poor performance mentioned earlier. In the final analysis, the algorithm can best be described as an automation of the way a human dispatcher might go about finding a solution somewhat better than the naive approach, and therefore its usefulness is probably limited to finding feasible solutions to larger-sized problems.

In 1993, Jarrah et al [12] introduced the paradigm of using minimum-cost network flow models to find an operable, system-balanced flight schedule when aircraft shortages disrupt an airline's scheduled network of flights. They present two models, one that uses flight delay (only) at a single airport to absorb the aircraft shortages and one that uses flight cancellations (only) across multiple airports to absorb the shortages; both models allow aircraft swapping between flights and the use of spare aircraft. They use a recursive cost equation to capture the downline effects of delay, and the equation includes the cost of missed connections. When discussing the limitations of their work, they acknowledge the hub-and-spoke system, but conclude that including bank-related dependencies when determining cancellation and delay costs is too difficult. The delay model was implemented at United Airlines [16]; the implementation allows human controllers to use their knowledge and experience to tailor the inputs of the model.

Cao and Kanafani [3, 4] are extending Jarrah et al's work to simultaneously consider both cancellations and delays, and the trade-off between them, when finding an operable, system-balanced flight schedule. In essence, they have taken Jarrah et al's delay model's single-airport representation and extended it to include multiple airports, flight cancellations, and ferrying of spare aircraft. Their objective function is profit maximization, and they compute explicit downline delay costs.

Yan and Yang [21] use a time-space network as a framework for simultaneously making delay, cancellation, and ferrying decisions to recover from a single aircraft shortage. Their model returns fleet flows, not aircraft routings, but the biggest deficiency is in their objective function. Their stated primary objective is to minimize the time that the airline's schedule is perturbed, but in reality they only bound this recovery time. Their method is to allow the repaired aircraft to be ferried to any airport in the airline's network, guaranteeing a feasible solution, with the recovery ending time being a function of these ferry times, among other things. Inside the resulting time bound, their model now finds the fleet flow that maximizes operating profit, their secondary objective. The deficiency is the primary importance given to minimizing the recovery time, for it is easy to see that this strategy omits from consideration potential low-cost solutions. For instance, it may be the case that the repaired aircraft can be absorbed back into the schedule at the station it was repaired at, but at a time shortly past the computed time bound. Unfortunately, the model is not allowed to consider this solution, and instead may suggest ferrying the aircraft to a distant location, certainly a more expensive solution.

Clarke [5, 6] is developing an ambitious model to assign aircraft to flight sequences when severe weather disrupts an airline's operations. The objective function of his model minimizes cost, and the model considers such factors as loss of revenue due to spill, arrival slot constraints, crew availability, airport ground capacity, seating capacity, and maintenance-related aircraft utilization bounds, although not all as hard constraints. Clarke is also developing both heuristic and optimal solution methodologies; the results of applying the methodologies to realistic-sized problems have not yet been reported.

None of the models discussed thus far considers banks and their special dependencies. While some of the models include the concept of connecting passengers, most of the models consider all passengers to be terminating at the destination airport. This approach is particularly dangerous for models whose objective function is the minimization of passenger delay. Reducing the delay of an inbound flight full of connecting passengers by 15 minutes is a hollow improvement if the outbound flights those passengers are connecting to are delayed even longer.

One of the first decision support models to consider banks and their special dependencies is found in a 1994 MIT term paper by Mette [14]. He developed a sequential heuristic algorithm for scheduling the flights of an outbound bank when the inbound bank experienced delays. The heuristic, which minimizes cost, considers the trade-off between the costs of delaying outbound flights to allow connections from the inbound bank and the costs of missed passenger connections if the outbound flights are not delayed. While the heuristic does include the availability of spare aircraft, it does not take into account the possibility of flight cancellations in either the inbound or outbound banks. Another shortcoming is the heuristic's naive use of the First-Idle-First-Used principle when assigning aircraft from the inbound bank to flights of the outbound bank; the assumption is made that an aircraft can be assigned to an arbitrary flight simultaneously and at zero cost. Finally, as befits a term paper, the heuristic is not particularly elegant and it was not executed on examples of realistic size.

Milner [15] took an extensive look at how the airlines might both participate in and respond to the allocation of arrival slots in a partially-decentralized ATFM environment. To increase airline participation, he proposed a market-based approach to allocation—an auction of arrival slots, administered by the FAA. To determine airline response, he formulated a number of models, intended for various users and uses, that account for the bank structure of flights scheduled into hub airports. For his Cancellation/Delay Model with Connection Information—discussed in the next chapter and improved on in Chapter 4—he devised a heuristic solution technique based on dynamic programming. His work is the basis for a large majority of the research described in this thesis.

2.3 Classification of Models for Airline Use

Table 2.2, found at the end of this chapter, gives a classification of the irregular operations decision support models intended for airline use developed to date. For each model we give the author and year of publication, the classification of the model according to the problem scenario the model is intended to solve, the constrained resources of the problem scenario that prevent the schedule of flights from being executed as planned, the solution strategy used to solve the model, the objective function of the model (primary objectives are denoted by (1), secondary objectives by (2)), an indication of whether the model considers banks of flights and their dependencies, and an indication of whether the model has, to our knowledge, been used operationally by an airline.

The first category of models reschedules flights when a resource shortage occurs. Schedule reduction is not considered; in other words, all flights are scheduled. It should be noted that there are many similar models intended to be used by the FAA that could be modified for airline use. The second category of models does include flight cancellations in the models' decision-making process, allowing schedule reduction. Like the rescheduling models, the scope of the reduction models developed so far has been limited to flights at a single airport;

the effects of solutions on the airline's network of flights is not directly considered, but can be indirectly considered through the input costs of the models. Since improvements in the reduction process have the greatest pay-off at airports with a large number of operations, these reduction models have been a natural fit to consider bank operations at hub airports. Although the two models presently in this category both assume a partially-decentralized ATFM environment, a reduction model could be formulated to operate in the present environment. The third category of models addresses schedule recovery in the face of aircraft and/or crew shortages. Since the goal of these models is an operable, system-balanced network of flights, their scope is the entire network of flights. Although the shortage is typically absorbed into the present schedule via flight cancellations and/or delays, it should be noted that schedule reduction is by no means an inherent part of the solutions these models return. Ideally, the aircraft shortage can be absorbed by utilizing the slack time built into the operating plan of the airline.

Are the recovery models a generalization of the reduction models? There does not appear to be a uniform answer. It does seem that arrival slot constraints could, in theory, be added to some of the recovery models, but only in the form of side constraints, which will affect the performance of the minimum-cost network flow models. Also, it is not obvious how to include such concepts as banks, bank spread, and flight separation in the existing recovery models. Instead, in their current form the reduction and recovery models can be applied sequentially to determine an overall solution. A reduction model will assign banks of flights to limited arrival slots, and due to flight cancellations and delays, its solution will result in aircraft shortages at the destination airport and aircraft surpluses at the originating airports. Using these shortages and surpluses and the reduction model's flight schedule as input, a recovery model can then produce an operable schedule for all remaining flights. The reduction model can assist in the recovery step by including the costs of recovery in its input costs. For example, flights from an airport that has very little activity may have high cancellation costs, to reflect the fact that recovering a surplus aircraft from that airport may be difficult. Of course, keep in mind that in those situations in which the airline has no control over the reduction step, for instance when an aircraft becomes unavailable due to mechanical problems, the recovery model can be executed directly.

Unfortunately, applying the reduction and recovery models sequentially is not ideal. This gives rise to the final category in our classification: models that solve schedule reduction and recovery simultaneously. To our knowledge, Clarke's is the first attempt to solve the problem in this way.

2.4 Review of Models for FAA Use

Models that assist the FAA in improving air traffic flow management, most frequently by suggesting improved ground-holding programs, are for the most part beyond the scope of this thesis. The objective of those models is generally to assign every flight to an arrival slot, minimizing delay in some way. Rarely are flight cancellations an option or bank-induced dependencies considered. There are a couple exceptions, which we review in this section.

Hoffman and Ball [11] present and compare four model formulations, each having an alternative way of adding banking constraints to the Single-Airport Ground Holding Problem. Since the output of their models is a ground-holding program, flight cancellations are not an option; that is, all flights must be assigned to an arrival slot. Their approach to banks is to enforce a fixed-length time window for the arrival of bank flights—called the bank width—to prevent bank spread. Unfortunately, explicitly enforcing a bank width prevents the model from considering more general solutions, of which one may be optimal. For example, consider the scenario described in Table 2.1, in which the bank width is set to three time periods. Now, due to the bank width requirement, the fifteen bank flights are forced to arrive only during time periods two through four, which has the following undesirable results: first, the three arrival slots during time period one are left unused, needlessly wasting three scarce resources. Second, even though there are 18 arrival slots and 18 flights, the three independent flights will be unnecessarily delayed to time period five or later as it is unlikely that any of those flights can be moved up to arrive in time period one, three time periods earlier than scheduled. Third, at least three of the bank flights (for example, the three that could have arrived during time period one) are unnecessarily delayed at their origin airport. While it is true that the inbound bank cannot be completed until time period four in either case, given the presence of terminating passengers, the shuffling of connecting passengers and their baggage, and the importance of on-time dependability statistics, it seems preferable for flights to arrive at the hub as early as possible and wait there for the inbound bank to complete. Finally, we note that if the capacity in time periods two through four is reduced any further, the models will find no feasible solutions.

Time Period	Capacity	Scheduled Activity
1	3	Five arriving bank flights
2	5	Five arriving bank flights
3	5	Five arriving bank flights
4	5	Three arriving independent flights

Table 2.1: A Schedule with Bank Width of Three Time Periods

In their paper on the air traffic flow management problem with enroute capacities, Bertsimas and Stock [2] discuss how their model can be extended to account for banks of flights (the extension has apparently not been implemented). Their approach is to minimize the time between the arrival of the first flight and the last flight of the inbound bank. Although this approach is slightly different from that used by Hoffman and Ball, it suffers from the same shortcomings: arrival slots may be needlessly wasted, and both bank flights and independent flights may be unnecessarily delayed. However, their approach does not suffer from the problem of finding no feasible solutions, as the bank is allowed to spread out, if needed. Finally, although they do not mention it, when adding the minimization of bank spread to the objective function they will probably need to introduce a bank spread cost to maintain compatibility with the current objective of cost minimization.

Author and Year	Classification	Constrained Resources	Solution Strategy	Objective Function	Banks?	Implemented?
Vasquez-Marquez 1991	Reschedule	Arrival Slots	Network-based Heuristic	Min passenger delay	No	Yes
Mette 1994	Reschedule	Inbound Flights (Delays to)	Heuristic Algorithm	Min cost	Yes	No
Milner 1995	Reduction	Arrival Slots	Heuristic Algorithm	Min cost	Yes	No
Carlson 1997	Reduction	Arrival Slots	Integer Programming Optimization	Min cost	Yes	No
Teodorovic et ux 1990	Recovery	Aircraft	Heuristic Algorithm	Min cancellations (1) Min passenger delay (2)	No	No
Teodorovic et ux 1995	Recovery	Aircraft, Crew	Heuristic Algorithm	Min cancellations (1) Min passenger delay (2)	No	No
Jarrah et al 1993	Recovery	Aircraft	Network-based Optimization	Min cost	No	Yes
Cao and Kanafani 1996	Recovery	Aircraft	Integer Programming Approximation	Max profit	No	No
Yan and Yang 1996	Recovery	Aircraft (single)	Network-based Heuristic and Optimization	Min time of recovery (1) Max profit (2)	No	No
Clarke ??	Reduction and Recovery	Arrival slots, Aircraft	Not yet reported	Min cost	No	No

Table 2.1: A Classification of Irregular Operations Decision Support Models for Airline Use

Chapter 3

Milner's Cancellation/Delay Model with Connection Information

3.1 Introduction

In 1995, Milner [15] presented his Cancellation/Delay Model with Connection Information. (Although the name suggests it, Milner's model does not include any connection information and will hereafter be referred to as the Cancellation/Delay Model.) Milner's model does represent a significant advance over prior work, as it explicitly includes the bank-induced dependencies of hub operations. Unfortunately, Milner's model contains rather severe errors and other shortcomings; fortunately, these errors and shortcoming are correctable, as will be seen in the models presented in the next chapter. This chapter describes the Cancellation/Delay Model, and discusses the opportunities for improvement associated with it.

3.2 Model Description

The Cancellation/Delay Model's input and decision variables are shown in Figure 3-1, and the model formulation is given in Figure 3-2 (see pages 109–113 of [15]). The input variables are mostly straightforward. Note that every flight belongs to a bank (i.e., there are no independent flights). Rescheduling a canceled flight is an indirect method of separating the

Input Variables

F	a set of flights, each having a scheduled arrival time
B	a set of banks, each having a scheduled completion time equal to the scheduled arrival time of the latest flight in the bank
T	a set of discrete time periods (each 15 minutes in length)
$\mathrm{m_t}$	the number of arrival slots allocated to the airline in time period t
$\mathrm{w_{bt}}$	the cost of spreading bank b to complete at time period t
$\mathrm{c_{f0}}$	the cost of canceling flight f
$\mathrm{d^s_{ft}}$	the benefit of rescheduling the canceled flight f at time t

Decision Variables

z_{bt}	1 if bank b is assigned to be completed *at or before* time t; 0 otherwise
y_{ft}	1 if flight f is assigned to arrive *at or before* time t; 0 otherwise
y_{f0}	1 if flight f is canceled; 0 otherwise
y^s_{ft}	1 if canceled flight f is rescheduled to arrive *at or before* time t; 0 otherwise

Figure 3-1: The Cancellation/Delay Model's Input and Decision Variables

flight from its bank; of course, the rescheduling benefit should decrease over time. Similarly, we expect the bank spread cost to increase over time. Since severe weather may completely shut down a runway, we allow the number of arrival slots in a given time period to be greater than or equal to zero. Three of the four decision variables have the so-called 'step function' behavior (enforced in the model's constraints). That is, when they become 1 at some time period t, they stay 1 through all successive time periods.

The objective function minimizes the costs of bank spread and flight cancellation; a benefit is credited if a canceled flight is rescheduled for a later time. Constraint 1 ensures that all banks are eventually completed. Constraint 2 ensures that a bank is not completed until all its flights have either arrived or, in conjunction with Constraint 3, have been canceled. Constraint 3 prevents banks from overlapping; that is, a flight in a given bank cannot arrive until all preceding banks have been completed. Constraint 4 prevents a flight from being rescheduled (a positive action, since a benefit occurs) unless the flight has been canceled. Constraint 5 is the set of arrival slot capacity constraints. Constraints 6 through 8 enforce the step-function behavior of three of the decision variables. Constraint 9 enforces integrality of all decision variables.

The model also implicitly constrains flights from arriving before their scheduled arrival times and banks from being completed before their scheduled completion times. This is

$$\text{Minimize } \sum_b \sum_t \mathrm{w_{bt}}(z_{bt} - z_{bt-1}) + \sum_f \mathrm{c_{f0}} y_{f0} - \sum_f \sum_t \mathrm{d^s_{ft}}(y^s_{ft} - y^s_{ft-1})$$

subject to

$$z_{bT} = 1 \quad \forall\, b \tag{1}$$

$$z_{bt} \leq y_{ft} \quad \forall\, f \in b, \forall\, b, t \tag{2}$$

$$y_{ft} \leq z_{b't} + y_{f0} \quad \forall\, b' < b, \forall\, t, \forall\, f \in b, \forall\, b \tag{3}$$

$$y^s_{ft} \leq y_{f0} \quad \forall\, f, t \tag{4}$$

$$\sum_f ((y_{ft} - y_{ft-1}) + (y^s_{ft} - y^s_{ft-1})) \leq m_t \quad \forall\, t \tag{5}$$

$$y_{ft} \geq y_{ft-1} \quad \forall\, f, t \tag{6}$$

$$y^s_{ft} \geq y^s_{ft-1} \quad \forall\, f, t \tag{7}$$

$$z_{bt} \geq z_{bt-1} \quad \forall\, b, t \tag{8}$$

$$z_{bt}, y_{ft}, y^s_{ft}, y_{f0} \in \{0, 1\} \tag{9}$$

Figure 3-2: The Cancellation/Delay Model

enforced by not defining the y_{ft}'s and y^s_{ft}'s for time periods before flight f's scheduled arrival time and the z_{bt}'s for time periods before bank b's scheduled completion time.

3.3 Errors in the Model

The formulation of the Cancellation/Delay Model shown in Figure 3-2 contains two errors, one that causes the model to incorrectly have no feasible solutions in certain instances, and one that causes the model to produce output that in almost all instances must be post-processed to be useful.

The first error occurs in Constraints 2 and 3. Notice that, according to Constraint 2, for a bank to be completed *every flight in that bank must be assigned to an arrival slot, even if that flight is canceled.* Proof that this is possible is seen in Constraint 3, which allows a canceled flight to be assigned to an arbitrary arrival slot. The result is that canceled flights are unnecessarily consuming a scarce resource—arrival slots—and when the total number of flights exceeds the total number of arrival slots (over all time periods), the model can find no feasible solution (incorrectly so, since the cancellation of all flights should always be a feasible solution).

This error can be corrected by changing Constraints 2 and 3 as shown in Figure 3-3.

$$z_{bt} \leq y_{ft} + y_{f0} \qquad \forall\, f \in b, \forall\, b, t \qquad (2)$$

$$y_{ft} \leq z_{b't} \qquad \forall\, b' < b, \forall\, t, \forall\, f \in b, \forall\, b \qquad (3)$$

Figure 3-3: The Corrected Cancellation/Delay Model

Now, as seen in Constraint 2, a bank is completed when all its flights have either arrived or been canceled. Furthermore, there is no reason for y_{f0} to appear in Constraint 3, and it has been modified accordingly. Now, when the total number of flights exceeds the total number of arrival slots (over all time periods), some flights are canceled and some assigned to the available arrival slots.

The second error of the Cancellation/Delay Model is that flights are allowed to 'slip', capacity permitting, inside the time span of their bank. A simple example will elucidate this point. For simplicity, let us assume that each time period has infinite capacity. Let us further assume that a bank has three flights, scheduled to arrive one each at time periods 1, 2, and 3, and thus the bank is scheduled to be completed at time period 3. *A solution in which the flights arrive at their scheduled arrival times has the same objective function value as a solution in which all three flights 'slip' to arrive at time period 3, the bank's scheduled completion time.* (Note that the bank itself is not spread in either case, nor are any flights canceled; in fact, both solutions have an objective function value of zero.) Due to this slipping behavior the model cannot be used operationally without undergoing some type of post-processing, since its solution includes (possibly large) amounts of unnecessary flight delay. This error can be corrected by including flight delay costs in the objective function.

3.4 Shortcomings and Other Suggestions

The Cancellation/Delay Model has a handful of other shortcomings that are not severe enough to be labeled errors but which are important enough to be addressed by newer models (and indeed are addressed by the models described in the next chapter). Specifically, newer models should:

- be more general, by being able to accommodate schedules having both flights that cannot be separated from their bank and independent flights. (The Cancellation/Delay

Model probably can accommodate independent flights, but only through the construction of 'one-flight banks'. However, this technique requires artificial bank information and is inefficient, since the artificial bank unnecessarily expands the number of decision variables and constraints.)

- be more realistic, by incorporating flight delay costs—surely a cost of interest to the airlines—in their objective functions. (And as just discussed, including flight delay costs will correct one of the two errors in the Cancellation/Delay Model.)

- be more intuitive and easier to understand. For instance, the Cancellation/Delay Model's unusual 'cancellation–rescheduling' method, which is intended to model flight separation, requires flights to incur their cancellation cost prior to receiving a reward for being rescheduled and adds complexity to understanding the model.

- refrain from enforcing administrative policies through 'soft' constraints (i.e., constraints that do not reflect any physical laws) when those policies can be enforced through appropriate costs in the objective function. For example, Constraint 3 in the Cancellation/Delay Model prevents banks from overlapping; in Milner's words, "this models the goal of airlines to keep banks from spreading so that passengers may transfer quickly." However, bank spread is quite different from bank overlap (after all, if a preceding bank is spread an arbitrary amount, bank overlap is avoided by simply delaying the start of the succeeding bank), and it seems that if low bank spread is desirable, this can be reflected in high bank spread costs.

We conclude that while Milner understood the importance of explicitly considering bank-induced dependencies when solving the schedule reduction phase of irregular operations, his models contain numerous opportunities for improvement.

Chapter 4

Some New Bank Scheduling Models

4.1 Introduction

We present three new bank scheduling models in this chapter; all three models solve the problem scenario described in Section 1.5. The first model is the result of applying the opportunities for improvement outlined in the previous chapter to Milner's Cancellation/Delay Model. The second model is the result of replacing the step-function variables of the first model. The third model takes advantage of the fact that the cancellation decision variables turn out to be unnecessary and is thus the result of removing those decision variables from the second model. After presenting the models we fully explain the meaning of the input cost variables that are so important to both the solution and the solution run-time of the models. We then describe a suite of small test cases that validate the behavior of the models. We conclude the chapter by describing how the models were implemented and explaining an implementation technique that turns out to be crucial to obtaining model solutions in real-time.

4.2 Improving Milner's Cancellation/Delay Model

The model presented in Figures 4-1 and 4-2 is the result of improving Milner's Cancellation/Delay Model by correcting the two errors and rectifying the four shortcomings identified in the previous chapter.

Input Variables

F	a set of flights, each having a scheduled arrival time a_f
Ψ	the subset of independent flights (all other flights belong to a bank)
Φ	the subset of flights that cannot be separated from their bank
B	a set of banks, each having a scheduled completion time c_b equal to the scheduled arrival time of the latest flight in the bank
T	a set of discrete time periods (each 15 minutes in length)
m_t	the number of arrival slots allocated to the airline in time period t
d_{bt}	the cost of spreading bank b to complete at time period t
d_{ft}	the delay cost of flight f arriving at time period t
c_f	the cost of canceling flight f
s_f	the cost of separating flight f from its bank

Decision Variables

z_{bt}	1 if bank b is assigned to be completed *at or before* time t; 0 otherwise
y_{ft}	1 if flight f is assigned to arrive *at or before* time t; 0 otherwise
x_f	1 if flight f is canceled; 0 otherwise
w_f	1 if flight f is separated from its bank; 0 otherwise

Figure 4-1: The First Model's Input and Decision Variables

The input variables have been expanded. First, the set of flights now includes a subset of independent flights, or flights that are not part of a bank. These might be flights from origins whose demand to the hub is large enough to fill an aircraft without connecting passengers, or flights that have connecting passengers but also have travel times, restrictions at the origin airport, or other scheduling problems that prevent the flight from being a formal part of the bank. In either case, since these independent flights are also competing for the scarce number of arrival slots, it is important that the model include them. Second, the set of flights also includes a subset of flights that belong to a bank that cannot be separated from their bank. These might be a small number (four or five) of large-capacity flights in each bank that historically have carried and exchanged such a volume of connecting passengers that the bank has no meaningful definition without them. Since these flights cannot be separated, unless they are canceled (a necessary option for feasibility in complete shutdown situations) the bank must be spread, if necessary, to accommodate their arrival. We point out that this subset can always be empty if no flights have these characteristics. Third, we have added flight delay costs and an explicit flight separation cost, which allows us to remove the rescheduling benefit. The full meaning of all four cost variables are discussed in

$$\text{Minimize } \sum_{b=1}^{B}\{\mathrm{d_{bc_b}} z_{bc_b} + \sum_{t=c_b+1}^{T} \mathrm{d_{bt}}(z_{bt} - z_{bt-1})\} + \sum_{f=1}^{F}\{\mathrm{d_{fa_f}} y_{fa_f} + \sum_{t=a_f+1}^{T} \mathrm{d_{ft}}(y_{ft} - y_{ft-1})\}$$
$$+ \sum_{f=1}^{F} \mathrm{c_f} x_f + \sum_{f=1}^{F} \mathrm{s_f} w_f$$

subject to

$z_{bT} = 1$	$\forall\ b$	(1)
$z_{bt} \leq y_{ft} + x_f + w_f$	$\forall\ f \in b, \forall\ t \geq c_b, \forall\ b$	(2)
$y_{fT} \geq w_f$	$\forall\ f \notin \{\Psi \cup \Phi\}$	(3)
$y_{fT} + x_f = 1$	$\forall\ f \in \Psi$	(4)
$\sum_{f=1}^{F} y_{f1} \leq \mathrm{m_1}$		(5A)
$\sum_{f=1}^{F} (y_{ft} - y_{ft-1}) \leq \mathrm{m_t}$	$\forall\ t \geq 2$	(5B)
$y_{ft} \geq y_{ft-1}$	$\forall\ f, t \geq 2$	(6)
$z_{bt} \geq z_{bt-1}$	$\forall\ b, t \geq 2$	(7)
$w_f = 0$	$\forall\ f \in \Phi$	(8)
$z_{bt}, y_{ft}, x_f, w_f \in \{0, 1\}$		(9)

Figure 4-2: The First Model

Section 4.5.

Flight separation is now included as a decision variable. Note that the notation for a handful of the input and decision variables has been changed to be, we believe, more intuitive.

The objective function minimizes the total cost of changes to scheduled operations by minimizing the sum of the costs of spreading a bank, delaying a flight, canceling a flight, and separating a flight. Constraint 1 ensures that all banks are eventually completed. Constraint 2 ensures that a bank is not completed until all its flights have either arrived, been canceled, or been separated. Alternatively, this constraint ensures that all flights belonging to a bank either arrive or are canceled or separated. Constraint 3 ensures that all flights that are separated from their bank eventually arrive (our definition of separation implies eventual arrival; of course, flights that are canceled are in some sense separated from their bank as well). Since so far we have only considered flights belonging to a bank, Constraint 4 ensures that all independent flights are either flown or canceled. Constraints 5A and 5B represent the set of arrival slot capacity constraints. Constraints 6 and 7 enforce the step-function behavior of two of the decision variables. Constraint 8 enforces the inability to separate those flights so designated. Constraint 9 enforces integrality of all decision variables.

We also implicitly constrain flights from arriving before their scheduled arrival times and banks from being completed before their scheduled completion times. This is enforced by not defining the y_{ft}'s for time periods before flight f's scheduled arrival time and the z_{bt}'s for time periods before bank b's scheduled completion time. If it turns out we want flights to possibly arrive or banks to possibly complete earlier than scheduled, it is a simple matter to define these decision variables for the appropriate time periods.

The output of the model is the optimal-cost bank and flight schedule. All flights will be either flown or canceled. If a flight originally in a bank is flown, it will either remain in its bank or be separated. Each flown flight is assigned an actual arrival time; if this time is later than its scheduled arrival time, the flight has been delayed. Each bank is assigned an actual completion time equal to the latest actual arrival time of flights that have remained in the bank; if this completion time is later than its scheduled completion time, the bank has been spread.

4.3 Replacing the Step-Function Variables

After applying the first model to various test cases and real-world scenarios, and analyzing the structure of the model itself (i.e., examining the number and type of constraints and decision variables), it became clear that the step-function behavior of the y_{ft} and z_{bt} decision variables is very costly to implement. For instance, when applying the first model to the real-world scenario described in the next chapter (8 banks, 304 flights, and 64 time periods), Constraints 6 and 7—those that enforce the step-function behavior—account for 51% of the constraints. So it is clear that if we can remove those constraints without adding additional ones, the size of the model will be significantly reduced.

Are the step-function variables necessary? An examination of the first model reveals that they are not. Are the step-function variables desirable? Milner does not indicate why he defined the decision variables of his Cancellation/Delay Model this way, but we suspect that he was influenced by Bertsimas and Stock [2]. Bertsimas and Stock discovered their model had numerous constraints of the form $\sum_t y_{ft} = 1$. After introducing the step-function behavior, these constraints can be rewritten as $y_{fT} = 1$; then, the constraints and an equal number of decision variables can be removed from the formulation, as $y_{fT} = 1$ can be handled as known parameters of the model. However, our ability in the present case to take advantage of this simplification is limited. While Constraint 1, $z_{bT} = 1$, does have the required form and can indeed be treated as a parameter, these constraints account for a very small percentage (less than 0.05% in the real-world scenario) of the total number of constraints. The overwhelming majority of constraints deal with flights and we do not have any constraints of the form $y_{fT} = 1$, since due to the possibility of cancellation we are not certain, unlike Bertsimas and Stock, whether any given flight will eventually arrive. (Incidentally, Bertsimas and Stock also state that the step-function variables of their model define connectivity constraints that are facets of the convex hull of solutions, a situation that they believe is responsible for an LP relaxation solution that is almost always integral. However, anecdotal evidence has suggested that a new model with the step-function variables replaced outperforms their original model.)

In any case, it seems worthwhile to compare the performance of a model with the step-function variables replaced to that of the first model. The second model, shown in Figures 4-3 and 4-4, is equivalent to the first model with the step-function variables replaced. The

Decision Variables

z_{bt} 1 if bank b is assigned to be completed at time t; 0 otherwise
y_{ft} 1 if flight f is assigned to arrive at time t; 0 otherwise
x_f 1 if flight f is canceled; 0 otherwise
w_f 1 if flight f is separated from its bank; 0 otherwise

Figure 4-3: The Second Model's Decision Variables

$$\text{Minimize } \sum_{b=1}^{B}\sum_{t=c_b}^{T} \mathrm{d_{bt}} z_{bt} + \sum_{f=1}^{F}\sum_{t=a_f}^{T} \mathrm{d_{ft}} y_{ft} + \sum_{f=1}^{F} \mathrm{c_f} x_f + \sum_{f=1}^{F} \mathrm{s_f} w_f$$

subject to

$$\sum_{t=c_b}^{T} z_{bt} = 1 \qquad \forall\, b \qquad (1)$$

$$z_{bt} \le \sum_{j=1}^{t} y_{fj} + x_f + w_f \qquad \forall\, f \in b, \forall\, t \ge c_b, \forall\, b \qquad (2)$$

$$\sum_{t=a_f}^{T} y_{ft} \ge w_f \qquad \forall\, f \notin \{\Psi \cup \Phi\} \qquad (3)$$

$$\sum_{t=a_f}^{T} y_{ft} + x_f = 1 \qquad \forall\, f \in \Psi \qquad (4)$$

$$\sum_{f=1}^{F} y_{ft} \le \mathrm{m_t} \qquad \forall\, t \qquad (5)$$

$$w_f = 0 \qquad \forall\, f \in \Phi \qquad (6)$$

$$z_{bt}, y_{ft}, x_f, w_f \in \{0,1\} \qquad (7)$$

Figure 4-4: The Second Model

objective function and the constraints of the second model have precisely the same meaning as those of the first model.

4.4 Removing the Cancellation Decision Variables

A further simplification can be made when one realizes that the cancellation decision variables are not explicitly needed. Instead, a flight is canceled if it does not arrive during any time period; mathematically, $x_f = 1 - \sum_t y_{ft}$. The third model, shown in Figures 4-5 and 4-6, is the result of this new approach and some algebraic substitution. This third model prompts us to think differently about flight scheduling. Whereas before our paradigm

Decision Variables

z_{bt}	1 if bank b is assigned to be completed at time t; 0 otherwise
y_{ft}	1 if flight f is assigned to arrive at time t; 0 otherwise
w_f	1 if flight f is separated from its bank; 0 otherwise

Figure 4-5: The Third Model's Decision Variables

$$\text{Minimize } \sum_{b=1}^{B}\sum_{t=c_b}^{T} \mathrm{d_{bt}} z_{bt} + \sum_{f=1}^{F}\sum_{t=a_f}^{T} \mathrm{d_{ft}} y_{ft} + \sum_{f=1}^{F} \mathrm{c_f}(1 - \sum_{t=a_f}^{T} y_{ft}) + \sum_{f=1}^{F} \mathrm{s_f} w_f$$

$$\text{rewritten as } \sum_{b=1}^{B}\sum_{t=c_b}^{T} \mathrm{d_{bt}} z_{bt} + \sum_{f=1}^{F}\sum_{t=a_f}^{T} (\mathrm{d_{ft}} - \mathrm{c_f}) y_{ft} + \sum_{f=1}^{F} \mathrm{c_f} + \sum_{f=1}^{F} \mathrm{s_f} w_f$$

subject to

$$\sum_{t=c_b}^{T} z_{bt} = 1 \qquad \forall\, b \qquad (1)$$

$$\sum_{t=a_f}^{T} y_{ft} \leq 1 \qquad \forall\, f \qquad (2)$$

$$z_{bt} + \sum_{j=t+1}^{T} y_{fj} \leq 1 + w_f \qquad \forall\, f \in b, \forall\, t \mid c_b \leq t < T, \forall\, b \qquad (3)$$

$$\sum_{f=1}^{F} y_{ft} \leq \mathrm{m_t} \qquad \forall\, t \qquad (4)$$

$$w_f = 0 \qquad \forall\, f \in \Phi \qquad (5)$$

$$z_{bt}, y_{ft}, w_f \in \{0, 1\} \qquad (6)$$

Figure 4-6: The Third Model

was that "all flights will arrive as scheduled, subject to delay or cancellation due to arrival capacity," now our paradigm is that "all flights will be canceled, but we try to reduce this number by scheduling flights, subject to arrival capacity, to arrive as close to their scheduled arrival times as possible." Note that the problem scenario has not changed at all, but the way the model is formulated has changed our view of the scenario's solution methodology.

The new objective function is the result of the algebraic substitution. It is still the minimization of the sum of the same costs as before, but now the cancellation cost of each flight represents a fixed cost always present; if we schedule a flight, a credit equal to its cancellation cost is received. Note that the fixed cancellation costs could be removed from the objective function, but we keep them in to make the objective function value meaningful.

Constraint 1 still ensures that all banks are eventually completed. Constraint 2 is needed to ensure that the implicit cancellation decision variables $(1 - \sum_t y_{ft})$ remain binary; it essentially replaces $x_f \in \{0, 1\}$. Without this constraint, due to the cancellation cost credit in the objective function, the model will schedule the same flight multiple times to receive multiple credits. Constraint 3 is the result of algebraic substitution. It says that if a flight in a bank is scheduled to arrive after its bank has been completed, that flight must be separated from its bank. Constraints 4, 5 and 6 are the same as before.

Note that Constraints 3 and 4 of the second model have disappeared. Constraint 3, which said that a separated flight must eventually arrive, is now captured in the new Constraint 3 of the third model, which reverses the rule and says that any flight scheduled to arrive outside the bank must be separated. Constraint 4 has been reduced to a tautology, as the algebraic substitution produced $1 = 1$.

4.5 Explaining the Cost Variables

The models in this chapter have four mutually-exclusive input cost variables: bank spread costs, flight cancellation costs, flight delay costs, and flight separation costs. Bank spread costs represent the costs incurred when an inbound bank is late in being completed. When this happens, the 25 to 50 flights in the corresponding outbound bank must now be delayed to wait both for connecting passengers from the inbound bank and possibly for aircraft from the inbound bank assigned to these flights. Bank spread costs can be avoided by separating the delayed inbound flights from their bank. The cost of separation is that connecting passengers on the separated inbound flight will not make their connections, as the outbound bank will now not be delayed for them. The models in this chapter capture the trade-off between the cost of flight separation and of bank spread.

Any flight that is not flown incurs a flight cancellation cost. Any flight that arrives after its scheduled arrival time incurs a flight delay cost. It is important to understand the distinction between flight delay and flight separation costs. Delay costs are independent from separation costs, since passengers on a flight that is delayed but not separated will make their connections to the outbound bank, as the completion time of the inbound bank will be delayed, if necessary to include the flight (recall that if the inbound bank is spread, the start of the outbound bank is delayed accordingly). Any flight that is separated is by

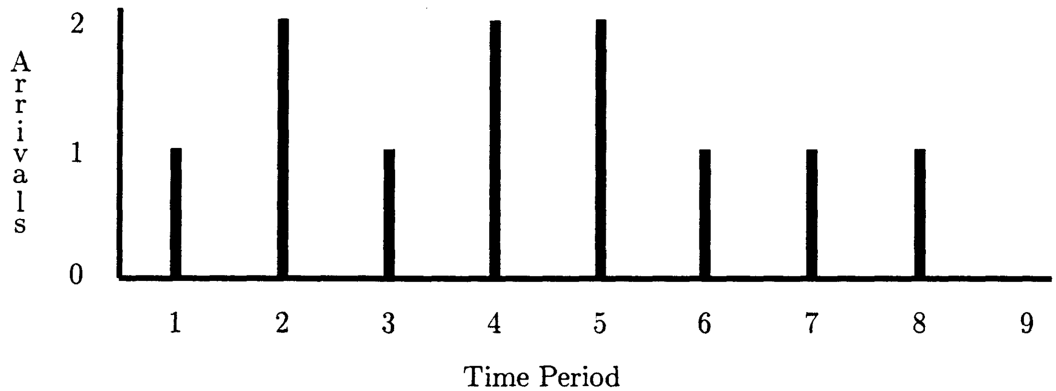

Figure 4-7: Scenario Used For Behavioral Validation

definition also delayed, and will incur each of the two costs.

4.6 Validation Through Behavioral Analysis

Before applying the models of this chapter to real-world scenarios, we wish to satisfy to ourselves that the models behave as expected when applied to some small test cases having obvious solutions. In this section we describe a suite of six test cases, and report that all three models behave as expected when applied to the suite.

The scenario that produced the test cases is described in Figures 4-7 and 4-8. It involves nine time periods, eleven flights and three banks each having three flights. Two of the flights are independent and two cannot be separated from their bank. The test cases are generated by varying arrival capacities and costs; unless otherwise stated, all costs are greater than zero.

CASE 1

Every time period was given ample capacity. As expected, the models returned a schedule having zero costs, all banks completing on time, and all flights arriving on time.

CASE 2

No capacity was given at all. As expected, all three models returned a schedule having all flights canceled and a cost equal to the sum of the cancellation costs of all flights.

CASE 3

Capacity for eleven flights was given to time period nine, and both bank spread and flight delay costs were set to zero. As expected, all three models returned a schedule having no cancellations, zero cost, and all flight arrival times and bank completion times delayed

Arrival Time	Flight Number	Bank Number
1	1	1
2	2	1
2	3	1
3	4	—
4	5*	2
4	6	2
5	7	—
5	8	2
6	9*	3
7	10	3
8	11	3

* inseparable flights

Figure 4-8: Specific Flight Arrival Times

until time period nine.

Case 4

Capacity for eleven flights was given to time period nine, and both flight cancellation and bank spread costs were set to zero. As expected, all three models returned a schedule having zero costs and all flights canceled.

Case 5

Capacity for eleven flights was given to time period nine, with costs set as follows: high bank spread costs, zero flight delay costs, and higher flight cancellation costs than flight separation costs. As expected, all three models returned a schedule having independent flights delayed until time period nine, separable flights separated from their bank and delayed to arrive in time period nine, and inseparable flights canceled.

Case 6

Capacity for eleven flights was given to time period nine, with costs set as follows: high bank spread costs, zero flight delay costs, and higher flight separation costs than flight cancellation costs. As expected, all three models returned a schedule having independent flights delayed until time period nine, and all bank flights canceled.

4.7 Implementing the Models

The three models described in this chapter were solved using CPLEX Version 4.0 on a Sun SPARCstation 20. A C program was used to generate the inputs to CPLEX, to call the CPLEX Mixed Integer Solver, and to print out the results. This section gives some insight into how the models were implemented.

One implementation technique we used that reduces the size of the models is to remove from the solution process those decision variables having a constant value. For example, constraint (1) of the first model is $z_{bT} = 1, \forall\ b$, which says that all banks must eventually be completed. Since the value of this set of decision variables is already known, the z_{bT}'s can be removed from the objective function and any z_{bT}'s in the constraints can be replaced by the value 1. This allows constraint (1) to disappear. (Note that when we report the final objective function value, we account for the z_{bT}'s to keep the value meaningful.) Similarly, all three models have the constraint $w_f = 0, \forall\ f \in \Phi$, which enforces the inability to separate those flights so designated. This constraint is implemented by not defining the appropriate w_f's, which in effect gives them the value 0. The w_f's are also not defined for all independent flights, since separation is not a possible decision for those flights. Finally, recall that the y_{ft}'s are not defined for time periods before flight f's scheduled arrival time, and the z_{bt}'s are not defined for time periods before bank b's scheduled completion time.

But the most important implementation technique is our ability to relax the integrality constraint for a significantly-sized group of decision variables. Specifically, in each of the three implementations we relaxed the y_{ft} decision variables from binary to continuous. (We did discover that it is crucial to performance to use the CPLEX bounding mechanism to bound the continuous variables between 0 and 1.) The y_{ft}'s make up the large majority of the decision variables, and for the full scenario this relaxation applies to over 90% of the decision variables. We found this relaxation to produce a tremendous improvement in the models' solution run-times. Some quantitative examples of the improvement due to this relaxation are shown in the next chapter.

Of course, we are able to make this relaxation only because we are confident that the y_{ft}'s will be either 0 or 1 without being constrained to those values. There are three requirements that ensure that this will likely happen: first, flight delay costs must strictly increase as a function of time (which mirrors reality); second, the arrival slot capacities of each time period must be integer (which also mirrors reality); and finally, the other decision variables must continue to be constrained to binary. When these requirements are satisfied, the models have no incentive to assign fractional values to the y_{ft}'s. However, since we can construct cases in which a solution having fractional values has an equal objective function value as a solution having only integer values, we explicitly check all solutions for fractional results; to date, none have been observed.

Chapter 5

Computational Results Using A Real-World Scenario

5.1 Introduction

In this chapter we use a real-world scenario to both compare the three models to each other and to test their limits. First we describe the real-world scenario and how we constructed our test cases from it. After a short section describing the size of the models when solving the test cases, we move into comparisons and testing and we give the results of four different arrival slot capacity levels. We conclude by giving some quantified benefits of the integrality relaxation implementation technique.

5.2 A Real-World Scenario

We chose US Airways' flight schedule at Pittsburgh International Airport for our real-world scenario. While there are larger hub airports (for example, American at Dallas/Fort Worth and Delta at Atlanta Hartsfield), US Airways' operations at Pittsburgh International involve some of the largest banks in the country [1]. We obtained data for both US Airways' scheduled operations and actual operations for the month of August 1996 (Consolidated Operations and Delay Analysis System (CODAS) data, through the Airline Service Quality Performance (ASQP) program). After examining the data we decided to focus on Wednesday, August 7 and Thursday, August 8. US Airways' schedule is identical on both days, consisting of 306 inbound flights (and 306 outbound flights). The schedule contains two

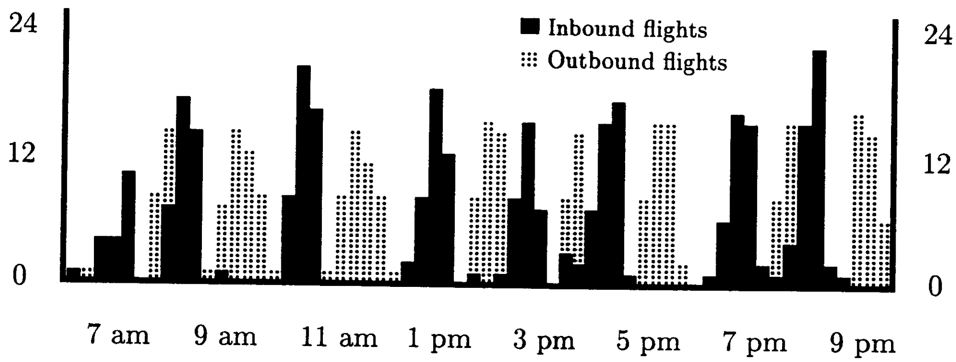

Figure 5-1: Scheduled Operations for August 7 and 8, 1996

Bank	Flights	Start	End
1	18	7:05 am	7:45 am
2	38	8:23 am	9:00 am
3	44	10:37 am	11:15 am
4	40	1:00 pm	1:45 pm
5	30	2:55 pm	3:29 pm
6	40	4:26 pm	5:05 pm
7	40	6:49 pm	7:36 pm
8	41	8:15 pm	8:50 pm

Table 5.1: Inbound Banks

outliers that we did not include in our scenario: the first flight of the day, which arrives at 5:33 am, 62 minutes earlier than the second flight, and the last flight of the day, which arrives at 11:08 pm, 110 minutes later than the penultimate flight. It did not seem worthwhile to extend the timeline from 15 hours to 18 hours simply to include those two flights.

Figure 5-1 shows the number of scheduled arrivals per fifteen-minute time period, with the scheduled departures shown in the background. The definition of the banks is remarkable, and it is obvious that US Airways has eight inbound/outbound banks scheduled. Tables 5.1 and 5.2 show how we broke the schedule into banks; the 304 inbound flights consist of 291 bank flights and 13 independent flights. The nearly-uniform time gap of thirty minutes between the last arrival of the inbound bank and the first departure of the outbound bank was not our doing, the schedule actually works that way.

We chose August 7 and 8 for our scenario because of the significant difference in actual operations on the two days. August 7 was sunny and dry, while August 8 had afternoon

Bank	Flights	Start	End
1	24	8:15 am	8:35 am
2	41	9:30 am	10:15 am
3	41	11:45 am	12:30 pm
4	42	2:15 pm	2:55 pm
5	28	3:55 pm	4:25 pm
6	40	5:30 pm	6:05 pm
7	35	8:00 pm	8:30 pm
8	36	9:20 pm	9:55 pm

Table 5.2: Outbound Banks

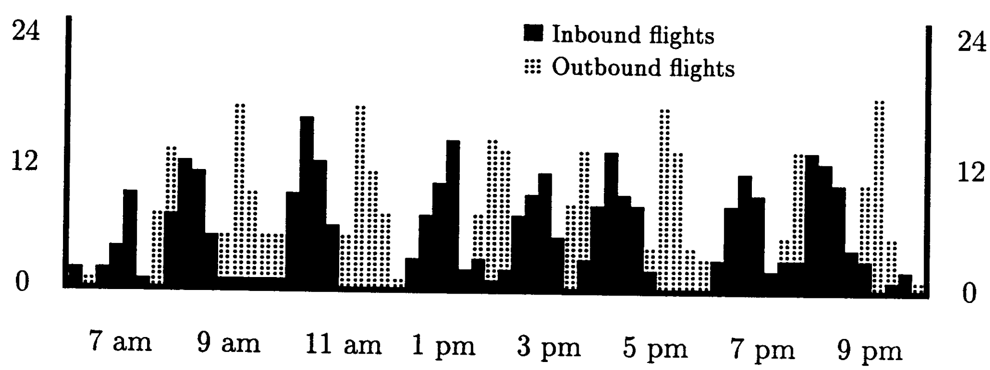

Figure 5-2: Actual Operations for August 7, 1996

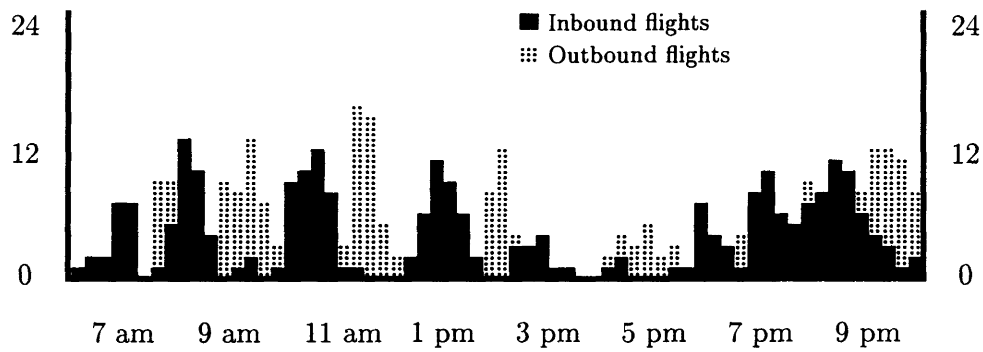

Figure 5-3: Actual Operations for August 8, 1996

thunderstorms and 1.28 inches of rain [10]. Figure 5-2 shows the actual operations for August 7. Only one of the 304 arrival flights was canceled. The banks are still well-defined, although not as precise as in Figure 5-1. This flattening of the banks is inevitable, as US Airways routinely schedules more flights to arrive in a fifteen-minute time period than is physically feasible. For example, on August 7 there was one actual instance of 17 arrivals in a fifteen-minute period, but for the most part 14 to 16 arrivals per fifteen-minute period is the maximum observed. But because the schedule includes instances of 18, 20, and 22 arrival flights per fifteen-minute period, flattening must occur.

Figure 5-3 shows the actual operations for August 8. It is easy to see when the thunderstorms arrived. Fifty-two of the 304 arrival flights were canceled; of the 52 cancellations, four occurred before noon, for reasons presumably not related to the weather.

Our scenario thus consists of eight inbound banks, each having a scheduled completion time, and 304 flights, 13 of which are independent, each having a scheduled arrival time. The scheduled arrivals start at 6:35 am and end at 9:20 pm. To allow for delays, the timeline of our model will consist of 64 fifteen-minute time periods, starting at 6:35 am and ending at 10:35 pm. For each bank, we arbitrarily designate approximately 15% of the flights to be inseparable.

Since we wish to see how the performance of our models is affected by the size of the flight schedule, we construct a series of partial schedules from our real-world scenario. Table 5.3 shows the characteristics of each partial schedule. Note that each partial schedule includes one additional bank and its corresponding flights, and is cumulative of all the schedules before it. Schedule8 is the full scenario.

Name	# of Banks	# of Flights*	# of Time Periods
Schedule1	1	19	10
Schedule2	2	57	15
Schedule3	3	102	24
Schedule4	4	142	34
Schedule5	5	174	41
Schedule6	6	219	48
Schedule7	7	259	58
Schedule8	8	304	64

*includes those independent flights scheduled to arrive before the last bank's completion time

Table 5.3: The Partial Schedules

We still must assign values to the four cost variables. Since the relative values of the cost variables will affect the character of the output (for example, high flight cancellation costs and high bank spread costs will induce an output having many flight separations), we construct three cost structures. Having multiple cost structures allows us to test whether each one has a different best-performing model, or whether a single model performs best regardless of the cost structure.

The first cost structure, labeled Cost1, is our intuitive attempt to mirror reality. Recall that all the model requires is relative, not absolute, cost values, so we arbitrarily choose flight delay costs and go from there. Specifically, we let flight delay costs range from \$8 to \$12 per time period of delay (costs in a range are chosen randomly from a uniform distribution over that range, rounded to the nearest whole dollar). A flight delayed by three hours might as well be canceled, so we set flight cancellation costs to be twelve times flight delay costs, or ranging from \$96 to \$144. The aircraft, crew, and connecting passengers of a separated flight may need to wait two hours for the next outbound bank before continuing, so we set flight separation costs to be eight times flight delay costs, or ranging from \$64 to \$96. Finally, since each time period of inbound bank spread will cause all the flights in the corresponding outbound bank to be delayed by one time period, we set bank spread costs to be the summation of the resulting delay of the outbound flights (the delay costs for each outbound flight range from \$8 to \$12 per time period of delay). Although we have used a linear function for flight delay costs and bank spread costs, our models can handle any general cost function.

Test Case	Model 1	Model 2	Model 3
Schedule1	169 variables (76.9%) 249 constraints 899 entries (2.1%)	170 variables (76.5%) 135 constraints 1019 entries (4.4%)	151 variables (86.1%) 120 constraints 701 entries (3.8%)
Schedule2	603 variables (80.3%) 929 constraints 3466 entries (0.62%)	605 variables (80.0%) 491 constraints 4472 entries (1.5%)	548 variables (88.3%) 444 constraints 3225 entries (1.3%)
Schedule3	1534 variables (85.4%) 2547 constraints 9660 entries (0.25%)	1537 variables (85.2%) 1307 constraints 16119 entries (0.80%)	1435 variables (91.3%) 1223 constraints 12743 entries (0.73%)
Schedule4	2943 variables (88.7%) 5125 constraints 19600 entries (0.13%)	2947 variables (88.6%) 2592 constraints 41422 entries (0.54%)	2805 variables (93.0%) 2474 constraints 34629 entries (0.50%)
Schedule5	4259 variables (90.0%) 7545 constraints 28956 entries (0.09%)	4264 variables (89.9%) 3788 constraints 69661 entries (0.43%)	4090 variables (93.8%) 3644 constraints 59627 entries (0.40%)
Schedule6	5957 variables (90.9%) 10594 constraints 40755 entries (0.06%)	5963 variables (90.8%) 5265 constraints 108958 entries (0.35%)	5744 variables (94.3%) 5087 constraints 94722 entries (0.32%)
Schedule7	8595 variables (92.1%) 15501 constraints 59791 entries (0.04%)	8602 variables (92.0%) 7650 constraints 185579 entries (0.28%)	8343 variables (94.8%) 7438 constraints 164326 entries (0.26%)
Schedule8	10659 variables (92.4%) 19271 constraints 74409 entries (0.04%)	10667 variables (92.3%) 9483 constraints 247083 entries (0.24%)	10363 variables (95.0%) 9236 constraints 220270 entries (0.23%)

'Entries' refers to the number of non-zero entries in the CPLEX input matrix

Table 5.4: Size of the Models

The second cost structure, labeled Cost2, discourages flight cancellations by increasing cancellation costs by a factor of five, to range between \$480 and \$720. The third cost structure, labeled Cost3, encourages flights to remain in their bank by both increasing flight separation costs by a factor of two, to range between \$128 and \$192, and by lowering bank spread costs by a factor of two, by decreasing outbound flight delay costs to range between \$4 and \$6 per time period of delay.

We now have eight test schedules and three cost structures, giving us 24 test cases to apply to different arrival capacity scenarios. Before giving the computational results of applying different capacity scenarios, we discuss how the size of the models varied for the different test schedules.

5.3 Size of the Models

	Model 1			Model 2			Model 3		
	Cost1	Cost2	Cost3	Cost1	Cost2	Cost3	Cost1	Cost2	Cost3
Schedule1	0	0	0	0	0	0	0	0	0
Schedule2	2	2	2	0	0	0	0	0	0
Schedule3	10	14	9	0	0	0	0	1	0
Schedule4	30	53	24	1	1	1	1	2	1
Schedule5	55	114	53	2	2	2	2	4	2
Schedule6	99	n/a	90	5	5	5	3	8	3
Schedule7	187	n/a	177	9	9	9	5	13	5
Schedule8	n/a	n/a	n/a	16	16	16	6	19	6

For entries marked 'n/a', CPLEX
terminated unnaturally

Table 5.5: Results of Test Suite 1, in CPU seconds

Table 5.4 shows the size of the models for the eight test schedules. As expected, due to the replacement of the step-function variables, Model 2 has far fewer constraints and far more non-zero matrix entries than Model 1; the number of decision variables is virtually identical, as is the percentage of those variables (shown next to the number of variables) that are being relaxed from binary to continuous (essentially all the y_{ft}'s). Comparing Model 2 to Model 3, we see that the effect of removing the cancellation decision variables is to decrease the number of decision variables, constraints, and non-zero matrix entries. Furthermore, the percentage of decision variables that are relaxed from binary to continuous has increased.

Finally, Table 5.4 also shows the density of each model instantiation, given after the number of non-zero entries as the percentage of non-zero entries in the CPLEX input matrix (density % = entries/(variables * constraints)). Here the effect of including step-function variables is quite noticeable. For Schedule8, Models 2 and 3 are six times as dense as Model 1.

5.4 Results of Test Suite 1: Ample Arrival Capacity

For our first test suite we allocated 25 arrival slots to each time period. Essentially the models are simply verifying that the schedule can be executed as planned. The cost of the solution in all cases was zero. As shown in Table 5.5, we see that Models 2 and 3 are much faster than Model 1, that for the second cost structure Model 2 slightly outperforms Model 3, and that each model took longer to solve the second cost structure. Recall that the

	Model 1			Model 2			Model 3		
	Cost1	Cost2	Cost3	Cost1	Cost2	Cost3	Cost1	Cost2	Cost3
Schedule1	0	0	0	0	0	0	0	0	0
Schedule2	3	3	2	0	0	0	0	0	0
Schedule3	3600+	3600+	14	3600+	3600+	2	1	2	1
Schedule4	3600+	3600+	43	3600+	3600+	8	2	4	2
Schedule5	3600+	3600+	72	3600+	3600+	15	3	8	3
Schedule6	3600+	3600+	105	3600+	3600+	15	5	12	4
Schedule7	3600+	3600+	n/a	3600+	3600+	59	6	20	7
Schedule8	3600+	3600+	n/a	3600+	3600+	80	8	27	8

For entries marked 'n/a', CPLEX terminated unnaturally

Table 5.6: Results of Test Suite 2, in CPU seconds

second cost structure discourages flight cancellations. Our speculation is that this forces the models to spend more time considering the trade-off between flight separation and bank spread.

5.5 Results of Test Suite 2: Normal Arrival Capacity

For the second test suite we allocated 15 arrival slots to each time period, representing regular operations on a sunny, dry day. Table 5.6 gives the computational results. As expected, solution times are slower than those in the first test suite. Now that the capacity constraints are active, Model 3 is showing its superiority. In fully half the cases, Models 1 and 2 took over an hour to solve (we used CPLEX's time limit function to halt their execution at the one hour point), while Model 3 took 27 seconds at most. It is not clear why Model 2, after outperforming Model 3 in the first test suite for the second cost structure did so poorly on the second cost structure here. Nor is it clear why Models 1 and 2 were able to solve the third cost structure in real-time, but not the first or second. What is clear is the marked effect that different cost structures can have on a model's solution time.

The specific objective function value of each solution does not add any insight, but we do point out that since the three models are equivalent, for a given schedule/cost pair the three models all have the same objective function value. But if the three models are equivalent, how can their solution times vary so dramatically? The answer is that while the models have an equivalent set of integer feasible solutions, their formulations and thus

	Model 1			Model 2			Model 3		
	Cost1	Cost2	Cost3	Cost1	Cost2	Cost3	Cost1	Cost2	Cost3
August 8	3600+	3600+	3600+	3600+	3600+	98	4	124	4

Table 5.7: Results of Test Suite 3, in CPU seconds

	Bank Spread	Flight Delays	Flight Cancellations	Flight Separations
Cost1	0	34	53	10
Cost2	1	51	3	54
Cost3	0	34	62	0

Table 5.8: Solution Characteristics of the Cost Structures

their linear programming relaxations are very different, which affects the performance of the branch-and-bound solution methods used by the integer optimization.

Finally, had US Airways' planned schedule been more realistic and limited planned arrivals to 15 arrivals per time period, we would expect the solution run-times to be similar to those for the first test suite. The lesson learned here, which is no surprise, is that for problem scenarios in which the arrival slot capacity constraints will be active due to bad weather, allowing the planned schedule to also activate the constraints can severely hamper the models' performance.

5.6 Results of Test Suite 3: August 8, 1996

Our third test suite is based on the actual operations of August 8, 1996. Since the model will be executed over the period of irregular operations, we start the timeline at 2:50 pm, the estimated start of the ground hold program, and continue until our typical ending time of 10:35 pm. Since we assume that all the banks and flights scheduled before 2:50 pm have arrived (Figure 5-3 indicates that this was the case), our scenario consists of the flight schedule starting at 2:50 pm. Specifically, it consists of 31 time periods, four banks, and 160 flights. We assume that US Airways' actual arrivals during the ground hold program is representative of the number and distribution of the arrival slots they were allocated, and thus the models' arrival slot constraints are set to mirror the actual arrivals. Specifically, from 2:50 pm until 7:20 pm we allocate a total of 32 arrival slots, an average of less than two slots per time period. Since the arrival slots mirror the actual landings, the distribution

	Model 1			Model 2			Model 3		
	Cost1	Cost2	Cost3	Cost1	Cost2	Cost3	Cost1	Cost2	Cost3
Schedule1	0	0	0	0	0	0	0	0	0
Schedule2	3600+	3600+	6	3600+	3600+	2	2	2	2
Schedule3	3600+	3600+	46	3600+	3600+	26	19	27	11
Schedule4	3600+	3600+	253	3600+	3600+	265	129	294	61
Schedule5	3600+	3600+	641	3600+	3600+	1713	374	1398	136
Schedule6	3600+	3600+	3600+	3600+	3600+	3600+	1315	3600+	414
Schedule7	3600+	3600+	3600+	3600+	3600+	3600+	3600+	3600+	1033
Schedule8	3600+	3600+	3600+	3600+	3600+	3600+	3600+	3600+	2685

Table 5.9: Results of Test Suite 4, in CPU seconds

of the slots can be seen in Figure 5-3 between 2:50 pm and 7:20 pm. We assume that the ground hold program ended at 7:20 pm, and we allocate the normal fifteen arrival slots per time period from that point onward.

The results of this test suite are given in Table 5.7. With one exception, only Model 3 was able to find an optimal solution in less than one hour. For cost structures one and three, it only required four seconds, and for cost structure two, 124 seconds. Model 2 was able to solve the third cost structure in 98 seconds.

We use the solutions of this test suite to examine the different solution characteristics induced by the cost structures. Table 5.8 shows the number of banks spread, flights delayed, flights canceled and flights separated for each solution. Using the first cost structure resulted in 53 of the 160 flights being canceled, whereas US Airways actually canceled 43 of the 160 flights. As expected, the increased cancellation costs of the second cost structure resulted in a large reduction of cancellations, and thus an increase in the number of flights either delayed or separated. The increased separation costs of the third cost structure had a similar expected result. Finally, we note that none of the cost structures induced significant bank spread.

5.7 Results of Test Suite 4: Restricted Arrival Capacity

Our fourth test suite tests the performance limits of all the models by allocating just 5 arrival slots per time period for the entire day. As shown in Table 5.9, beyond Schedule4 even the real-time performance of Model 3 begins to degrade. While Model 3 performed well under the 4 1/2 hours of severely reduced capacity found in the August 8 test suite, we conclude from this example that the real-time performance limit of Model 3 is reached

Test Suite #	Model #	Schedule #	Cost #	Relaxed Time	Unrelaxed Time
1	1	4	1	30	29
1	1	4	2	53	55
1	1	4	3	24	25
1	2	8	1	16	19
1	2	8	2	16	20
1	2	8	3	16	19
1	3	8	1	6	7
1	3	8	2	19	20
1	3	8	3	6	7
2	1	2	1	3	44
2	1	2	2	3	40
2	1	2	3	2	33
2	2	2	1	0	13
2	2	2	2	0	16
2	2	2	3	0	29
2	3	8	1	8	3600+
2	3	8	2	27	3600+
2	3	8	3	8	162
August 8	3	-	1	4	3600+
August 8	3	-	2	124	3600+
August 8	3	-	3	4	3600+

Table 5.10: Effects of Integrality Relaxation; Times are in CPU seconds

when severe reduction of capacity extends to eight hours or more.

5.8 Effects of Integrality Relaxation

Finally, we wish to quantify the benefits of relaxing the y_{ft} decision variables from binary to bounded continuous. Table 5.10 shows the comparison of the relaxed solution times to the unrelaxed solution times for a sample of the test cases described in this chapter. For the first test suite involving ample arrival capacity, we discern no substantial benefit from using the relaxed version. For the second test suite involving normal arrival capacity, we begin to see some differences; in particular, Model 3's solution run-times increase dramatically for Schedule8 under the first and second cost structures. Similarly, for the August 8 test suite, the unrelaxed version of Model 3 is unable to duplicate—or even come close to—the results of the relaxed version. We conclude that relaxing the y_{ft} decision variables is a significant factor in being able to obtain solutions in real-time.

Chapter 6

Conclusion and Future Research

6.1 Conclusion

We presented in this thesis a series of tactical optimization models that an airline can use to assign banks of flights to scarce arrival slots in a partially-decentralized ATFM environment. We used a real-world scenario involving over 300 flights and 16 hours of operations to show that models that account for the bank-induced dependencies of flights into a hub airport can find optimal solutions in real-time. Our first model is an improvement to Milner's Cancellation/Delay Model, our second model is the result of replacing the costly step-function variables, and our third model is the result of removing the unnecessary cancellation decision variables. Even though all three models are equivalent in integer feasible solutions, the third model has the fewest constraints and decision variables and has a structure that results in very quick solution times. Crucial to the performance of all three models is our ability to relax the integrality constraints of over 90% of the decision variables. Additionally, we saw that by varying the relative values assigned to the cost input variables, both the character of and the time needed to obtain the solutions changed significantly.

Since the solutions of the models in this thesis reduce a flight schedule through flight delays, cancellations and separations when constraints are placed on arrival capacity, we classify them as schedule reduction models. Our models can handle any degree of arrival capacity scarcity, which is important since "the successes and failures of ATFM programs are most visible—and most critical to system performance—when the demand-capacity imbalance is most unfavorable" [13]. However, while our models provide local solutions to quickly-developing situations by assigning flights to limited arrival slots, the solutions do not result in an operable, system-balanced schedule of flights. This type of solution is best obtained by using schedule recovery models, which modify an airline's flight schedule and aircraft rotation when unforeseen perturbations to the established schedule arise. Unfortunately, it does not appear that these recovery models can be easily extended to account for either bank-induced dependencies or limitations on arrival slots. Instead, the overall problem can be solved by decomposing it into separate reduction and recovery steps, facilitated by including recovery costs in the cost variables of the schedule reduction models.

Even though a partially-decentralized ATFM environment in which the airlines have greater decision-making power is not in place yet, we believe our models make a contribution today by modeling the airlines behavior in and quantifying the benefits of the new environment. Indeed, [13] says that "modeling airline behavior and quantifying the benefits of alternative ATFM concepts are the most challenging aspects of evaluating decentralized ATFM." Once the environment these models are formulated under is in place, we hope that the models can be developed into operational decision support tools to assist the airlines in the resolution of irregularities.

6.2 Future Research

There are many ways the models presented in this thesis can be improved, for example by increasing the realism they capture and by expanding the scope of the problem scenarios they solve. We describe some of these opportunities for improvement in this section.

6.2.1 Dealing with Uncertainty

Our models make the implicit assumption that the arrival slot allocations made by the FAA are deterministic but we know that this is not true in reality since the allocations depend on uncertain weather forecasts. Dealing with uncertainty in ATFM is the topic of much previous research; as a result, there are many starting points for considering how to incorporate uncertainty into our models. One interesting approach involves the use of scenario analysis [9].

An interesting problem arises when an airline makes an operational decision based on a given allocation and then that allocation is subsequently reduced. The schedule reduction model should be re-optimized given the new allocation, but some of the affected flights might already be airborne. Fortunately, our models can account for this through the cost variables; specifically, we will assign a greater delay cost and a much greater cancellation cost to airborne flights than to flights still on the ground. Any delay the model assigns to an airborne flight will obviously be taken in the air, while a cancellation decision for that flight will mean either returning to the origin or diverting to another destination.

6.2.2 Disaggregating the Banking Operations

While our models assume that each bank flight has a fixed separation cost, in reality the separation cost of a flight is not known until the disposition of all the other flights in the bank is known. For example, consider two through flights, broken into inbound flight A1 and outbound flight A2 and inbound flight B1 and outbound flight B2. When a large number of passengers on flight A1 are connecting to flight B2, if flight B1 is canceled or separated we expect the separation cost of flight A1 to be reduced, since it now does not matter if those passengers connecting from flight A1 to flight B2 arrive in the bank (assuming that no spare aircraft are available to fly flight B2 as scheduled). A similar argument can be made for bank spread costs: as more and more inbound flights involving aircraft assigned to outbound flights are canceled and separated, we expect the bank spread costs to be reduced as there are fewer outbound flights that will be delayed due to the bank spread.

Essentially, our current models approximate these dynamic costs through the use of aggregate bank spread costs and fixed flight separation costs. An opportunity for improvement is to explicitly account for the changing costs and dependencies just mentioned. A preliminary model considering these factors has been formulated, and since some non-linearities in the objective function are introduced, the next step is defining new decision variables to remove the non-linearities.

6.2.3 Estimating the Input Cost Variables

As we saw in this thesis, the input costs drive both the solution characteristics and the solution run-times of the models. Unfortunately, many of these costs are unknown, even to the airlines; however, as we argued in the first chapter, we do not believe that this justifies a departure from a cost minimization objective function. Instead, we propose that greater effort should be made in estimating the costs. Since most irregular operations decision support models make the shaky assumption that these costs are available, such an effort should have great benefit to the aviation operations research community. Indeed, it seems futile to continue to devote a great deal of effort to developing new and improved models without recognizing that such models are of little value without reasonably accurate cost estimates.

Schedule reduction models that only address problem scenarios of limited scope are especially dependent on accurate costs. For example, since the reduction models in this thesis do not explicitly account for some important operating constraints, we would like to implicitly represent them in the input costs. For instance, we would like the cancellation cost of a flight to be a function of the difficulty of recovering the stranded aircraft back into the schedule (among other things) and the delay cost of a flight to be a function of the amount of available downline slack time it its aircraft's rotation (among other things). Thus, an important topic of future research is devising a set of cost-estimation equations that are functions of known data such as passenger totals, individual passenger connections, critical departure times, etc.

6.2.4 Expanding the Problem Scenario

The models of this thesis schedule the arriving flights at a hub airport, a problem scenario of fairly limited scope. While we believe that solving this scenario in real-time is a contribution and that many of the operating factors missing from the problem scenario can be implicitly represented in the cost variables (recall the previous section), extending the models to expand the scope of the problem scenarios considered is a significant opportunity for improvement. For instance, it should be straightforward to expand the models to also schedule the outbound banks at the hub airport, especially if done in conjunction with disaggregating the banking operations. The new models can consider aircraft rotations, crew rotations, passenger connections and the availability of spare aircraft at the hub, as appropriate, and will of course explicitly include the effects of inbound cancellations and delays on outbound operations.

Bibliography

[1] Air Transport World, December 1996, page 1.

[2] Dimitris Bertsimas and Sarah Stock. The air traffic flow management problem with en-route capacities. Working Paper, Operations Research Center, Massachusetts Institute of Technology, 1995.

[3] Jia-Ming Cao and Adib Kanafani. Real-time decision support for integration of airline flight cancellations and delays—part I: Mathematical formulation. Working Paper UCB-ITS-WP-96-1, Institute of Transportation Studies, University of California at Berkeley, April 1996.

[4] Jia-Ming Cao and Adib Kanafani. Real-time decision support for integration of airline flight cancellations and delays—part II: Algorithm and computational experiments. Working Paper UCB-ITS-WP-96-2, Institute of Transportation Studies, University of California at Berkeley, May 1996.

[5] Michael Dudley Delano Clarke. Solving the problem of irregular airline operations. Presented at the INFORMS New Orleans Fall 1995 Meeting, MIT Flight Transportation Laboratory, October 1995.

[6] Michael Dudley Delano Clarke. Development of heuristic procedures for flight rescheduling in the aftermath of irregular airline operations. Presented at the INFORMS Atlanta Fall 1996 Meeting, MIT Flight Transportation Laboratory, November 1996.

[7] Nigel Dennis. Scheduling strategies for airline hub operations. *Journal of Air Transport Management*, 1(3):131–144, 1994.

[8] A. Ghobrial and A. Kanafani. Future of airline hubbed networks: Some policy implications. *Journal of Transportation Engineering*, 121(2):124–134, 1995.

[9] Gregory D. Glockner. Effects of air traffic congestion delays under several flow management policies. Logistics Engineering Center, School of Industrial and Systems Engineering, Georgia Institute of Technology, January 1996.

[10] Greg Goodrich, AccuWeather, Inc.

[11] Robert Hoffman and Michael O. Ball. A comparison of formulations for the single-airport ground holding problem with banking constraints. Working Paper, Institute for Systems Research, University of Maryland, 1997.

[12] Ahmad I. Jarrah, Gang Yu, Nirup Krishnamurthy, and Ananda Rakshit. A decision support framework for airline flight cancellations and delays. *Transportation Science*, 27(3):266–280, 1993.

[13] The Charles Stark Draper Laboratory. Evolution toward a decentralized air traffic flow management system. Final Report for National Aeronautics and Space Administration, Cambridge, Massachusetts, November 1996.

[14] Matthias Mette. Sequential heuristic algorithm for minimum-cost rescheduling of connecting complexes in airline hub operations. Term Paper, Flight Transportation Operations Analysis, Massachusetts Institute of Technology, May 1994.

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

[16] Ananda Rakshit, Nirup Krishnamurthy, and Gang Yu. System operations advisor: A real-time decision support system for managing airline operations at united airlines. *Interfaces*, 26(2):50–58, 1996.

[17] Dusan Teodorovic and Goran Stojkovic. Model for operational daily airline scheduling. *Transportation Planning and Technology*, 14:273–285, 1990.

[18] Dusan Teodorovic and Goran Stojkovic. Model to reduce airline schedule disturbances. *Journal of Transportation Engineering*, 121(4):324–331, 1995.

[19] Alberto Vasquez-Marquez. American Airlines arrival slot allocation system (ASAS). *Interfaces*, 21(1):42–61, 1991.

[20] Michael Wambsganns. Collaborative decision making/dynamic information transfer. Working Paper, Metron Corporation, Washington D.C., 1996.

[21] Shangyao Yan and Dah-Hwei Yang. A decision support framework for handling schedule perturbation. *Transportation Research*, 30B(6):405–419, 1996.