Contents

HANDBOOK OF COMBINATORIAL OPTIMIZATION
D.-Z. Du and P.M. Pardalos (Eds.) pp. – – –

Optimization Applications in the Airline Industry

Gang Yu
Department of Management Science and Information Systems and
Center for Management of Operations and Logistics
The University of Texas at Austin, Austin, TX 78712
E-mail: yu@uts.cc.utexas.edu

Jian Yang
Department of Management Science and Information Systems and
Center for Management of Operations and Logistics
The University of Texas at Austin, Austin, TX 78712
E-mail: jiany@uts.cc.utexas.edu

References

1 INTRODUCTION

The quality of an airline's product is measured by its timeliness, accuracy, functionality, quality, and price. For the air transportation customers, these criteria translate into flexible schedules, on-time flights, safety, satisfactory in-flight services, proper baggage handling, reasonable prices, and convenient ticket purchases. To provide this high-quality, low-cost product, airlines rely on optimization-based decision support systems to generate profitable and cost-effective fare classes, flight schedules, fleet plans, aircraft routes, crew pairings, gate assignments, maintenance schedules, food service plans, training schedules, and baggage handling procedures.

The airline industry has always been an exciting arena for interplay between optimization theory and practice. This attractiveness can be attributed to the nature of airline business, including:

- Severe competition among airlines and between air and other transportation vehicles;
- Large operational scale and scope;
- Tightly coupled resources such as aircraft, crew, maintenance facilities, airports, etc.;
- Active interactions and close dependencies among all involved components;
- Dynamic environment;
- Sophisticated customer behavior;
- Complicated company policies, business rules, and tight control by the Federal Aviation Administration (FAA);
- Complex operational plan, schedules, routes, task assignments, and control mechanisms; and
- Real-time and mission-criticality of decisions.

The challenges above are faced by applied mathematicians, management scientists, operations research analysts, and system engineers who apply their optimization concepts and ideas to the airline industry.

In the past few decades, optimization has made an unprecedented impact on the airline industry. This is largely due to the evolution of computing technology, rapid advancement of optimization methodology, a better grasp of airlines' business rules by operations researchers and practitioners, and the driving demand from the airline industry caused by competition and customers' high expectations.

In this paper, we survey the major optimization areas that have had a great impact on the airline industry. The following topics are covered in separate sections: network design, yield management, flight planning and fleet assignment, crew scheduling, air traffic flow control, and irregular operations control. In the conclusion, we comment on some of the missing topics and point out future research and development directions based on our knowledge and experience.

2 NETWORK DESIGNS

The network design problem in the airline industry can be stated as follows: find the optimal network structure and optimal routes to carry the targeted passenger flow at the lowest total transportation cost.

Among all the factors, including air fares, quality of service, yield management and route structures, the routing system has been proven to be a critical element impacting an airline's market share, and thus its ability to compete. As a result, the optimal network design problem together with an efficient routing structure are critical issues facing airlines today.

Prior to airline deregulation in 1978, the routes of U.S. airlines were controlled by the Civil Aeronautics Board which was established by the Civil Aeronautics Act of 1938. In order to add new routes, airlines were required to prove to the Civil Aeronautics Board that the proposed new services would benefit the public and that competing airlines already serving the routes would not be adversely affected. Therefore, developing long main routes was a primary objective, resulting in a "linear" pattern for the airline network structure [14]. Figure 1 illustrates such a "linear" pattern network structure which was commonly employed by all the airlines before deregulation.

Figure 1: A point-to-point network structure

By eliminating entry restrictions, deregulation gave the airlines increased freedom and flexibility in restructuring their networks. It allowed the airlines, for the first time in forty years, to establish routes and fares freely. Tumultuous changes ensued in the industry with profound impact on the most basic aspects of airline operations, including fares, services, quality, and safety. In fact, eighteen months after deregulation, 106,000 city-pair authorizations had been issued, in marked contrast to the 24,000 authorizations granted during the eighteen-month period immediately preceding deregulation.

Perhaps one of the most significant developments in the airline industry was the hub-spoke network structure. Hub-spoke was conceived to protect and increase an airline's market share. In the competitive environment encouraged by deregulation, the development of a hub-spoke routing system was used as a cost reduction approach, but also as an essential marketing tool. To date, all major U.S. carriers except Southwest Airlines have established hub-spoke network systems, including Dallas-based American Airlines, Chicago-based United Airlines, Houston-based Continental Airlines, St. Paul-based Northwest Airlines, and Atlanta-based Delta Airlines.

Figure 2: A typical hub-spoke network system.

Figure 2 is an illustration of a typical hub-spoke network system commonly seen after deregulation. A centrally-located airport serves as an airline's hub. The airline offers flights between its hub and the airports on the periphery. The solid lines in Figure 2 represent these routes. Flights from various origins (spokes) arrive at a hub as an intermediate point from which passengers change planes to proceed to their ultimate destinations. This strategy targets passengers traveling between origins and destinations for which the traffic volume is not sufficient to establish frequent non-stop flights. However, by consolidating passengers with different origins and destinations, the hubbing airline is expected to be able to serve more passengers on its flights, using larger and more efficient aircraft. Take American Airlines' hub-spoke network system as an example, at its main hub in Dallas/Fort Worth, passengers from one flight can connect to any of 30 or more other flights.

Essentially, centralization and the broader scope of operations associated with a hub-spoke system permit the airlines to take advantage of economy of scale. Consider American Airlines as an example. In 1980, approximately 10 percent of its traffic consisted of connecting passengers. By the mid-1980s, about 66 percent of the passengers on a typical flight to a hub airport were connecting to other flights to varying destinations [130].

In addition to becoming popular in airline operation, the concept of hub-spoke networks has also been applied to air cargo delivery, ground transportation, satellite communication, telephone networks, and other logistical systems. For example, WalMart has been successfully using the concept in retailing since the business grew from a small regional retailer to a national chain with more than 890 stores. Today, about 80 percent of its merchandise is delivered directly to the individual retail outlets from eight distribution centers in six hub cities [92].

Despite the economic significance and popularity of hub-spoke systems, there has been little analysis of the effectiveness of this type of network. Insufficient systematic mathematical analysis has been conducted for the purpose of justifying and evaluating the effectiveness of hub-spoke systems currently in use. Hence, the questions of whether or not hub-spoke is a better network design remains to be further investigated and correctly answered.

Further, very limited systematic mathematical studies has been performed in order to provide an optimal network structure. Most studies in the past only focused on the construction of the hub-spoke system in a restrictive way. Solution procedures were presented under the assumptions that (1) there must exist a certain number of hubs in the underlying network and (2) each node must be connected to a hub. Obviously, with a restriction regarding the number of hubs that must exist and a restriction of the fixed structure of the hub-spoke framework, this type of analysis only provided a sub-optimal solution.

Past research on hub-spoke systems were conducted from two different aspects. On the one hand, many researchers analyzed and assessed the advantages of a hub-spoke system in terms of airline economics. On the other hand, some focused on mathematical models for identifying optimal hub locations and designing routing policies within a given hub-spoke network structure.

To assess the effect of hub-spoke systems on an airline's cost structure, McShan and Windle [97] measured the change in the extent of hub-spoke routing since deregulation. Their comments on the hub characteristics were: (i) Hubbing is likely to result in more frequent flights and should therefore improve service; (ii) the central location is clearly preferable, since the hub minimizes total distance traveled on each spoke; and (iii) the hub city size is also important in order to minimize total distance traveled by all passengers. Overall, they suggested that hubs should be located centrally, thereby also attaining substantial local traffic.

Brown [26] developed an economic model to examine the effectiveness of the hub-spoke system. His study showed that the welfare effects of airlines' hub-and-spoke system are ambiguous. On one hand, hubbing route patterns represent an improvement over the linear route structures that existed before deregulation, since they permit a more efficient use of aircraft and lower the fixed cost on a per-passenger basis. On the other hand, however, he concluded that hub-and-spoke structures result in the transfer of consumer surplus to airlines.

In their analysis of airline competitive behavior, Bailey, Graham and Kaplan [14] studied airline cost and profitability using a simultaneous regression model. They concluded that the hub-spoke system allows airlines to have more frequent flights with larger aircraft and a higher percentage of seat occupancy, and is, therefore, cost-effective. Morrison and Winston [102] looked at the effects of hub-spoke systems on passenger welfare, finding that, on average, passengers benefited from the switch to hub-spoke networks by receiving more frequent flights with lower fares and slightly shorter travel times. Reynolds-Feighan [115] studied the efficiency of airline network routing systems and found that the overall returns to scale for hub routes and higher than that of nonhub routes. He concluded that the increasing returns to scale associated with each of the hub subsystems encourages airlines to expand the number of connections to hub airports and discourages adding 'spoke' routes which are not directly connected to one of the hubs.

Aside from the research on the economic effects of the hub-spoke systems, there are few studies which focused on analyzing a hub-spoke system in order to better understand it.

Ghobrial and Kanafani [56], Hansen [61], and Hansen and Kanafani [62][63] are among the researchers who primarily modeled the airline hub-spoke system as an approach in which equilibration rather than optimization is stressed. It is noted that hubbing causes unevenness in the distribution of both the benefits and the costs associated with air transportation, while at the same time, it closely couples that distribution with the competitive fortunes of individual airlines. As summarized by Hansen and Kanafani [62], the net result is increased uncertainty and rivalry among airports, as well as the communities they serve.

In terms of constructing a hub-spoke system, some studies were conducted solely for the purpose of finding the optimal hub locations. These studies either analyzed the factors affecting the hub location decisions or used mathematical modeling to derive the optimal hub locations.

In an effort to better understand the hubbing phenomena, Bauer [17] looked for the main factors that airlines consider in evaluating existing and potential hubs and investigated the impact of the hubbing decision on airport traffic. His results indicated that population is the most important factor determining hub location among all the characteristics that influence hub location and the effect on airport traffic as a result of hub activity. One of his most interesting findings was that the creation of a hub at a city leads to a more-than-doubling of revenue generated by passenger emplanements at that city.

O'Kelly [107] developed several hub-location models for a hub-spoke system. O'Kelly concluded that if the cost of setting up each intercity route is ignored, then there is no rational reason to construct a hub-oriented system. If, on the other hand, there is a cost associated with each intercity route, as long as the incremental transportation costs are less than the savings in link costs in order for a single hub to emerge, the one-hub system is preferred to the no-hub system.

Another assumption used in O'Kelly's model for the two-hub system, which was implicitly reflected in the model as pointed out by Aykin [10], is that the destinations in O'Kelly's model are all in a Euclidean space. The model and the Euclidean distance are appropriate if all destinations cover a small area on the earth's surface. When destinations are widely separated, the Euclidean space is no longer a suitable approximation due to the curvature of the earth's surface. Aykin further generalized O'Kelly's model into a formulation that can be used to solve multi-hub location problems in the

Euclidean space and gave the condition under which a destination is optimally assigned to a particular hub facility.

O'Kelly [108] subsequently formulated a more generalized hub location problem in a discrete solution space with a quadratic integer programming model. With the goal of minimizing the total transportation cost, the model was formulated to yield the optimal hub locations.

$$\min \sum_{i=1} \sum_{j=1} W_{ij} (\sum_{k} X_{ik} C_{ik} + \sum_{m} X_{jm} C_{jm} + a \sum_{k} \sum_{m} X_{ik} X_{jm} C_{jm})$$

subject to

$$\begin{aligned}
(n-p+1)X_{jj} - \sum_{i} X_{ij} &\geq 0 \qquad && j = 1, ..., n \\
\sum_{j} X_{ij} &= 1 && i = 1, ..., n \\
\sum_{j} X_{jj} &= p && \\
X_{ij} &\in \{0, 1\} &&
\end{aligned}$$

where binary variable X_{ij} indicates if node i is linked to a hub at j, and $X_{ii} = 1$ if node i is a hub. W_{ij} is the demand flow between any two cities, and C_{ij} is the transportation cost of an unit flow between node i and j. The total number of cities to be interconnected is n, and p is the total number of hubs to be constructed. O'Kelly's model is strictly limited to a hub-spoke network structure. On the assumption that there is no direct connection between any two spoke cities, the possible alternative route structure is completely ignored in the model.

To introduce a more comprehensive framework with networking policies, Aykin [11] developed several integer programming models considering different networking policies. Each of his models deals with a special case in which a specified networking policy is applied. Basically, he considered the following two networking policies: (i) nonstrict hubbing, where flows between any pair of nodes are not required to go through a hub. Even there exist hubs, flows can go through a hub only when it is cost effective; (ii) strict but nonrestrictive hubbing, where all flows are channeled through hubs. The "nonrestrictive" hubbing implies that flows from the same spoke node are not required to go through the same hub, but different hub if desirable;

and (iii) strict and restrictive hubbing, where not only flows are required to go through hubs, but also flows from the same spoke node have to be served by the same hub. Aykin concluded that nonstrict hubbing is more cost-effective than strict but nonrestrictive hubbing policy. And among the three, the strict and restrictive hubbing is the least desirable.

Jaillet et al. [70] [71] addressed the network design problem from the perspective of a given airline, assuming that an airline serves a fixed share of the market. To be more specific, they assumed that the demand would be given, and dependent on the resulting design of the network. This is an idealized situation. Demand is only captive in a monopoly situation (e.g., Olympic Airways exclusively serving the Greek Islands). Within a given set of service policies that an airline may provide, i.e., non-stop flights versus hub-connecting flights, their network design problem can be stated as follows:

> Given a fixed origin-destination flow demand matrix, and the capacities and the mileage cost of different types of aircraft, design a network which satisfies the demand and minimizes the total transportation cost.

In the Jaillet et al. model, the following notations and definitions were used. Models are all prefixed with NWD, standing for network design problems.

Policy Classifications:

1) One-Stop: The airline provides two possible services for each route it serves: a non-stop flight and a flight with one connection. *NWD(1)* is used to denote this basic model from which several more complex models are developed. The index (1) implies that only up to one stop flights are allowed.

2) Two-Stop: Similar to the one-connection case, the airline now provides an additional two-connection flight. This is the most common type of service in the U.S. airline industry. With an extra stop is permitted, it is expected that the solution in terms of total operational cost to this model should be at least as good as that of model *NWD(1).*

3) All-Stop: The airline is assumed to be in a monopoly situation. It serves the entire market exclusively without any competition. Therefore, if it would make the airline more profitable, flows on any single route could be channeled through as many stops as there are cities involved in that market. Although this policy is not practical to airline industry, it has various important applications in other fields such as telecommunication networks, air cargo delivery, and other logistical systems.

This model will be denoted as *NWD(n-2)*. The index $(n-2)$ implies that flights with up to $(n-2)$ stops are allowed, where n denotes the number of cities involved. There is a close relationship between *NWD(n-2)* and *NWD(1)*. Solutions to *NWD(n-2)* can be obtained by solving a transformed model of *NWD(1)*.

Parameters:

K = number of aircraft types (e.g. Boeing 747s or DC 10s)
n = number of cities the airline serves
f_{ij} = number of passengers who desire to fly from city i to city j per day
d_{ij} = air distance from city i to city j
c_k = cost per mile for aircraft type k
b_k = capacity of aircraft type k

Decision variables:

x_{ij} = fraction of the flow f_{ij} served by a direct flight from i to j
x_{ilj} = fraction of the flow f_{ij} served by an indirect flight from i to an intermediate city l and then to j
x_{iltj} = fraction of the flow f_{ij} served by an indirect flight from i to j by going through two intermediate cities at l and t
y_{ij}^k = number of aircraft of type k used on the route from city i to city j

Model *NWD(1)*

The basic model focuses on minimizing the total in-flight transportation cost. It ignores the fixed cost for aircraft purchasing/leasing, it is believed that this model will reveal a basic hub-spoke pattern if it is cost-effective. The solution of this model will not only answer the question of which network structure an airline should adopt, but also yield an optimal routing strategy for the airline to apply.

$$NWD(1) \qquad \min \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{K} d_{ij} c_k y_{ij}^k$$

subject to

$$\begin{aligned}
f_{ij}x_{ij} + \sum_{l=1}^{n} f_{lj}x_{lij} + \sum_{l=1}^{n} f_{il}x_{ijl} &\leq \sum_{k=1}^{K} b_k y_{ij}^k \qquad && i,j = 1,...,n \\
x_{ij} + \sum_{l=1}^{n} x_{ilj} &= 1 && i,j = 1,...,n \\
x_{ij}, x_{ilj} &\geq 0 && i,j,l = 1,...,n \\
y_{ij}^k &\in Z_+ && i,j = 1,...,n; k = 1,...,K
\end{aligned}$$

The objective function minimizes the total transportation cost with respect to the number of aircraft used and the distances to be traveled by the aircraft. First set of constraints restricts the number of flow units carried on each arc so that the capacity of a selected set of aircraft on that arc will not be exceeded. Second constraint set ensures that demand between any given pair of cities is satisfied. Together with the integrality constraints, the model depicts the airline's operations under the one-stop policy. Note that the following is not taken into account in this simplified model:

1. Fixed cost for purchasing/leasing aircraft.

2. Limit on total number of available aircraft. Omission of this limit is not an unreasonable approximation, since it is assumed the possibility of leasing aircraft at any city in any amount, and the objective function also gives an incentive to reduce the number of aircraft.

3. Periodic airline operations. The airline operations should be scheduled on a regular basis. After a certain length of time, crews must return to their bases.

Despite the omission of these considerations, this model captures the essence of the intended goal in the following sense: (i) if a hub-spoke system is desirable, this model will reveal the pattern; and (ii) by revealing a basic network structure, it provides directions for both routing passengers and selecting aircraft for each route.

Model *NWD(n-2)*

Define S_{ij} as a set of paths from city i to city j, and if a path $P \in S_{ij}$, then $|P|$ is defined as the number of cities traversed by the path. Also, the authors assume that if $P \in S_{ij}$, then $|P_{ij}| \geq 3$, so that S_{ij} does not include the one-arc path x_{ij}. Mathematically, any path $P \in S_{ij}$ can be represented by a sequence $v_1, v_2, ..., v_{|P|}$ with $v_1 = i, v_{|P|} = j$. Expression $(ij) \in P$ where $P \in S_{al}$ is used to indicate that the arc (i, j) is in the path P from a to l, or that i and j are two consecutive cities in the path. The model *NWD(1)* is then extended to describe the all-stop policy:

$$NWD(n-2) \qquad \min \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{K} d_{ij} c_k y_{ij}^k$$

subject to

$$f_{ij} x_{ij} + \sum_{a,l} \sum_{\forall P \in S_{al}:(ij)\in P} f_{al} x_P \leq \sum_{k=1}^{K} b_k y_{ij}^k \qquad i, j = 1, ..., n$$

$$x_{ij} + \sum_{\forall P \in S_{ij}} x_P = 1 \qquad i, j = 1, ..., n$$

$$x_{ij}, x_P \geq 0 \qquad i, j = 1, ..., n$$

$$y_{ij}^k \in Z_+ \qquad i, j = 1, ..., n; k = 1, ..., K.$$

Since a path may involve up to n cities, a large number of real variables is expected in the model. In fact, if all possible paths are considered, the number of real variables X will be exponential with respect to the number of cities n. In turn, this would make the model intractable.

By using a transformation described in the following, Jaillet et al. showed the relationship between model *NWD(n-2)* and model *NWD(1)*.

Transformation of NWD(n-2): Given k types of aircraft and the demand matrix f_{ij}, denote by k^* the most economic aircraft (i.e., aircraft k^* has the lowest cost per seat c_{k^*}/b_{k^*}), by m a large integer number such that $mb_{k^*} \geq$

$\sum_{ij} f_{ij}$. Define a transformation of the demand matrix by $\bar{f}_{ij} = f_{ij} + mb_{k^*}$, then by adding an additional constraint:

$$y_{ij}^{k^*} \geq m \qquad i,j = 1,...,n.$$

a transformed model of *NWD(n-2)* is obtained:

$$TNWD(n-2) \qquad \min \sum_{i=1}^{n}\sum_{j=1}^{n}(\sum_{k=1}^{K} d_{ij}c_k y_{ij}^k - d_{ij}c_{k^*}m)$$

subject to

$$\begin{aligned} f_{ij}x_{ij} + \sum_{P_{al}:(ij)\in P_{al}} f_{al}x_{P_{al}} &\leq \sum_{k=1}^{K} b_k y_{ij}^k && i,j = 1,...,n \\ x_{ij} + \sum_{P_{ij}} x_{P_{ij}} &= 1 && i,j = 1,...,n \\ y_{ij}^{k^*} &\geq m && i,j = 1,...,n \\ x_{ij}, x_{P_{ij}} &\geq 0 && i,j,l = 1,...,n \\ y_{ij}^k &\in Z_+ && i,j = 1,...,n; k = 1,...,K \end{aligned}$$

Lemma 1 *Let the solution to* $TNWD(n-2)$ *be* $\bar{x}_{P_{ij}}, \bar{y}_{ij}^k$*, then*

$$\hat{y}_{ij}^k = \begin{cases} \bar{y}_{ij}^k & \text{if } k \neq k^* \\ \bar{y}_{ij}^{k^*} - m & \text{if } k = k^* \end{cases}$$

is an optimal solution to $NWD(n-2)$*.*

This transformation is valid for the all-stop case only, since Lemma 1 may not hold under any restricted cases; for any arc the demand mb_{k^*} is not guaranteed to be carried directly. Because the key requirement for a valid transformation is that added demand be directly carried between each pair of nodes, solution equivalence of the transformation for either one-stop or two-stop cases cannot be established. However, by applying the transformation to the model *NWD(1)*, we can show that the solution space of the transformed model *TNWD(1)* is defined the same as that of *TNWD(n-2)*. Hence, the model is reduced significantly in size, and an optimal solution to *NWD(n-2)* can then be obtained through solving a much simplified model, *TNWD(1)*. Applying the transformation to the model *NWD(1)*, we have the following transformed model *TNWD(1)*:

$$TNWD(1) \qquad \min \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{K} d_{ij}c_k y_{ij}^k$$

subject to

$$\begin{array}{rcll}
\bar{f}_{ij}x_{ij} + \sum_{l=1}^{n}\bar{f}_{lj}x_{lij} + \sum_{l=1}^{n}\bar{f}_{il}x_{ijl} & \leq & \sum_{k=1}^{K} b_k y_{ij}^k & i,j=1,...,n \\
x_{ij} + \sum_{l=1}^{n} x_{ilj} & = & 1 & i,j=1,...,n \\
y_{ij}^{k^*} & \geq & m & i,j=1,...,n \\
x_{ij}, x_{ilj} & \geq & 0 & i,j,l=1,...,n \\
y_{ij}^k & \in & Z_+ & i,j=1,...,n; k=1,...,K
\end{array}$$

and the following theorem for establishing the relationship between the transformed model *TNWD(1)* and *NWD(n-2)*:

Theorem 2.1 *Let an optimal solution to* $TNWD(1)$ *be* $\bar{x}_{ij}, \bar{x}_{ijl}, \bar{y}_{ij}^k$, *then*

$$\hat{y}_{ij}^k = \begin{cases} \bar{y}_{ij}^k & \text{if } k \neq k^* \\ \bar{y}_{ij}^{k^*} - m & \text{if } k = k^* \end{cases}$$

is the optimal solution to $NWD(n-2)$.

Based on Theorem 2.1, the following procedure can be used for solving the model *NWD(n-2)*:

Step 1: Transform *NWD(1)* into *TNWD(1)*;

Step 2: Solve the transformed model *TNWD(1)* to get solution $\hat{y}_{ij}^k$;

Step 3: Use the longest path in the graph determined by nonzero $\hat{y}_{ij}^k$ to limit the possible path length; and

Step 4: Solve *NWD(n-2)* as a linear program with y_{ij}^k fixed at $\hat{y}_{ij}^k$ and the path length limited by what has been found in Step 2.

Model *NWD(2)*

In practice, almost all the airlines in the U.S. have adopted this type of policy. The rationale is obvious. Using as many stops as needed would certainly be more profitable if the market were fixed. However, in a realistic situation, a flight with more connections would certainly make air travel less desirable to passengers. An airline not operating in a monopoly situation would lose its market share, and therefore this is not acceptable.

On the other hand, the model with at most two intermediate stops will both be adequate to describe realistic situations and reasonable enough for the fixed demand assumption to be valid.

$$NWD(2) \qquad \min \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{K} d_{ij} c_k y_{ij}^k$$

subject to

$$\begin{aligned}
& f_{ij}x_{ij} + \sum_{l=1}^{n}\sum_{t=1}^{n} f_{lj}x_{ltij} + \sum_{l=1}^{n}\sum_{t=1}^{n} f_{il}x_{ijtl} \\
+\sum_{l=1}^{n}\sum_{t=1}^{n} f_{lt}x_{lijt} \quad &\leq \quad \sum_{k=1}^{K} b_k y_{ij}^k \qquad i,j = 1,...,n \\
x_{ij} + \sum_{l=1}^{n}\sum_{t=1}^{n} x_{iltj} \quad &= \quad 1 \qquad i,j = 1,...,n \\
x_{ij}, x_{iltj} \quad &\geq \quad 0 \qquad i,j,l,t = 1,...,n \\
y_{ij}^k \quad &\in \quad Z_+ \qquad i,j = 1,...,n; \quad k = 1,...,K.
\end{aligned}$$

Jaillet et al. developed various heuristics to solve $NWD(1), NWD(2)$ and $NWD(n-2)$. The heuristics all contain an initial feasible solution construction phase based on LP relaxation and then followed by various solution improvement procedures. Some benchmark testing problems have been formed. One problem contains 39 cities selected from top 100 U.S. cities. These cities are chosen in such a way that all major geographical areas of U.S. are covered. The distance between any two cities is measured as a direct line in miles. Intercity passenger travel demand is estimated based on some gravity model. Those models in general are constructed using two sets of variables, namely socioeconomic variables and supply variables. Among the socioeconomic variables, choices are the following:

- Population. The population of the total metropolitan area served by an airport;
- Employment. The total employment in a metropolitan area is a measure of the level of economic activities that generate travel or attract travel; and
- Disposable income. This variable is usually measured on a per capita basis, and it is used as a measure of potential of travel.

The choice of supply variables are commonly used in city-pair models are:

- Airfare;
- Travel time;
- Distance;
- Frequency of service; and
- Other level of service attributes.

The detailed model and methods for setting parameters can be found in Song [131]. The following list the major criteria defined for analysis of the network structure based on the solutions:

- Degree of the nodes (cities) in the solution graph, i.e., the number of arcs connecting to a city. The higher degree of a city in the resulting network, the more likely the city is a hub;
- Number of aircraft flying in and out of a city. Since the degree of a node does not account for the volume of flow, this criterion is used to account for this factor. The larger the number of aircraft goes in-and-out of a city, the more likely the city serves as a hub;
- Difference between the number of aircraft going in a city and the minimum number of aircraft that is required for satisfying the demand of that city. The larger the difference is, the more likely the corresponding city is a hub;

- Number of passengers to whom a city is served as an intermediate stop, i.e., the total number of passengers going through a city for connecting flights to destinations other than this city. The larger this number is, the higher chance for the city to be a hub;

- Proportion of total passengers traveling directly to destinations without going through intermediate stops. The larger this percentage is for a given city, the more likely that the city serves as a hub. This is because for any spoke city, a large proportion of passengers originating from it are likely to go through some hub cities in order to reach their destinations, while for a hub city this proportion is less.

Conclusions drawn from Jaillet et al. study can be summarized as follows. As a combination of hub-spoke and alternative arrangement, a cost effective network design appears to be more hub-spoke structured. Location of potential hubs seem to be more geographically dependent than they are on the density of the demand levels. In addition, hub positions can be located differently depending on the policy adopted. With a relative high level of demand flows, the difference among the policies is insignificant. Thus, the one-stop policy is recommended to account for social factors.

3 YIELD MANAGEMENT

Airline deregulation has remarkably raised the level of competition in the air transportation market. In order to maintain and improve market share, airlines make tremendous effort in yield management (it is also referred to as revenue management in recent literature). In order to survive in such a competitive environment and as result of yield management research, many airlines offer a wide variety of fares, ranging from deeply discounted fares to higher priced coach, business, and first class fares. The fare levels offered for a flight are directly affected by the pressure to match competitors' fares in the same market. Since little room is left for improving yield management in terms of better pricing due to low profit margins, balancing the number of discount and full-fare reservations accepted for a flight so as to maximize total passenger revenue became the focus of airlines' yield managers [19]. The need for a balanced solution comes from the fact that lower fares attract more passengers, thus create greater load factors, while they also take away seats which could have been sold at higher fares to increase revenue. The payoff for effectively managing the seat inventory is substantial. Delta

Airlines estimated that selling one seat per flight at full fare rather than at a discounted fare would add over $50 million to its annual revenue [85].

Research on reservations and booking control dates far back to before deregulation. Etschmaier and Rothstein [46] and Gasco [52] gave complete surveys. Beckmann [18] and Thompson [148] studied the problem in a very simplified manner. Taylor's work [140] can be considered as a pioneering research in booking level control. Based on Taylor's work, a whole family of models were developed for controlling overbooking levels. Several airlines have implemented Taylor's ideas in their booking systems. Rothstein and Stone [123] described one of the implementations by American Airlines. Rothstein [122] also gave a survey of the application of operations research to airline overbooking.

Most airlines and researchers deal with the seat allocation problem flight by flight. Even for a single flight leg, the problem is very complex. On the same flight, there are passengers with various origin-destination (O-D) itineraries each of which generates a different amount of revenue. For a major airline practicing hub-and-spoke operations, every flight to the hub can have passengers destined to almost all of its spoke stations; every flight from the hub can have passengers departing from almost all of its spoke stations. In addition, every itinerary has several different fare levels. So, there can be hundreds of fare class/itinerary combinations for each flight leg, each having its own desirability to the airline. The essential factor in determining the seat allotment is passenger demand. Passenger demand is not deterministic, but its trend is reflected in past records and in the number of reserved seats for the current flight. To build and solve a model optimizing seat utilization which covers the decisions of all the combinations, fully utilizes historical passenger demand, and dynamically adjusts its decision with the evolving reservation data is out of the question. All the models which deal with this problem make certain simplifying assumptions.

Belobaba [19] discusses a very simple static model to decide the seat allocation for a flight with two fare classes. Passenger demand for each fare class is assumed to be an independent random variable. f_i is defined as the revenue generated per passenger in fare class i and $\bar{b}_i(S_i)$ the expected number of passengers that will make reservations in fare class i when S_i reservations have been made in class i. The constraint is due to the preexisting cabin capacity: $C = S_1 + S_2$. The ideal allocation that maximizes the total expected revenue

$$\bar{R} = \bar{R_1}(S_1) + \bar{R_2}(C - S_1) = f_1\bar{b_1}(S_1) + f_2\bar{b_2}(C - S_1)$$

is the solution of equation

$$\partial\bar{R}/\partial S_1 = \partial\bar{R}/\partial S_2.$$

Hence, we have the optimal allocation S_1^*, S_2^* satisfying

$$f_1\bar{P}_1(S_1^*) = f_2\bar{P}_2(S_2^*),$$

because we have

$$\bar{b_i}(S_i) = \int_0^{S_i} x\bar{p_i}(x)dx = \int_0^{S_i} \bar{P_i}(x)dx,$$

where $\bar{p_i}(x)$ is the probability density at the ticket selling level of x seats in fare class i, and $\bar{P_i}(x)$ is the probability of selling x or more seats in fare class i. Therefore, seats are optimally allocated between the fare classes such that the marginal expected total revenue with respect to additional seats in each class is equal to zero. In this model, the $\bar{P_i}(S_i)$ data are generated from historical records.

Equating marginal revenues in each of two fare classes to find the optimal seat allocation for a flight can be extended to dynamic models. Littlewood [84] suggests that low-fare passengers paying f_2 should be accepted as long as:

$$f_2 \geq \bar{P_1}(S_1)f_1,$$

where $\bar{P_1}(S_1)$ is the probability of selling all remaining S_1 seats to high-fare passengers paying f_1 per ticket. The implicit assumptions made in the model are: low-fare passengers book first, there are no cancellations of bookings, and a rejected request is regarded as revenue lost to the airline. Wang [154] uses the same idea to dynamically allocate seats for different fare class-itinerary combinations in a flight. If the current number of bookings for combination (j, k) is b_{jk} for every feasible combination (j, k), then the next seat will be allocated to the combination (j, k) with the largest $f_{jk}P[r_{jk} > b_{jk}]$, where $P[r_{jk} > b_{jk}]$ is the probability that another request for (j, k) will be received given b_{jk} bookings are accepted for combination (j, k).

Demands for both discount-fare class and full-fare class on a flight are closely related to the overall passenger demand for the flight, and passengers who are not able to get the discount fare often upgrade their bookings to the full-fare class. A more accurate treatment of the allocation problem should consider the correlation between the demands for the two fare classes. Brumelle et al. [27] propose a static model to determine the discount booking limit η. It uses the distribution of demand for full-fare tickets conditioned on the demand for discount-fare tickets. Furthermore, consideration of the loss in revenue reflecting the loss of goodwill of passengers who are turned away from full-fare class is incorporated into a second model. A third model even considers the situation where some passengers are upgraded from discount-fare class to full-fare class. All the models are built on the following framework.

Suppose B is the demand for discount fare class; $Y(\eta)$ is the demand for full fare class given discount booking limit η; κ is the cabin capacity; and ρ_B, ρ_Y, and ρ_G stand for the revenue generated by carrying each discount fare passenger, each full fare passenger, and goodwill-related loss for each passenger being denied full fare class booking, respectively. Then the total revenue as a random variable $R(\eta)$ is:

$$R(\eta) = \rho_B(B \wedge \eta) + \rho_Y(Y(\eta) \wedge (\kappa - (\eta \wedge B))) - \rho_G(Y(\eta) - (\kappa - (\eta \wedge B)))^+,$$

where $a \wedge b = \min(a, b)$, $a \vee b = \max(a, b)$, and $a^+ = a \vee 0$. To determine whether the ηth request for the discount fare class should be accepted at the time $\eta - 1$ bookings for the class have been made, the expected incremental gain $G(\eta) = E[R(\eta) - R(\eta - 1) \mid B \geq \eta]$ shall be assessed as:

$$G(\eta) = \rho_B + \rho_Y E[(Y(\eta) \wedge (\kappa - \eta) - Y(\eta - 1) \wedge (\kappa - \eta + 1) \mid B \geq \eta]$$

$$-\rho_G E[Y(\eta) - (\kappa - \eta))^+ - (Y(\eta - 1) - (\kappa - \eta + 1))^+ \mid B \geq \eta].$$

If $G(\eta)$ is nonnegative for all η up to some η^*, and nonpositive thereafter, then η^* will be optimal. For all the three models the η^* can be found.

The first model neglects the goodwill factor, so its optimal discount booking limit is:

$$\eta^* = \max\{0 \leq \eta \leq \kappa : Pr[Y > \kappa - \eta \mid B \geq \eta] < \frac{\rho_B}{\rho_Y}\},$$

and $\eta^* = 0$ if $Pr[Y > \kappa] \geq \frac{\rho_B}{\rho_Y}$. Probability $Pr[Y > \kappa - \eta \mid B \geq \eta]$ can be interpreted as the maximal ratio of passengers being turned away in the full-fare booking process. When the limit η is reached in practice, this probability is just the ratio. So, closeness of the spill rate to the ratio of unit revenues $\frac{\rho_B}{\rho_Y}$ somehow evaluates the closeness of the η in use to its optimal value.

For the second model, the optimal value is:

$$\eta^* = \max\{0 \leq \eta \leq \kappa : Pr[Y > \kappa - \eta \mid B \geq \eta] < \frac{\rho_B}{\rho_Y + \rho_G}\}.$$

Because the probability is a monotonously-increasing function of η, the incorporation of the goodwill consideration decreases the optimal limit.

When considering the third model, every customer seeking a discount-fare after its booking limit η has been reached independently has a probability of γ of seeking a booking in the full-fare class. That is, $Y(\eta) = Y + U(\eta)$, where $U(\eta) = \sum_{i=\eta+1}^{B} D_i$, with D_i's being independent and identically-distributed Bernoullie random variables having $ED_i = \gamma$. The optimal discount-fare booking limit is:

$$\eta^* = \max\{0 \leq \eta \leq \kappa : Pr[Y + U(\eta) > \kappa - \eta \mid B \geq \eta] < \frac{\rho_B - \gamma\rho_Y}{(1-\gamma)\rho_Y}\}.$$

If ρ_Y is replaced by $\rho_Y + \rho_G$, then the consideration of upgrading is incorporated into the model. All these models provide tighter limits than what would have been provided by Littlewood's model, thus produce lower full-fare passenger spill rates.

Glover et al. [57] present a time-space-network-based seat-allocation model to find the mix of passenger itineraries flowing over the airline's network in independent fare classes that maximizes total revenues. Nodes in the network are associated with flight segments' departure time-station pairs and arrival time-station pairs. One set of arcs in the time-advancing direction represents the flight legs and has the flight legs' capacities as the arcs' flow capacities, and another set of arcs in the time-reversing direction represents passenger itinerary-fare class combinations (PIs) and has their demands as the arcs' flow capacities. The last set of arcs links nodes at the same stations in the time-advancing direction that represents passengers transferring flights. With the fare for each PI being designated as a negative cost, the problem of globally allocating each flight segment's seats to various PIs is solved as a minimum cost network flow problem with side constraints. The model accommodates up to 600 daily flights, 30,000 passenger itineraries, and five fare classes. The model uses each PI's deterministic demands as the upper limit for its level of allocation. This is a good approximation if demand far exceeds supply.

Curry [35] combined the marginal seat revenue and mathematical programming approaches to find the optimal seat allocations when fare classes are nested on an origin-destination itinerary, and the seats are not shared among different origin-destinations. Classes with different fares and the same origin-destination are allocated to nests, with higher fare classes' nests containing lower fares. A fare class can take the seats allocated to classes with lower fares. The mathematical model maximizes the expected total revenues generated from all the nests, subject to the constraints imposed by the limited capacity of each flight leg. Using the marginal seat allocation approach, the expected revenue from each nest is found to be a convex function of the number of seats allocated to the nest. The problem is then approximated as a piecewise linear programming problem and solved by using linear programming techniques.

Mayer [96] addressed the two-fare nested case where both classes are allowed to be filled at the same time. Titze and Griesshaber [149] simulate the nested case when low-fare passengers book before high-fare passengers book. Belobaba [20] presented the Expected Marginal Seat Revenue (EMSR) algorithm for finding an approximately optimal policy for seat allocations when demand for each fare class is normally distributed. Wollmer [158] gave a simpler and faster algorithm for finding the optimal seat allocation policy that maximizes mean revenue by establishing a critical value for each fare class. Booking requests for a particular fare class are accepted if and only if the number of empty seats is strictly greater than its critical value.

In practice, passengers are allowed to cancel their reservations booked in advance or not to show up at all without penalty. Even when a flight is solidly booked, it is likely to have vacant seats at take-off time. To offset the revenue loss incurred by this practice, airlines have adopted the overbooking policy, which allows the numbers of bookings to exceed the available seats on a flight. The level of allowable overbooking must be determined by the airlines. Too tight a level often produces a low occupancy rate of the airlines' perishable seats, while too loose a level decreases passenger goodwill of those who are denied seats they have booked and creates expense for the airline as it tries to compensate those passengers. The difficulty of determining a good policy of overbooking is due to the randomness and unpredictably of passenger cancellations and no-show rates.

Shlifer and Vardi [129] let the upper bounds of the number of bookings $N^*(t)$ allowed at various stages of the booking process be the decision variables. They used the fact that the ratio of show-ups at take-off time versus the number of bookings at any given time is independent of the elapsed time since the booking was made. They made a justifiable assumption that the number of show-ups at take-off time, when $N(t)$ is the number of bookings at time t, is a normal random variable with expectation $N(t)a(t)$ and variance $N(t)b(t)$, where $a(t)$ and $b(t)$ are experimentally-observed parameters. Three criteria were used to determine $N^*(t)$ at any time t, with it being the largest number that simultaneously satisfies all the criteria. The three criteria were:

- η^* = maximum allowable probability of show-ups exceeding the capacity of the plane M.
- θ^* = maximum allowable ratio of expected rejections over expected show-ups.
- ξ^* = ratio of the loss C_2 incurred by rejecting a passenger and the profit C_1 incurred by carrying one. Given this ratio, the model looks for the policy that maximizes the expected revenue of a flight.

In the case of a single-itinerary flight carrying a single type of passengers, the probability of show-ups in excess of capacity $\eta(N)$, given N reservations booked at time t, is expressed as:

$$\eta(N) = \phi(\frac{Na - M}{\sqrt{Nb}}),$$

where $\phi(\cdot)$ is the standard normal distribution. We can find N^{η^*} by solving $\eta(N) = \eta^*$. The expected number of rejected passengers $D(N)$, given N reservations booked at time t, is written as:

$$D(N) = (Na - M + \sqrt{Nb})\phi(\frac{Na - M}{\sqrt{Nb}}).$$

With $\theta(N) = D(N)/Na$, N^{θ^*} is the solution of $\theta(N) = \theta^*$. The expected gain of a flight $\bar{R}(N) = C_1(Na - D(N)) - C_2 D(N)$. Denoting $R(N) = R(\bar{N})/C_1$, we have:

$$R(N) = Na - (1 + \xi^*)D(N).$$

In a certain range, both $D'(N)$ and $D''(N)$ are larger than zero. Thus $R(N)$ has a single maximum with the optimal solution N^{ξ^*}. The optimal policy is defined as

$$N^* = \min\{N^{\eta^*}, N^{\theta^*}, N^{\xi^*}\}.$$

For other more complicated cases involving more than one itinerary and passenger type, the approaches go along the same line. The implementation of the decision rules substantially reduces the instances when passengers were affected by overbooking and at the same time maintains a high utilization of aircraft capacity.

Rothstein [121] applies dynamic programming to solving a Markovian sequential decision process for an airline's overbooking policy when there is only one fare class. The policy's decision variables $k_n(t)$'s denote the number of additional reservations allowed to be made from $t = 1$ (the period just before departure) to $t = T$ (the period starting T units before departure). At each t, the decision variable depends on the previous booking level n, where n ranges from 0 to n_t (the largest booking level prior to t which has a positive probability). The optimal policy is the one that maximizes the expected total revenue, which takes into account the penalty incurred for denying passengers from boarding.

The following parameters, which are all constructed from experimental data, are used in the model:

$\psi_n(T)$ = the probability that n passengers are booked at the beginning of period T, $n = 0, ..., n_T$.

$d_i(t)$ = the probability that there are i reservation demands during period t, $t \geq 1$, for $i = 0, .., i_t$, with i_t being the maximum number of demands with positive probability in period t.

$c_{j|n}(t)$ = the probability of j cancellations during period t out of n passengers already booked at the beginning of this period. No-shows are included among the cancellations in period 1.

$u_h(t)$ = the probability that h unconfirmed reservations will be recorded during period t, and in the period of each t, for $h = 0, ..., h_t$, where h_t is the maximum value of h with positive probability.

Rothstein defined $V_n(t)$ to be the maximum expected gain achievable by any booking policy, given that n passengers are already booked at the beginning of period t for $t \geq 1$, and $V_n(0)$ is the maximum gain when n passengers with recorded reservations, plus any no-records and standbys, arrive for boarding. $V_n(0)$ is fairly accurately estimated, given the airplane's cabin size, distribution of the number of standby passengers, revenue gain c per passenger being carried, and revenue loss b per passenger being denied boarding. $V_n(t)$ for $t \geq 1$ and the optimal policy are reached by using the following recursive relations over time from $t = 0$ to $t = T$, and at each t, from $n = 0$ to $n = n_t$:

$$V_n(t) = \max_k \{ \sum_{i=0}^{k} d_i(t) \sum_{h=0}^{h_t} \sum_{j=0}^{n} u_h(t) c_{j|n}(t) V_{n+i+h-j}(t-1)$$

$$+ \sum_{i=k+1}^{\infty} d_i(t) \sum_{h=0}^{h_t} \sum_{j=0}^{n} u_h(t) c_{j|n}(t) V_{n+k+h-j}(t-1) \}.$$

The unconditional expected revenue gained from the optimal policy is:

$$E(V) = \sum_{n=0}^{n_T} \psi_n(T) V_n(T).$$

Rothstein also dealt with the situation where a maximum expected denied boarding ratio is enforced. The ratio $R_n(0)$ at departure time, given that there are n recorded reservations, can be evaluated. If $R_n(t)$ is defined as the expected denied boarding ratio when there are n reservations at the beginning of period t, for $t \geq 1$, then there is a recursive relation for these $R_n(t)$'s based on conditional probability, given that the optimal policy $(k_n(t), n = 0, ..., n_t, t = 1, ..., T)$ is at hand:

$$R_n(t) = \sum_{i=0}^{k_n(t)} d_i(t) \sum_{h=0}^{h_t} \sum_{j=0}^{n} c_{j|n}(t) u_h(t) R_{n+i+h-j}(t-1)$$

$$+(\sum_{i=k_n(t)+1}^{\infty} d_i(t)) \sum_{h=0}^{h_t} \sum_{j=0}^{n} c_{j|n}(t) u_h(t) R_{n+h+k_n(t)-j}(t-1).$$

When the optimal policy is used, the unconditional expected denied boarding ratio is:

$$E(R) = \sum_{n=0}^{n_T} \psi_n(T) R_n(T).$$

Rothstein argued that the unit denied boarding penalties b and $E(R)$ move in opposite directions. He suggestd that b be tuned to get an optimal policy such that $E(R)$ is less than and as close to the enforced maximum ratio r_0 as possible. In so doing, he treatd b as some intermediate value to reach his optimal policy, while b has its true meaning in reality. The result achieved by manipulating b is rather doubtful.

Alstrup et al. [2] extended Rothstein's model to the case of two-fare classes. They employed a two-variable stochastic dynamic programming model developed by Ladany et al. [65][78][79] to decide the optimal overbooking policy for flights with two types of passengers. The policy is set to be the number of reservations allowed to be made for two classes (C and M) in the time period t, $UC(t)$ and $UM(t)$. The state variables are the booking levels at the beginning of various time periods for the two classes, $BC(t)$ and $BM(t)$. The goal is to find the optimal values for UC and UM, so that the expected total cost when the booking levels at the beginning of period T are given as BC and BM; namely, $V(BC, BM, T)$ is minimized.

Alstrup et al. defined $PBC(i,t)$ as the probability of i reservation demands during period t (t units before departure) for class C passengers, $PBM(j,t)$ as the probability of j reservation demands during period t for class M passengers, $PCC(BC, k, t)$ as the probability of k cancellations during period t for class C passengers when the booking level is BC at the beginning of the period, and $PCM(BM, g, t)$ as the probability of g cancellations during period t for class M passengers when the booking level is BM at the beginning of the period. The recursive relation for $V(BC, BM, t)$ is:

$$V(BC, BM, t) =$$

$$\min_{UC,UM} \{ \sum_{i=0}^{UC} \sum_{j=0}^{UM} \sum_{k=0}^{BC} \sum_{g=0}^{BM} PBC(i,t) PBM(j,t) PCC(BC,k,t)$$

$$PCM(BM,g,t)V(BC+i-k,BM+j-g,t-1)$$

$$+(\sum_{i=UC+1}^{\infty} PBC(i,t))\sum_{j=0}^{UM}\sum_{k=0}^{BC}\sum_{g=0}^{BM} PBM(j,t)PCC(BC,k,t)$$

$$PCM(BM,g,t)V(BC+UC-k,BM+j-g,t-1)$$

$$+(\sum_{j=UM+1}^{\infty} PBM(j,t))\sum_{i=0}^{UC}\sum_{k=0}^{BC}\sum_{g=0}^{BM} PBC(i,t)PCC(BC,k,t)$$

$$PCM(BM,g,t)V(BC+i-k,BM+UM-g,t-1)$$

$$+(\sum_{i=UC+1}^{\infty} PBC(i,t))(\sum_{j=UM+1}^{\infty} PBM(j,t))\sum_{k=0}^{BC}\sum_{g=0}^{BM} PCC(BC,k,t)$$

$$PCM(BM,g,t)V(BC+UC-k,BM+UM-g,t-1)\}.$$

To calculate $V(BC,BM,T)$ for every integer combination of (BC,BM) in the rectangle area determined by the lower-left point $(0,0)$ and the upper-right point $(C_{upperbound}, M_{upperbound})$, calculations have to proceed from $t=0$ to $t=T$, by using the recursive relation. For each t, $V(BC,BM,t)$ for every (BC,BM) in the rectangle has to be computed and the optimal (UC,UM) associated with this (BC,BM,t) triplet needs to be stored as that instance's solution. The authors aggregated passengers into batches of size 5, and used the rule-of-thumb that optimal $(UC,UM)=(0,0)$ for (BC',BM') leads to optimal $(UC,UM)=(0,0)$ for $BC>BC'$ and $BM>BM'$, to reduce computation time. For $t=0$, $V(BC,BM,0)$'s estimation takes into account the probability of passenger no-shows:

$$V(BC,BM,0)=\sum_{i=0}^{BC}\sum_{j=0}^{BM} PNC(i,BC)PNM(j,BM)cost(BC-j,BM-j),$$

where $PNC(i,BC)$ is the probability for i no-shows out of BC booked class C passengers and $PNM(j,BM)$ is the probability for j no-shows out of BM booked class M passengers, $cost(BC^*,BM^*)$ is the cost incurred on the airline when the real number of show-up passengers for the two classes are BC^* and BM^*, respectively. This value includes the impact of turning down excessive passengers and downgrading classes for passengers. All the possibility functions, PBC, PBM, PCC, PCM, PNC, and PNM, are estimated from the real data provided by Scandinavian Airlines Systems (SAS). Simulations and comparisons with other simpler models indicate that the decision tables obtained from the model define an efficient booking policy.

None of the models mentioned above lets passenger demand depend explicitly on fares. Weatherford [156] proposed a model in which fares are decision variables along with seat allocation upper limits, while passenger demands are random variables affected by fares. Furthermore, his model takes into account the effect of cross-elasticity, i.e., the demand for a fare class depends not only on the price of this class, but also on prices of other competing fare classes. Weatherford assumed that passenger demand for each fare class is a normal distribution whose mean is a linear function of its own and neighboring classes' prices. The objective of his model is to maximize the expected total revenue. By making certain assumptions conceiving the situations when passenger demand exceeds booking upper limits, the model is solved by nonlinear optimization tools.

To be effective, the above models all rely on accurate passenger demand and no-show forecasts. Typical forecasting methods are simple time series models like moving averages and exponentially smoothed averages, causal modeling that seeks quantitative relationships between the dependent variables being predicted and any cross sectional factors that might have impact on them, and time-series cross-sectional modeling like Kalman filters. Sun et al. [136] introduced the adaptive neural networks modeling approach. The neural network consists of three layers of nodes: one input layer, one hidden layer, and one output layer. The input information for dynamic forecasting such as day of the week, origin-destination pair, and current booking level is transformed into values on the input nodes. The final outputs are uniquely determined from values on output nodes which are transformed from values on input nodes through the three layers. The processes that transform the input values to the output values are simple weighted averages. The weights are stored on the arcs between different layers. The arcs' linkage and the weights are subject to change so as to lessen the difference between the output and the actual outcome during the adaptive training process. The approach compares favorably to other traditional approaches in a comparative study. BANKET, the forecasting module that employs this approach, has been implemented on a daily basis for ten airlines since 1989.

4 FLIGHT PLANNING AND FLEET ASSIGNMENT

Flight planning is a critical stage of in airline's planning process. When a flight schedule is given, a major proportion of costs and revenues are fixed. All the subsequent planning stages have to optimize the use of resources in the space restricted by the schedule. Therefore, optimization of the flight schedule is of great importance to an airline. In most airlines, the flight schedule is drafted several months before it is put into execution. When the first draft comes out, it is studied by the various departments involved in the work of fleet assignment, crew scheduling, aircraft maintenance, and other resource allocation processes. After the draft's feasibility and economics are evaluated and changes are recommended, it is sent back to the flight planning department for revision. Normally, a flight schedule goes through an iterative procedure of this kind many times before reaching its final, ready-to-execute form.

The factors that must be considered to efficiently and effectively draft a flight schedule include [45]: the demand function and associated revenues for each origin-destination market over the time-of-the-day and the day-of-the-week of the planning cycle; features of the routes such as distances, operational restrictions, and aircraft characteristics such as capacities, speeds, fuel costs, crew assignments, etc.; and other operational constraints. To capture all the details of an airline's operations and produce an ultra-optimal solution for the flight schedule amounts to an intractably-complicated task. All the currently-existing planning models only take into consideration some simplified functions of passenger demand, aircraft operation costs, route characteristics, and other restrictions into consideration. Also, the flight schedule essentially provides a friendly framework on which other downstream planning processes are based. Most planners plan the frequency of service on each route first, then determine the departure times on the basis of the time-of-the-day variability of demand and connectivity of flights.

Since an airline's profit is monotonously dependent on its market share, to maximize its market share with limited aircraft capacity is its primary goal. Simpson [128] and Teodorovic [141] showed that market share on routes with a large number of competitive carriers is determined largely by flight frequency. The other factor an airline has to consider during its flight schedule planning is passengers' satisfaction levels, because of its effect on the airline's long-term profitability. There are various measures of this level, while only the timing of departures and frequency of flights are related to the flight schedule. Making some simplifying assumptions of timings, it can also be transformed into a frequency issue. Therefore, deciding the frequency of flights in an airline's serviced network is a key decision-making process crucial to its short-term and long-term success.

Teodorovic and Krcmar-Nozic [143] proposed a multi-criteria model that incorporates the major considerations in determining a good flight schedule in a competitive environment. The model's first objective is to maximize a designated airline p's total net profit. The profit on each route i is the revenue minus the cost. Revenue is proportional to the total number of passengers captured, and cost is proportional to the frequency of flights on i. According to Powell [113], the number of passengers captured by airline p on route i is:

$$V_{ip} = \mu_i N_{ip}^{\alpha} / \sum_{j=1}^{m} N_{ij}^{\alpha},$$

where μ_i is the average number of passengers to take route i, α is an empirical parameter between 1 and 2, m is the number of competitive carriers including p, and N_{ij} is carrier j's frequency of flights on route i. Therefore, the total net profit that p can make from operating on all the k routes with flight frequency of N_{ip} for route i when others fly the same route with frequencies $N_{ij}, j \neq p$ is:

$$P_p = \sum_{i=1}^{k} [(c_i \mu_i N_{ip}^{\alpha} / \sum_{j=1}^{m} N_{ij}^{\alpha}) - C_i N_{ip}],$$

where c_i is the average ticket price on route i, and C_i is the total cost per flight on route i.

The model's second and third objectives are related to the passengers' satisfaction levels. The second objective is to minimize the average scheduled delay of passengers due to the discrepancies between scheduled departure times and passengers' desired departure times. Based on some stochastic assumptions, the model takes $sd_{ip} = T/(4N_{ip})$ as the average scheduled delay per passenger when served by airline p on route i, where T is the time interval during which passengers express a desire to fly. When taking into account the number of passengers on each route taken by p, the total scheduled delay $Sd_p(=\sum_{i=1}^{k} V_{ip} sd_{ip})$ is:

$$Sd_p = \sum_{i=1}^{k} [\mu_i N_{ip}^{\alpha-1} T / (4 \sum_{j=1}^{m} N_{ij}^{\alpha})].$$

The third objective is to maximize the number of passengers captured by airline p. This is for the far-flung purpose of minimizing the number of passengers being turned away due to lack of seats. Carrier p's total expected number of passengers is:

$$T_p = \sum_{i=1}^{m} (\frac{\mu_i N_{ip}^{\alpha}}{\sum_{j=1}^{m} N_{ij}^{\alpha}}).$$

In building the model, the constraint that must be considered is the airline's limited capacity; that is, it has limited resources required to produce a certain number of flying seat-hours. This constraint can be expressed as:

$$\sum_{i=1}^{k} N_{ip} n t_{B_i} \leq S_p,$$

where n is the number of seats in the plane used for executing the flights, and t_{B_i} is the time needed to fly route i. On a certain set A of routes, some flight frequencies (the maximum allowed) cannot be exceeded. So, there is a group of constraints that reads like this:

$$N_{ip} \leq N_{ip}^{*}, \qquad \forall i \in A,$$

where N_{ip}^{*} is the maximum allowed flight frequency on route i. On some other set B of routes, net profits are required to be nonnegative. Using the previously described expression for the net profits, we have yet another group of constraints:

$$N_{ip}^{\alpha-1} / \sum_{j=1}^{m} N_{ij}^{\alpha} \geq C_i / (c_i \mu_i), \qquad \forall i \in B.$$

The complete model is shown below:

$$\max P_p = \sum_{i=1}^{k}[\frac{c_i \mu_i N_{ip}^{\alpha}}{\sum_{j=1}^{m} N^{\alpha} ij} - C_i N_{ip}]$$

$$\max T_p = \sum_{i=1}^{k}(\frac{\mu_i N_{ip}^{\alpha}}{\sum_{j=1}^{m} N_{ij}^{\alpha}})$$

$$\min Sd_p = \sum_{i=1}^{k}[\frac{\mu_i T N_{ip}^{\alpha-1}}{4\sum_{j=1}^{m} N_{ij}^{\alpha}}]$$

subject to

$$N_{ip} \leq N_{ip}^{*} \qquad \forall i \in A$$

$$\frac{N_{ip}^{\alpha-1}}{\sum_{j=1}^{m} N_{ij}^{\alpha}} \geq \frac{C_i}{c_i \mu_i} \qquad \forall i \in B$$

$$\sum_{i=1}^{k} N_{ip} n t_{B_i} \leq S_p$$

$$N_{ip}\ (i = 1, 2, ..., k)\ integer.$$

This is a nonlinear integer multi-criteria model. The competitors' flight frequencies are assumed to be known parameters in the model, and hence are excluded from the decision variables.

To obtain a solution which optimizes all of the objectives simultaneously is impossible in the general case. For this kind of multi-criteria problem, only a Pareto-optimal solution, the solution where no criterion can be improved without simultaneously worsening at least one of the remaining criteria, can be achieved. More precisely, $x^* \in X$ is a Pareto-optimal solution if there is no other $x \in X$ such that:

$$f_i(x) \geq f_i(x^*) \qquad \forall i = 1, 2, ..., r$$

with at least one strict inequality being satisfied. In the expression, $f_1(x), ..., f_r(x)$ are the objective functions that are to be maximized. An objective function to be minimized can simply be converted to one that is to be maximized. In the paper, Teodorovic and Krcmar-Nozic tried to find a satisfactory solution $x \in X$ which holds the inequalities:

$$f_i(x) \geq \bar{f}_i \qquad \forall i = 1, 2, ..., r,$$

where $\bar{f} = (\bar{f}_1, ..., \bar{f}_r)$ is the aspiration level given by the decision maker. Their interactive solution method was based on the work of Nakayama and Sawaragi [103]. The brief layout of the algorithm is as follows:

Step 1. Assign an ideal point $f^* = (f_1^*, ..., f_r^*)$ where f_i^* is sufficiently large, for example $f_i^* = \max_{x \in X} f_i(x)$.

Step 2. In the kth iteration the decision maker is asked to give the aspiration level $\bar{f}_i^k < f_i^*$ for every criterion $f_i, i = 1, 2, ..., r$.
Set $W_i^k = 1/(f_i^* - \bar{f}_i^k), i = 1, 2, ..., r$, and solve the problem:

$$\min_x \max_i W_i^k \mid f_i^* - f_i(x) \mid \qquad\qquad i = 1, 2, ..., r$$

and get the solution x^k.

Note that the closer the aspiration level for i is to its ideal value, the bigger its weight W_i^k is.

Step 3. Based on the values of $f_i(x^k), i = 1, 2, ..., r$, the decision maker sets the aspiration level for the next iteration or quits the procedure with x^k being the final solution. If the decision maker wants to continue, then his options in setting $\bar{f}_i^{k+1}$'s are:

$\bar{f}_i^{k+1} > f_i(x^k)$, if the decision maker wants to improve the ith criterion,
$\bar{f}_i^{k+1} < f_i(x^k)$, if the decision maker is ready to worsen the ith criterion, or
$\bar{f}_i^{k+1} = f_i(x^k)$, if the decision maker accepts the ith criterion as it is.
To continue the procedure, put $k = k + 1$, and go to Step 2.

The subproblems to get the ideal point and the min-max problems in Step 2 are hard-to-solve nonlinear integer programming problems. The authors applied the Monte Carlo integer programming method [32] to get approximate solutions for them. The idea is to generate many feasible solutions and choose the best-performing one to be the output solution. The numerical experiments conducted by the authors on forty routes with two competing airlines achieved satisfactory solutions.

Levin [83] proposed several models to help determine the selection of flights. The variability of the schedule comes from the selection of flights from a larger set of potential flights. For each intended flight, a bundle of flights with departure times in a time neighborhood is considered. Conceivably the models are for the cases when the flight frequency for each origin-destination market is already determined. The models' objective is to minimize the use of aircraft to make all the connections to fulfill the intended flights.

One of Levin's models follows the following line of thinking. When treating each flight as a node i and considering an aircraft serving an earlier flight i being able to serve a later flight j as a directed arc (i, j), the problem becomes that of finding the least number of chains in the thus constructed network G, such that exactly one node in each bundle is included in one of the chains. Furthermore, the number of nodes required to be covered is a constant $|N|$, the number of intended flights, while every chain has arcs one less than its nodes. When network G is mapped to a bipartite graph $G^* = \{S, T, A^*\}$, such that each node i corresponds to a node $s_i \in S$ and another node $t_i \in T$ and each arc (i, j) corresponds to an arc (s_i, t_j), the problem of minimizing the number of aircraft being used can be formulated as maximizing the total 0-1 flow in graph G^* such that no more than one node in $S(T)$ in each bundle has outgoing (incoming) flow and each node sends out (receives) flow to (from) at most one other node. The bundle without in-going or outgoing flows can be thought of as having one aircraft committed to it, and there is no preference for any one of the alternate flights.

If we define the following sets and variables:

$$K_l = \{k \mid \ s_k \ \text{ belongs to the } l\text{th bundle}\},$$

$$A(i) = \{j \mid (s_i, t_j) \in A^*\},$$

$$B(j) = \{i \mid (s_i, t_j) \in A^*\},$$

$$x_{ij} = \text{the flow from } s_i \text{ to } t_j,$$

$$u_i = 0, 1 \quad \text{the indicator of whether the } i\text{th flight is selected},$$

then the model's mathematical formulation can be listed as below:

$$\max z = \sum_{l=1}^{|N|} \sum_{i \in K_l} \sum_{j \in A(i)} x_{ij}$$

subject to

$$\sum_{i \in K_l} u_i = 1 \qquad \forall l$$

$$u_i \geq \sum_{j \in A(i)} x_{ij} \qquad \forall i$$

$$u_j \geq \sum_{i \in B(j)} x_{ij} \qquad \forall j$$

$$u_i, x_{ij} = 0, 1 \qquad \forall i, j.$$

Levin suggested using the Land-and-Doig type [80] branch-and-bound algorithm to solve the model optimally. In the solution tree, if the LP solution of the terminal node with the maximum LP objective value among all the terminal nodes is integral, then the optimal solution found is claimed to be just this solution; otherwise, one u_i which is between 0 and 1 in the LP solution is fixed to 0 and 1, respectively, to cover this current terminal node with two new terminal nodes. After these new nodes' LP solutions are found, the terminal node with the largest LP objective is searched again, and the whole process repeats like this until the optimal solution is found. Because the algorithm keeps track of the problem's upper bound, the integral LP solution is actually the global optimal solution for the problem.

Another model Levin gave was built directly on the physical layout of the set of possible flights being considered. It was based on a time-space network. Each departure time and station pair constitutes a node of the network, so does each arrival time and station pair. Each potential flight links two nodes which correspond to its departure and arrival. All the nodes at the same station are linked in the time-advancing direction by the so called ground arcs, and the latest node σ_m at a station is linked back to the earliest one λ_m by an overnight arc to complete a cycle. A flow of 1 in one flight arc signals the selection of this flight and a flow of 0 otherwise, and a conserved flow is kept in the network. The problem of minimizing the use of aircraft is treated as minimizing the total amount of flow in all of the overnight arcs such that exactly one arc in one bundle has flow 1 while others have zero flow, and such that the flow is conserved. The flow conservation comes naturally due to the requirement that the number of aircraft is balanced. If this time, the set K_l is considered as the set of flight arcs in the network that corresponds to potential flights in bundle l, the layout of the model is as follows:

$$\min z = \sum_{(\sigma_m,\lambda_m)} f_{\sigma_m,\lambda_m}$$

subject to

$$\sum_{k\in B(i)} f_{ki} - \sum_{j\in A(i)} f_{ij} = 0 \qquad \forall i$$

$$\sum_{(i,j)\in K_l} f_{ij} = 1 \qquad \forall l$$

$$f_{ij} \in Z_+ \qquad \forall i,j.$$

Levin did not solve this model, but pointed out that this model has less rows than the previous one does, thus possibly rendering less LP solving time.

Other works on planning of flight frequencies can be found in Dantzig [36], Elce [42], Etschmaier [43][44], Kushige [76], Miller [99], Richardson [117], and Soudarovich [133]. The works on determining departure times, on the basis of the time-of-the-day variability of demand and of the connections of flights, can be found in Gagnon [49], Labombarda and Nicoletti [77], Loughran [86], Richter [119][120], Struve [134], and Tewinkel [147].

Fleet assignment is to assign fleet types to flight segments after the flight schedule is determined. The period for which the assignment is done is normally one day for domestic flights. The factors that influence schedulers when assigning fleet types to various flights are: passenger demand, seating capacity, operational costs, and availability of maintenance at arrival and departure stations. Actual aircraft are routed after the fleets are assigned to ensure the solution is operational. The operational requirements at that stage mainly come from the necessary maintenance for each individual aircraft. A good flight schedule should also provide sufficient flexibility to enable efficient crew scheduling to be done. On the other hand, flight schedules are often revised to facilitate feasible or more effective flight assignments.

One important requirement of the fleet assignment is that the aircraft must circulate in the network of flights. These so-called balance constraints are enforced by using time lines to model the activities of each fleet type. In the time line model, there is a network built on the flight schedule for every fleet type. The components of the network for each fleet type are as follows: Each flight's arrival corresponds to a node at the arrival station and at the ready time, the time after which the aircraft can start to fly after the previous flight, and each flight's departure also corresponds to a node at the departure station and at the departure time. The ready times may differ among different fleet types due to the fleets' individual physical conditions. Connection each flight's departure and arrival nodes is the flight arc, and between each node and its next adjacent node on the time line at the same station is the ground arc. There is a ground arc for each station connecting the last node to the first node for a station to complete a daily operational cycle.

For each assignment of a fleet type f to a flight i, there is a cost c_{fi} incurred. The direct operational costs are easy to estimate. These costs include fuel costs, crew costs, landing fees, etc. But the cost due to the not-so-perfect assignment of a fleet type to a flight segment is hard to assess. Aircraft seats are the most perishable goods: once the aircraft takes off, the empty seats do not generate revenue but incur operational costs. Therefore, assigning a bigger aircraft than is needed to a flight is a waste of airline assets. On the other hand, by assigning a smaller aircraft than is needed to a flight, potential passengers will be spilled because of the inadequate capacity. Some of the passengers will be recaptured by other flights of the same airline, but the rest will be won over by its competitors and other transportation means. Hence, a substantial amount of revenue that could have been generated by these passengers will be lost by the airline. The difficulty in deciding c_{fi}'s lies in the fact that passenger demand on a flight depends on extremely complicated factors that are hard to fully capture. It takes experience and insight to make good assessment of c_{fi}'s.

Hane et al. [60] used a model based on time line networks. For each fleet type, there is such a network. They denoted nodes by fot, where f is for fleet type, o is for station, and t is for time (discrete number of them for each fleet's network). A flight arc that enters node fot is denoted by $fdot$ where d stands for the departure station of the flight; a flight arc that leaves is denoted by $fodt$ in which d stands for the destination station of the flight. A ground arc that enters the node from the precedent time t^-_{fot} is denoted by $fot^-_{fot}t$, while another leaving it for the subsequent time t^+_{fot} is denoted by $fott^+_{fot}$. Thus, each flight i on the schedule is represented by an $fo_d o_a t_d$ once and an $fo_d o_a t_{af}$ once for each fleet type f, in which o_d is i's departure station and o_a is i's arrival station, and t_d is i's departure time, t_{af} is i's ready time for fleet type f. An additional market constraint considered in Hane et al.'s model came from the *required throughs*, a certain set of consecutive flight pairs $(i_1, j_1), (i_2, j_2), ...$ have to be serviced by the same aircraft.

Other notations for describing the model are as follows: L is the set of flights; F is the set of fleet types; N is the set of nodes in all the time line networks of all the fleets; M is the set of all ground arcs; H is the set of all required through flight pairs; and $O(f)$ is the set of flight arcs that crosses midnight. $t_1(fo)$ is the first node on the time line of station o in the network of fleet f, while $t_n(fo)$ is the last. C is the set of stations serviced by the schedule; $D_d(fot)$ is the set of stations d's that make $fdot$'s flight arcs; $D_a(fot)$ is the set of stations that make $fodt$'s flight arcs; and $S(f)$ is the number of aircraft available for fleet type f.

Decision variables of the model are: X_{fi} (X_{dot}, $X_{fodt'}$) is the $0-1$ variable that indicates the coverage of flight i (also represented by dot and odt') by flight f. Decision variable $Y_{fott'}$ represents the number of type f aircraft at station o between adjacent time points t and t'.

The model is presented below:

$$\min \sum_{i \in L} \sum_{f \in F} c_{fi} X_{fi}$$

subject to

$$\sum_{f \in F} X_{fi} = 1 \qquad \forall i \in L$$

$$\sum_{d \in D_d(fot)} X_{fdot} + Y_{fot^-_{fot}t} - \sum_{d \in D_a(fot)} X_{fodt} - Y_{fott^+_{fot}} = 0 \qquad \forall \{fot\} \in N$$

$$x_{fi} - x_{fj} = 0 \qquad \forall (i, j) \in H$$

$$\sum_{i \in O(f)} X_{fi} + \sum_{o \in C} Y_{fot_n(fo)t1(fo)} \leq S(f) \qquad \forall f \in F$$

$$Y_{fott'} \geq 0 \qquad \forall \{fott'\} \in M$$

$$X_{fi} \in \{0, 1\} \qquad \forall i \in L, f \in F.$$

The first set of constraints guarantees that each flight is covered by one fleet type exactly once; the second set preserves the balance of aircraft at each transition point; the third set enforces the required throughs; the fourth set limits the utilization of each fleet type to the availability level; and the last two sets specify the decision variable ranges. The integrality of Y variables is brought forth by the integrality of X variables through the set of constraints. It should also be noted that the ground arcs are always used in the time-forwarding direction so that aircraft can depart only after they are ready.

Hane et al. devoted much effort to making real-size problems solvable in reasonable time in the frame of this model. Their solution procedure goes through a branch-and-bound routine. Prior to entering branch-and-bound, the problem's LP relaxation is solved, and the X's whose fractional solutions exceed 0.99 are fixed to 1. The other work aimed at reducing the problem size includes:

- Node consolidation – An aggregated node is used to replace a sequence of nodes that consists of consecutive arrivals and consecutive departures. The aggregation is legal, because it does not affect the causal restriction on the flight connections; and thus does not alter the solution.

- Island construction – In a hub station's schedule, a whole day's activity can be decomposed into several islands, with each island having an equal number of arrivals and departures. And at any time, the number of arrivals in the island is always no less than the number of departures in the island. Ground arcs that connect the islands can be forced to be 0. If there is only one arrival and one departure in an island, both flights must be flown by the same fleet type.

- Eliminating missed connections – If fleet type f misses a connection in another fleet type g's island structure, then there are flights A and B in the same island and a station s. s is A's destination and B's origin. A is ahead of B in g's network, but after B in f's network. If we assign flight type f to either A or B, an extra aircraft must overnight at the station, which is undesirable. So we forbid A and B from being flown by f by fixing the corresponding X's to be 0.

A similar integer multiple commodity network flow model was proposed by Yan and Young [161]. The model is also based on the time-space (time line) networks for all the fleet types n. The major difference from the previous model is that this one not only incorporates costs incurred by assigning a fleet type to flights in the model explicitly but also other costs such as costs incurred by ground holdings and overnight stays. After grouping the constraints that limit the use of aircraft within each fleet type to its available number and the constraints that guarantee that every flight is covered by one and only one fleet type, the model is laid out as follows:

$$\min \sum_{n \in M} \sum_{(i,j) \in A^n} C_{ij}^n X_{ij}^n$$

subject to

$$\sum_{j \in N^n} X_{ij}^n - \sum_{k \in N^n} X_{ki}^n = 0 \qquad \forall i \in N^n, \forall n \in M$$

$$\sum_{(i,j,n) \in H_s} X_{ij}^n \le \delta_s \qquad \forall s \in R$$

$$0 \le X_{ij}^n \le U_{ij}^n \qquad \forall (i,j) \in A^n, \forall n \in M$$

$$X_{ij}^n \in Z_+ \qquad \forall (i,j) \in A^n, \forall n \in M.$$

In the model, M is the set of fleet types, and N^n and A^n are the sets of nodes and arcs, respectively, in the network for fleet type n. R is the index set of the constraints that are grouped. For each $s \in R$, H^s is the set of triplets of (i, j, n)'s that are involved in these side constraints. In the model, the first set of constraints is the flow conservation constraints for the networks. The second set of constraints is the side constraints just mentioned. The third set specifies the lower and upper bounds for each arc. The fourth is for the integrality requirements.

Yan and Young utilized the Lagrangian relaxation method to get the lower bounds of the original optimal solution. The Lagrangian multipliers are operated on the grouped side constraints. They are updated from iteration to iteration based on the subgradient method [47][28] . In each iteration, the Lagrangian relaxation problem is solved by the decomposition-based network simplex method. Then, a Lagrangian heuristic is used to perturb the solution for the relaxation to get a feasible solution of the original problem, along with an upper bound. The stopping criterion is set to be that the difference between lower and upper bounds falls into a small gap, or that the number of iterations surpasses a prescribed value. A drawback of this solution technique stems from the fact that even the optimal solution of the Lagrangian relaxation over all possible multipliers does not supply a tight bound for the original problem.

Daskin and Panayotopoulos [37] also utilized the Lagrangian relaxation method to optimize approximately a fleet assignment problem. Their objective was to maximize profits in a single-hub-and-spoke network. Like the above mentioned work, heuristics are used to get a primal feasible solution along with a lower bound, after each subgradient multiplier updating step in which the upper bound is upgraded.

As introduced by Abara, American Airlines' fleet assignment model is based on a network comprised of nodes of flights and stations and arcs of turns [1]. Turns exist between stations and flights that represent flight sequence origination and termination and between flights that can be consecutively served by a single aircraft. Costs or profits related to flights are assigned to turns linking the flights. The model considers constraints of flight coverage, continuity of equipment, schedule balance (total number of departures equals total number of arrivals at each station in a planning cycle), aircraft count, and other nonstructural constraints. The objective could be to minimize total operational costs, to maximize total profits, or to optimize the utilization of certain fleet types. Using the model helped reduce American Airlines' operational cost by 0.5 percent, which is on the order of tens of millions of dollars. Berge and Hopperstad [23] provided a similar model, which is part of a dynamic fleet assignment system.

Desaulniers et al. [38] proposed two equivalent models to determine daily fleet assignment with maximum anticipated profits. The merit of the models is that they incorporate a flexible timetable in the flight schedule. The first model is of the set partitioning type with other constraints isolated for each fleet type representing the balance of number of aircraft at each station and the fleet capacity constraints. The set partitioning part specifies that each flight is covered once and only once by a flight sequence assigned to a certain fleet type. The second model is a time constrained multicommodity network flow formulation based on a network similar to that of Abara [1]. The only difference is that now two flights are connected as long as the first one's destination is the second one's origination, regardless of their scheduled timings. The time constraints come from the requirement that each flight be flown within a prescribed time window. To solve the first model, a branch and bound method with variable fixing and premature stopping is used. The upper bound is gotten from the LP relaxation of the model. The LP problem is further solved by Dantzig-Wolfe decomposition or column generation technique, with the subproblems being the longest path problem with time windows in the same network as the second model is based on. Computational experiments are run on real airline data, and solutions with substantial profit improvements are achieved in a reasonable amount of time.

Recently Talluri [139] presented an algorithm that uses assignment swaps to improve daily fleet assignment. The algorithm finds swap opportunities that satisfy the requirements of flow balance, aircraft count, and flight coverage. The algorithm uses a small number of calls to a shortest-path algorithm, which exists and is very efficient. The author also gave two further applications of the daily swap algorithm in schedule development and shows the way to incorporate a number of other factors into the algorithm.

Among all the carriers, Delta Airlines was the first to solve to completion one of the largest and most difficult problems in this industry [135]. Delta's Coldstart project models fleet assignment based on the time line networks. Besides constraints for flow conservations, flight coverages, and fleet size restrictions, the model has additional features to capture other operational requirements. For example, maintenance arcs are added, extending from an evening arrival node to a morning departure node at a certain station, to capture the incident of an aircraft being assigned to this arc and going through a twelve-hour maintenance procedure. In addition, many soft constraints are introduced into the model in the form of penalties in the objective functions, to prevent hard-to-detect infeasibility from occurring. For instance, fleet sizes are not fixed. Instead, excessive use of one fleet type incurs penalties in the total cost. The ten-fleet, 450-aircraft, 2500-flight-per day Delta Airlines yields a model containing some 40,000 rows and 60,000 variables.

The solution strategy for Coldstart is to use the OB1 interior point code [89] to solve the problem as an LP problem, fix some or all of the binary variables that are at 1.0 level, and solve the smaller problem after the variable fixing with the OSL mixed integer programming code [39]. Before solving the LP problem, node aggregation and other reduction techniques are applied. The size of the resulting LP problem is some 10,000 rows and 30,000 variables. Also, not all the 1.0 variables in the LP solution are fixed due to the fear of infeasibility. A heuristic is used to select some of the variables.

At Delta, Coldstart was used for purposes even beyond its original scope [89]. By changing the objective function to reflect variable fleet sizes and include ownership costs, the model aided fleet planning. When the objective was changed from cost minimization to profit maximization, the model developed routes by considering the addition of new legs and deletion of existing legs. The traditionally time-consuming manual job of moving the flight schedule from season to season can now be done by a slightly different version of the model. The heavy use of the model by Delta for fleet scheduling has recorded cost savings of $220,000 per day. It is also estimated that the savings achieved by use of the model will accrue to $300 million in three years.

5 CREW SCHEDULING

For most airlines, crew expense is the second largest cost component, second only to fuel expense. The number of daily flights for the largest airlines like American and United is in the thousands [55][59]. A small improvement in crew scheduling can lead to savings of millions of dollars. This has driven academia and the airlines to devote a large amount of effort into research in this area.

Most airlines begin to devise their crew assignments for the next planning period right after the flight schedule for the current period comes out. Typically, a planning period is about two weeks or one month. When deploying pilots and flight attendants to flights, the airlines must conform to limitations set by aviation administrations, union contracts, and their own work rules. The maximum amount of work the airlines may assign to the crews in a certain period, the minimum amount of rest time the crews must have between two consecutive flights, payments and compensation the airlines must make to the crews according to their work types, etc., are regulated. Also, each crew has a home base. It must return to the base after a sequence of flights. So crew scheduling's primary effort is to find a sequence of connected flight segments for each crew that start and end at its base, so that each flight segment in the planning period is served exactly by one crew. Because flight schedules repeat every several days within one operational cycle, the much simplified problem of finding the sequences of flights for the several days can be solved first, then the schedule for the whole period can be completed by repeating the sequences in the remaining part of the period. The sequences are called crew pairings. A valid pairing is one that conforms to all the limitations and rules mentioned above.

When assigning a crew to a crew pairing, there is a real cost to an airline. The principal component of the cost is *pay and credit*: the guaranteed hours of pay minus the hours actually flown. Other components include hotel costs, per diem, etc. The airlines try to avoid an unnecessary cost resulting from an unwise crew assignment as much as possible. Therefore crew scheduling's secondary goal is to find the least expensive way of allocating crews to flights. Most airlines make the desirable crew pairings first, then pack them into bills of work, or bidlines, each containing consecutive pairings that cover a planning period. Individual crew members then bid on them. Thus, there is no individual consideration in the scheduling period.

In its primary form, the scheduling problem is just the set partitioning problem, partitioning the set of flight segments into disjoint pairings, each containing a valid sequence of flights, to make the total cost the minimum:

$$(SPP) \qquad \min \sum_{j=1}^{n} c_j x_j$$

subject to

$$\begin{aligned} \sum_{j=1}^{n} a_{ij} x_j &= 1 \qquad && i = 1, ..., m \\ x_j &\in \{0, 1\} \qquad && j = 1, ..., n. \end{aligned}$$

In the formulation, each a_{ij} is a 0-1 constant that specifies whether or not crew pairing j includes flight segment i, each c_j is the cost that pairing j incurs, and each x_j is a 0-1 integer variable that indicates the selection of crew pairing j into the solution set. A general survey of this problem was provided by Balas and Padberg [13].

Sometimes, a crew need not be assigned to a connected sequence of flights. The disconnection can be mended by deadheading the crew from one station to another. This practice can be reflected in the model by changing the equality constraints to inequality ones of "$\geq$". When a flight is covered more than once, then some extra crews are just deadheading on board. This change causes a problem: the costs for the pairings are still calculated as though every flight segment in each of them is a duty flight for the crew that covers it. So many models simply don't consider deadheading.

In reality, the aggregate number of hours that crews located at a certain base spend away from the base has to fall into specified limits during each planning period. Taking this into account, there will be one more group of so called domicile constraints in the model. The constraints are normally expressed as:

$$d_b^l \leq \sum_{j=1}^{n} D_{bj} x_j \leq d_b^u \qquad b = 1, ..., B.$$

Also, in some models, instead of strictly enforcing the legality constraints for each pairing, penalties are imposed on pairings so that violations of legality come at a price. This more realistically represents the actual practice.

There have been many attempts to solve the crew scheduling problem, and more generally, the set partitioning and set-covering problems. Theoretical works often tended to find the optimal solution, but the sizes of the problems they dealt with were too small for their methods to be directly applicable to real world problems. Nevertheless, these works generated useful techniques for and provided intuitive insights into the problems. This was very beneficial to the practitioners. The real problems airlines face are comprised of up to thousands of flight segments (rows) and billions of crew pairings (columns). The columns are partially or wholly generated before or during the optimization process. Heuristics have to be used to get near optimal solutions.

Because some of the literature we will present deals directly with the crew scheduling problem, some tackles the general-purpose set partitioning and set-covering problems, and most uses a linear algebraic formulation, the terminology used is not consistent. In the course of incorporating it, we will also have to inherit this inconsistency. We will use flight segments, elements, and row indices interchangeably, and crew pairings, sets, and column indices interchangeably. Whichever term explains most clearly will be used.

Garfinkel and Nemhauser [50] proposed an enumeration algorithm to tackle the set partitioning problem. Their algorithm groups all the sets into blocks that are one-to-one mapped to all the sets, each consisting of sets that contain the element corresponding to itself, but contain no lower-indexed elements. The algorithm goes through a depth-first search. In a solution, at most one set from each block is chosen. So the search for a solution is sequentially done on blocks. A set in the current block that overlaps with any selected sets will not be chosen. These limit the effort needed in the enumeration. Christofides and Korman [31] added a lower bound from dynamic programming to facilitate a potential cut at each partial selection. They showed that their branch-and-bound algorithm compares favorably with many other existing ones.

Rubin [124] gave a heuristic that solves massive crew scheduling problems with domicile constraints. When an initial feasible solution is given, the heuristic keeps on improving the pairing selection locally by solving a smaller set partitioning problem for the set of flight segments that are covered by several pairings in the solution of the previous iteration. By choosing the flight segments this way, no improvement affects the covering of other unselected flights. Before handing the small set partitioning problem to an optimizer, some matrix reduction techniques are used. Due to the presence of domicile constraints, the optimal solution for the subproblem on a subset of flights varies when the covering of the rest of the flights varies. So, the subproblem for the same subset may be solved more than once. A proper storage mechanism is used so that feasible pairings for that subset of flights can be retrieved easily once they are generated. Rubin also proposed the following method for producing the initial solution: surrounding a known sequence of flight sequences with high-cost deadheads to a single crew base. The artificial deadheads gradually disappear in the course of iterations owing to their high cost, and a real feasible solution is achieved. Combined with an efficient set partitioning optimizer, such as the ones that will be mentioned later, Rubin's heuristic will be able to solve real-sized crew scheduling problems.

Marsten [93] proposed a branch-and-bound algorithm based on the same block structure that Garfinkel and Nemhauser used. The branching nodes in the algorithm are not the conventional partial setting of the decision variables. Instead, the constraint matrix's columns are grouped into blocks that differ in their members' first non-zero row positions. The branching nodes are taken to be the partial mapping of rows to blocks indicating for each row concerned in the mapping the chosen block from which a column to cover it will be chosen. The choices of branching from a branching node, equivalently, the options for mapping a next row, are well limited by a simple examination of the incumbent setting. This is rendered by the judicious choice of the block expression. When a partial mapping is decided, a restricted region can be specified for each block so that only columns within this region are included in a feasible solution. At each branching node, LP relaxation of the subproblem involving only the limited set of columns is solved, and its optimal value is taken as the lower bound for the remaining branching tree. When all the rows are mapped to some blocks, a feasible solution is reached; the global upper bound is updated to be the solution's objective if it is lower than the incumbent global upper bound. This algorithm worked well on some realistic crew scheduling problems, but didn't solve a problem of this kind with 400 rows. The numbers of columns in the problems used to test the algorithm were not big enough to be realistic. In a real crew scheduling problem, crew pairings, i.e., the columns, are not given. Normally, astronomical numbers of them are generated from the informations about the flight segments and the rules governing the choice of pairings. Without modification, this optimal algorithm is not capable of solving real-size crew scheduling problems.

Hoffman and Padberg [68] also introduced a branch-and-bound algorithm to solve the crew scheduling problem with domicile constraints. Its merit is the utilization of an LP-based heuristic, a preprocessor for the LP solver, and a constraint generation mechanism. The constraint generator delves into mathematics of polyhedral theory.

The algorithm's major searching engine in each node is an LP-based heuristic. The heuristic solves a series of LP subproblems repeatedly, seeking a feasible integer solution. With the help of the preprocessor, the LP problems put into the LP solver are mostly easier than the ones that are intended to be solved, with the original ones having some variables fixed to 0, and the constraint matrix reduced. Before entering the heuristic, the LP relaxation of the overall problem with some fixed setting of variables is first solved. If the solution is fractional, then a bigger partial setting is achieved by fixing some other variables according to their values in the LP solution, their reduced costs, and the gap between the upper and lower bounds [34][112]. Afterward, a loop starts. Both at this time and inside the loop, whenever the current LP is infeasible or its objective is higher than the global upper bound, backtracking is induced; meanwhile, if a feasible integer solution is found for the entire problem, the global upper bound is updated and backtracking is induced. Inside the loop, a sequence of reducing LP problems are solved. After each LP solution, a decomposition routine divides the reduced LP basis into blocks, and some controlled setting is done in each block based on the LP result, so that a smaller reduced LP is achieved. As mentioned earlier, the preprocessor further fixes a partial setting and reduces the problem. This reduced problem is again put into the LP solver. The loop goes on until infeasibility is detected, some exiting criteria are met, or a feasible integer solution is reached. Also placed after the LP solver is the constraint generator, which produces facet cuts to the polytope of feasible integer solutions' convex hull on the fly. It is activated after solving the entire LP relaxation or exiting the heuristic without finding a feasible integer solution.

The preprocessor aims at reducing the LP problems being solved by fixing variables to 0, eliminating redundant columns, merging columns, and eliminating rows of the constraint matrices. The reduction is achieved by doing certain manipulations repeatedly. The manipulations stem basically from two observations based on the strict equality to 1 in the problems' constraints. The first observation is: If G_A is the intersection graph of the set partitioning constraint matrix $A = \{a_{ij}\}$, i.e., G_A's nodes correspond to A's columns and a pairing of nodes has an arc linking them if their corresponding columns have entries of 1 in a common row; and if K is the node set of a clique in G_A, then there is at most one variable that can be 1 among all the variables corresponding to nodes in K. The second observation is: If M_i is the set of all the columns that have entries of 1 at row i, then there is one and only one variable associated with a column in M_i to be 1.

The constraint generator is run on the fly, utilizing the solution of the current LP problem. It generates facet cuts that participate in defining the convex hull of the overall problem's feasible integer solutions, or approximate in doing so. The set partitioning problem's facet cuts were studied by Arabeyre et al. [7], Balas and Ho [12], Balas and Padberg [13], Gomory [58], Padberg [111], and Sassano [126]. At the same time, these cuts are not redundant in the sense that they are detectable by the current LP solution to be "piercing through" the LP's feasible region. In essence, the constraint generator tries to find inequalities like

$$\sum_{j=1}^{n} b_j x_j \leq b_0$$

that are satisfied by every point of the IP convex hull, while

$$\sum_{j=1}^{n} b_j \overline{x}_j > b_0$$

with $\overline{x}$ being the current LP solution. The cuts are built from the set partitioning constraints and the domicile constraints, respectively. Those from the set partitioning constraints are mainly from two sources, both relating to the constraint matrix A's intersection graph G_A. One is, if K is the node set of a clique in G_A, then $\sum_{j \in K} x_j \leq 1$. The other is, if C is an odd cycle without chords in G_A, that is C has an odd number of nodes, forming a cycle and having no other arcs linking any two of them, then

$$\sum_{j \in C} x_j \leq (|C| - 1)/2.$$

Because the number of nodes and arcs of G_A is too large to be handled efficiently, the generator first gets cuts involving only the variables in $F = \{j | 0 < \overline{x}_j < 1\}$. The cuts result from doing clique and odd cycle detections in F's intersection graph G_F. Afterward, these cuts are raised iteratively to include all the variables. The lifting is based on the following logic: If

$$\sum_{j \in F} b_j x_j \leq b_0,$$

then

$$\sum_{j \in F} b_j x_j + b_k x_k \leq b_0,$$

with $b_k = b_0 - \max\{\sum_{j \in F} b_j x_j | A_F x_F \leq e_m - a^k\}$, where A_F and a^k are submatrices of the constraint matrix A being limited on column index sets F and $\{k\}$ respectively, and e_m is the $m \times 1$ vector of 1's. The order of the remaining variables in the column index set being added to F is randomly chosen. Efficient heuristics are developed for clique and odd cycle detections and for estimating $z_k(= \max\{\sum_{j \in F} b_j x_j | A_F x_F \leq e_m - a^k\})$'s upper bound. By using z_k's upper bound, the cut is loosened. If a cut is so loosened that the current LP solution no longer violates its defining inequality, then it is dropped.

Ball, Bodin, and Dial [15] and Ball and Roberts [16] attacked the crew scheduling problem through the graph theoretical approach. Their procedure performs set partitioning on a graph whose nodes represent all flight legs and arcs represent all possible connections between the flight legs. The problem of crew pairing becomes the problem of finding the minimum matching for the graph, where a matching of a graph is a subset of that graph's arcs with the property that no two arcs in the subset are connected to an identical node. This is equivalent to partitioning the graph to paths that satisfy certain minimum condition. The algorithm to solve this graph partitioning problem consists of two phases: pairing construction and pairing improvement. Computational tests of the algorithm proved the algorithm to be fairly efficient in dealing with the crew scheduling problem.

The algorithm's implementation CREW_OPT uses the steepest-edge dual algorithm of CPLEX developed by Bixby [24] as its LP solver. In the computational experiments conducted by the authors, astonishingly few nodes were visited before optimal solutions were reached for most of the problems. The most difficult problems used to test CREW_OPT have hundreds of thousands of non-zero entries in the constraint matrix. They were all solved to optimality in a matter of hours on a CONVEX model 550 machine. Using Hoffman and Padberg's solution technique, one is on the verge of being able to solve the airlines' crew scheduling problem as a single problem to optimality.

Lavoie, Minoux, and Odier [81] proposed a column generation method to solve the linear relaxation of the crew pairing problem. In their paper, a flight service is a sequence of flight segments that can be performed one after another. The cost of a crew pairing is total absence time plus total rest time for the crew, which reflects the real cost to the airline. The authors invented different states for each flight service, so that the validity of a crew pairing can be checked by each of the two consecutive flight services with their states, instead of by the whole sequence of the flight services. The method requires a preprocessing to construct a graph whose nodes are all the flight services with all the possible states and whose arcs linked all the valid flight service-state pairings. The graph contains all the possible pairings. The problem itself was formulated as a set-covering problem, with each column being a potential pairing. The authors used the column generation method to solve its LP relaxation. The subproblem of finding the minimum reduced-cost column was found to be the shortest path problem in the aforementioned graph. The method's implementation efficiently produced optimal results. Most of the resulting optimal solutions were integer solutions. If a solution was noninteger, the authors suggested solving the actual problem in the restricted pairing set which the linear solution's basis defines. This method was applied to crew scheduling for Air France. Marsten and Shepardson [94] were also able to resolve the noninteger solutions satisfactorily.

Since the early 1970's, American Airlines (AA) has been using the trip reevaluation and improvement program (TRIP) to plan its crew assignments. Based on Rubin [124]'s methodology, TRIP improves an initial manually-made set of pairings by iteratively solving set-partitioning subproblems on a subset of flight segments covered by several incumbent pairings. Each iteration involves subproblem selection, pairing generation, and pairing optimization. Marsten [93]'s method, based on branch-and-bound, partial mapping, and LP relaxation, was used in optimization. Since 1986, major enhancements to TRIP have been made [3]. Due to improvements in pairing generation and optimization, each iteration's speed has been increased tenfold. The column screening technique made TRIP capable of solving subproblems with 100,000 generated pairings, as opposed to 5,000 previously. Other techniques were utilized to reduce the chance of being trapped into local minima. These enhancements generated savings of about $20 million annually in crew assignment costs for AA.

In spite of all the improvements made to TRIP, it could only provide local optimal solutions. To move another step toward the global optima, American Airlines Decision Technologies (AADT) joined forces with IBM to come up with a better methodology [4]. They first generated all the pairings with low costs from the flight segments under consideration, then used a column generation procedure to solve the LP relaxation of the resulting set-partitioning problem. The procedure involves iterations of column selection and LP solving. Using the last iteration's dual variables, all the generated columns are priced out, and some of the columns are selected for the next iteration's LP solving if their reduced costs are low enough. To find the feasible integer solution near the fractional LP optimal solution, follow-ons for each flight segment are sought after the overall problem's LP relaxation is solved. In essence, a flight's follow-ons are other flights that very possibly will be its next flight in the same pairing in the integer solution, judging from the LP solution. Because a flight can be assigned to different pairings (fractionally) in the LP solution, branching has to be done by locking some of the follow-ons to try all the alternatives. The reduced LP after locking is solved again using column generation until all the flights have their follow-ons. At this time, a feasible integer solution is found. This method was applied to the crew scheduling problems of AA's Super 800 and B 727 fleets. The resulting improvements amounted to savings of $300,000 on the TRIP solutions for a three-month period.

Graves et al. [59] used an elastic-embedded set-partitioning integer programming model to solve United Airline's (UA) crew scheduling problem. This model does not allow crew deadheading, but allows legality to be violated at a price. In real life, this price comes as "credit time" for which the crew is paid in addition to actual flight time, basically for the inconvenience incurred by the violation. Also, it considers the domicile constraints. Their solution method first generates all the feasible pairings and finds a disjoint or nearly disjoint set of pairings as an initial solution. It then makes 2-OPT or 3-OPT local improvements on all the possible combinations of two-pairing or three-pairing subsets of the solution set. The method has been in use since October 1989. It reportedly has been able to solve UA's narrow-body problem of about 1,716 flight segments in about 800,000 cpu seconds on an IBM 3090 mainframe computer. It is estimated that the system saves about $12 million annually in credit time and about $4 million in hotel costs.

6 AIR TRAFFIC FLOW CONTROL

Flight delays increase operational costs of the delayed flights and the affected downstream flights, inconvenience passengers, and thus damage the credibility of the airline and hurt passengers' goodwill for the airline. On the other hand, the high demand for air travel has brought airports to their saturation points. A small reduction in the airports' capacities can affect many flights' on-time service. Because the most common factor that impacts airports' capacities is bad weather, which is currently beyond human control, not much can be done to improve the stability of airports' capacities. What can be done is to mitigate the capacity reduction's impact on airline operations.

Delays are divided into two types, the less costly ground delays beyond takeoff times and the more costly airborne delays beyond landing times. When delay at the destination airport is foreseen before a flight has taken off, ground delay can be imposed on it to avoid more costly future airborne delay. In the United States, the Air Traffic Control System Command Center (ATCSCC) of the FAA is doing just that. It projects the capacity and demand at various major airports in the time horizon of several hours on a daily basis and makes ground holding assignments to various flights. But the decisions from the ATCSCC are primarily based on the expertise and judgement of human controllers. Much improvement can be made with more thorough and accurate flow management.

For the ground holding problem (GHP), work was first done on single-airport problems or problems with few airports involved. Odoni [106] gave a systematic description on the problem. Terrab [145] solved the static version of the single-airport GHP and gave several heuristics for the probabilistic version. He also presented formulations for static versions of the problem with two and three airports. Richetta [118] tackled the single-airport dynamic probabilistic GHP. Vranas et al. [152][153] first considered the network effect of GHP.

Andreatta and Romanin-Jacur [5] tried to find the best strategy to make ground holding assignments in the presence of airport congestions under some greatly simplified assumptions. The simplified problem instance being considered is: A group of flights $V = \{v_1, ..., v_n\}$ are all scheduled to land at the same destination airport at $t = 0$. The probability of the destination's capacity being $i, 1 \leq i \leq n$, is known a priori as $p(i)$ and known to be unchanged when time approaches 0. All delayed aircraft will be able to land at $t = 1$. For flight i, the ground delay cost is g_i, and the airborne delay cost is a_i. The authors first noticed that the delay cost was affected by the landing priorities of the flights not being held on the ground. For each aircraft v_i that reaches the destination at $t = 0$ among the q chosen in the set of flights U^q, it can land immediately if and only if the airport's capacity exceeds the number of airborne aircraft that have landing priorities higher than its own. If $P(k) = \sum_{h=0}^{k} p(h)$ is the probability that airport capacity is not higher than k, and U_i^q is a subset with cardinality B_i^q of flights in U^q whose priorities are higher than that of v_i's, then the total delay cost under the choice of U^q and priority π is:

$$C(q, U^q, \pi) = \sum_{v_i \in U^q} (a_i P(B_i^q) - g_i) + \sum_{v_i \in V} g_i.$$

The higher the priority v_i has, the smaller $P(B_i^q)$ is. So to minimize the cost, the optimal priority assignment should be that the flight with larger a_i receives higher landing priority.

When the landing priority is determined such that v_{i+1} has higher priority than v_i, $\forall i \in \{1, ..., n-1\}$, the number of aircraft that have landing priority over v_i is $S_i = n-i$. If d_i indicates the decision on whether or not v_i is to be held on ground, then D_i, the number of flights with priority over v_i that are held on the ground, can be expressed as $D_i = d_{i+1} + D_{i+1}$. If $d_i = 0$, then $S_i - D_i$ is the number of aircraft with priority over v_i that are not delayed on the ground, and the expected airborne delay cost caused by v_i is $a_i P(S_i - D_i) = a_i P(n-i-D_i)$. Taking into account the possibility of $d_i = 1$, the expected cost caused by v_i is $c_i(d_i, D_i) = d_i g_i + (1-d_i) a_i P(n-i-D_i)$. The authors introduce the quantity $\bar{C}_i(D_i)$ as the optimal value of the expected delay cost to the first i aircraft given that D_i aircraft with higher priority than i have been delayed on the ground. The following forward recursion equation exists for $\bar{C}_i(D_i)$:

$$\bar{C}_i(D_i) = \min_{d_i} \{c_i(d_i, D_i) + \bar{C}_{i-1}(d_i + D_i)\},$$

along with the initial condition:

$$\bar{C}_0(d_1 + D_1) = 0.$$

Using an $O(n^2)$ dynamic programming algorithm, $\bar{C}_i(D_i)$'s of all possible i's and D_i's are calculated. The optimal total expected cost is $C^* = \bar{C}_n(0)$, and the optimal decision is $d_n^* = \bar{d}_n(0)$, $d_{n-1}^* = \bar{d}_{n-1}(d_n^*)$, ..., $d_1^* = \bar{d}_i(\sum_{j=2}^n d_j^*)$.

The authors also prove that the optimal set of ground-delayed flights when the allowed number of ground delays is specified to be q contains the set when the allowed number is $q-1$. From this, a simple algorithm is devised for the case when the number of allowed ground delays is specified. Though the authors reached very satisfactory results, their assumptions are too simple for the results to render any practical use. In real life, there is more than one destination airport to be considered, time of delay cannot be set to be unique, and the prediction on airport capacity will change during the time when none of the aircraft being considered have taken off.

Terrab and Odoni [146] also addressed the single-airport ground holding problem with deterministic and stochastic capacity forecasts. The authors assume that airborne delay cost is always higher than ground delay cost and the takeoff capacities are unlimited. When the forecast is determined before any flight has been executed, any airborne delay at the destination can be obsorbed by the same amount of ground delay at the origin. So, only the amount of delay must be decided for each flight in the deterministic case. In the stochastic case, several scenarios with certain probability distribution exist for the outcomes of airports' capacities. At the time of an aircraft taking off, the particular scenario has not unfolded. They assumed that when an aircraft reaches an airport's airfield, the landing capacity at the airport is determined. In this case, a common decision for each flight at takeoff time is made anticipating any scenario, and an airborne delay decision is made for every scenario for the flight. The objective in the stochastic case is to minimize the expected total delay cost. Exact dynamic programming algorithms were devised for both cases. Realizing the impossibility of the exact solution for the stochastic case, the authors came up with several heuristics for the problem.

Considering the congestion at one single airport is not sufficient. The capacities of airports located near each other tend to be affected by similar weather patterns. Because runways are tightly-used resources, downstream airports' congestion will also cause congestion to upstream airports. Thus, congestion frequently happens to several airports simultaneously. Wang [155] decomposed congestion events into sets of events such that every event in each set has at least k impacted flights in common with any other event in that set and has less than k impacted flights in common with any event outside the set. Within each set, the order in which the events' congestions are to be erased is determined by running the shortest path algorithm on a network containing the information on relations between costs and order of congestion erasing. Strategies on how individual congestions are erased were not provided, and the sizes of networks are of exponential orders of the numbers of congestion events, which renders the process impractical.

Vranas et al. [152] considered the ground holding problem for a whole network of airports. They assumed that each airport's capacity at any time of the day is deterministic and known a priori. Another big step they took in modeling was to discretize the time horizon into many time points. The authors proposed three models: the first one considers both ground and airborne delays, the second one assumes that there are no airborne delays, and the third one adds flight cancellations to the second model. Here, we discuss only the authors' third model. Before presenting the model, however, we have to introduce the necessary notations. $\{1, ..., K\}$ is the set of airports, $\{1, ..., T\}$ is the set of time points, and $\{1, ..., F\}$ is the set of flights. For each flight f, r_f is its scheduled arrival time, and k_f^a its destination airport; the cost for delaying it a unit time on ground is c_f^g, the cost to cancel it is M_f, and the maximum allowed delaying time for it is G_f. Therefore, T_f^a, the set of possible arrival time for flight f, satisfies

$$T_f^a = \{t \in \{1, ..., T\} : r_f \leq t \leq \min\{r_f + G_f, T\}\}.$$

For each continued flight f' whose aircraft is used for a next flight f, $s_{f'}$ is the slack time built into the flight schedule. It is the surplus time between the two consecutive flights such that if f' is at most $s_{f'}$ late, f can still be executed on time. For simplicity, the next flight for f' is denoted by f instead of $f_{f'}$. For each airport k, its capacity at time t is $R_k(t)$. The decision variables are:

v_{ft}, which indicates whether flight k arrives at time t;
g_f, which represents flight f's delay time; and
z_f, which indicates whether flight k is delayed.

To represent both the situations of having and of not having spare resources, the authors partition the set of flights into two sets F'_1, the set of continued flights whose cancellations will not affect the next flight, and F'_2, the set of continued flights whose cancellations will.

The model is as follows:

$$\min \sum_{f=1}^{F} (c_f^g g_f + (M_f + c_f^g r_f) z_f)$$

subject to

$$g_f = \sum_{t \in T_f^a} t v_{ft} - r_f \qquad f \in \{1, ..., F\}$$

$$\sum_{f : k_f^a = k} v_{ft} \leq R_k(t) \qquad k \in \{1, ..., K\}, t \in \{1, ..., T\}$$

$$z_f + \sum_{t \in T_f^a} v_{ft} = 1$$

$$g_{f'} - s_{f'} + (s_{f'} + r_{f'} - r_f) z_{f'} \le g_f \qquad f' \in F_1'$$

$$g_{f'} - s_{f'} + (s_{f'} + r_{f'} + G_f + 1) z_{f'} \le g_f + (r_f + G_f + 1) z_f \qquad f' \in F_2'$$

$$v_{ft}, z_f \in \{0, 1\} \qquad f \in \{1, ..., F\}, t \in \{1, ..., T\}.$$

The first set of constraints stipulates that the definition of g_f is the amount of ground delay for each flight f if it is not canceled. If it is, $g_f = -r_f$ because none of v_{ft} takes the value 1. The objective function is just the total delay and cancellation cost. When $z_f = 1$, the part contributed by f is just M_f. The second set of constraints specifies that the number of arrivals at any time not exceed the airport's capacity at that moment. The third set of constraints states that a flight is either canceled or to arrive during the allowed period of time. When $z_{f'} = 0$, the fourth and fifth sets of constraints all become $g_{f'} - s_{f'} \le g_f$, which observe the constraints imposed by the minimum slack times. When $z_{f'} = 1$, the fourth set of constraints becomes $g_f + r_f \ge 0$, which are always true, and the fifth set of constraints becomes $G_f + 1 \le g_f + (r_f + G_f + 1) z_f$, enforcing $z_f = 1$.

The authors solved the models using a heuristic based on LP relaxations. In the heuristic, the set of flights, Φ, with fractional solutions is divided into classes, with each class containing all the flights in Φ to be flown by one aircraft. Each class is processed one at a time, and the flights in each class are also treated one at a time in the order in which they are to be flown by the aircraft. At the time flight ϕ in Φ is being treated, the times allowed for it to land at its arrival airport k_ϕ^a are examined. The first t that satisfies the airport capacity constraint $\sum_{f: k_f^a = k_\phi^a} v_{ft} \ge 1 - v_{\phi t}$ is assigned to be flight ϕ's arrival time; that is, $v_{\phi t}$ is set to be 1. If such a time is not found, then ϕ is canceled. If ϕ is assigned a time, the allowed time for its following flight is restricted to that satisfying the coupling constraint. The authors conducted computational experiments of the three models along with the modified versions of the coupling, network constraints being removed. From that, they drew the following conclusions:

- In general, network effects can be large, which implies that considering ground delaying independently at each airport is insufficient.
- Finite departure capacities have negligible impact if they are assumed not to influence arrival capacities.

- The heuristic performs well for low cancellation costs.

In another paper by the same authors [153], the dynamic ground holding problem was addressed. When new and more reliable information regarding airport capacities at subsequent times comes, new decisions are made to make the best use of them. In the paper, decisions are made at certain time points for those flights that have not taken off or landed. For those flights made to land based on previous decisions, the new decisions cannot affect them. At time τ when a decision is to be made, the set of flights that has not taken off is denoted by F_τ^g, and the set of flights that is in the air and has not landed is denoted by F_τ^a. The mathematical formulation for the decision making is similar to the static one. The only difference is that, in the dynamic model, the decision on ground holdings of flights in set F_τ^a is irrevocable, and the decision on ground holdings of flights in set F_τ^g can be made to override the previous one, while in the static model there is no such a disparity. Between two consecutive decision time points τ and $\hat{\tau}$, the sets F_τ^g and F_τ^a evolve according to the following rules:

$$F_{\hat{\tau}}^g = F_\tau^g / \ \{f \in F_\tau^g : d_f + g_f^\tau < \hat{\tau}\},$$

$$F_{\hat{\tau}}^a = (F_\tau^a / \ \{f \in F_\tau^a : r_f + G_f + a_f^\tau < \hat{\tau}\}) \cup$$
$$\{f \in F_\tau^g : (d_f + g_f^\tau < \hat{\tau})(r_f + g_f^\tau + a_f^\tau \geq \hat{\tau})\},$$

where all the decision variables are written in the same form as in the first paper, except for the additions of superscripts standing for decision making times. Also for $f \in F_\tau^a$, G_f is its ground holding time which is already irrevocable. The rules include: once a flight is planned at τ to take off before the current time $\hat{\tau}$, then its departure can no longer be held at $\hat{\tau}$. Also, if a flight is scheduled to land before the current time in the last decision period, no decision can be made on it; but, if a flight scheduled at τ is to take off before $\hat{\tau}$ and is to land no earlier than $\hat{\tau}$, then the decision on its landing is subject to revision at this point.

In a more restricted model, the authors assumed that the airports' take-off capacities are infinite. At the time decisions are made, the only decisions to be made are on ground holdings. The ground holdings are made to absorb airborne delays, though conditions at airports may be different from what is expected at the times of flights' arrivals. The gist is, whenever a takeoff will incur airborne delay according to the current capacity forecast, it is not allowed to take off. Airborne delays only occur to those flights that are already in the set F_τ^a, and only occur when they have to. Thus, airborne delays are not regarded as decisions to be made, and the minimum costs they incur are inevitable for the current stage of planning. To accommodate this strategy, for each airport k and time point t, an excess E_{kt}^τ is introduced, which represents the number of aircraft about to land in excess of the airport's landing capacity. Note that the excess is inherited from the last period's decision based on inadequate information. Another set of variables α_f^τ carries the length of unavoidable airborne delays for the flights. Whenever $E_{kt}^\tau > 0$, E_{kt}^τ of the flights about to land at airport k at time t are selected to be delayed another time unit. After the preliminary calculations, the authors arrived at such a model at planning time τ:

$$\min \sum_{f \in F_\tau^g} c_f^g g_f^\tau$$

$$g_f^\tau = \sum_{t \in T_f^a \cap T_\tau} t v_{ft}^\tau - r_f \qquad f \in F_\tau^g$$

$$\sum_{f \in F_\tau^g : k_f^a = k} v_{ft}^\tau \leq \max(-E_{kt}^\tau, 0) \qquad \forall k \in \{1,,,.K\}, t \in \{1, ..., T\}$$

$$\sum_{t \in F_f^a \cap F_\tau} v_{ft}^\tau = 1 \qquad f \in F_\tau^g$$

$$g_{f'}^\tau - s_{f'} \leq g_f^\tau \qquad f' \in F' \cap F_\tau^g$$

$$G_{f'} + \alpha_{f'}^\tau - s_{f'} \leq g_f^\tau \qquad f' \in F' \cap F_\tau^a$$

$$v_{ft}^\tau \in \{0, 1\},$$

where T_τ is the set of time points at and after time τ. The notations used here are essentially the same as in the previous paper. Only now, the time τ at which the decision is made has to be explicitly specified. $G_{f'}$ is the ground delay decided for flight f' before τ which cannot be revoked at τ.

In addition, the authors also considered extensions to the models when cancellations are incorporated and when airports' takeoff and landing capacities are interrelated. The corresponding changes to the models are insignificant. Another extension considers the probabilistic factor in capacity forecasting. The model assumes that a forecast is not deterministically made, rather probability distribution of several scenarios is given. Only during the short period of airborne holding will one of the scenarios be realized. A unique decision is made at takeoff time, but different decisions are made during flight corresponding to different realized scenarios. The goal is to minimize the total ground delay costs plus expected total airborne delay costs. The authors conducted extensive computational experiments. Following are some of their conclusions:

- If incorrect capacity forecasts made at the beginning of the day are corrected early enough, then their influence on the total cost of the dynamic problem can be minimized.
- A greedy dynamic FCFS heuristic simulating current practices concerning ground holding decision is highly inefficient compared to ground holdings based on the optimal solutions of the dynamic decision models presented in the paper.

Another model dealing with the multi-airport multi-period congestion problem was provided by Helme [64]. In addition, this model considers en route capacity restrictions. The drawback of the model is that it does not consider any downline effects beyond the destinations. The model treats flights as commodities differentiated by their destinations. The commodities form conserved flows in a space-time network. The network consists of three types of vertices. The origins are origin airport and time pairs; the destinations are destination airport and time pairs; and the fixes are en route space-time vertices that enable enforcement of en route capacities and inclusion of airborne delay costs. Time is discretized to be time points 15 minutes away from each other. There are ground arcs in time advancing direction linking origin vertices standing for the same origin airports, and linking destination vertices standing for the same destination airports. Also there are airborne arcs in time advancing direction linking fixes representing the same locations. It is assumed that all flights between two airports take the same amount of time without delay. The flight arcs linking origin, fix, and destination vertices correspond to flights linking one vertex representing one location and time pair to another vertex representing another location and time pair. The time span between two vertices linked by a flight arc is decided by the two locations. A unit flow on a ground arc incurs a unit ground delay, and a unit flow on an airborne arc incurs a unit airborne delay. Thus, delay costs can be placed on the ground and airborne arcs. Landing, arrival, and en route capacities are all placed on corresponding arcs. The model also assumes that all the flights being considered are executed within the time horizon it covers. The planned departures form flow sources at various origin vertices, and the required landings form flow sinks at destination vertices at the ending time.

The following notations are used in the model:

V = set of vertices in the network
A = set of arcs in the network
K = set of destinations, thus commodities
e_{ij} = minimum travel time on link from airport i to airport j
c_{it}^{ju} = cost of one unit of flow on the arc from vertex (i, t) to vertex (j, u)
$F(i, t) = \{(j, u) \in V \mid [(j, u) - (i, t)] \in A\}$ = set of "from" vertices for vertex (i, t)
$G(i, t) = \{(j, u) \in V \mid [(i, t) - (j, u)] \in A\}$ = set of "to" vertices for vertex (i, t)
r_{it}^{k} = net inflow of commodity k at vertex (i, t), for $k \in K$
b_{it}^{ju} = capacity of arc $[(i, t) - (j, u)]$

The decision variables are $x_{it}^{ju}(k)$, the number of aircraft headed for airport k, departing from location i in period t and arriving at location j in period u, $\forall [(i, t) - (j, u)] \in A, k \in K$. The formulation is listed below:

$$\min \sum_{[(i,t)-(j,u)] \in A} c_{it}^{ju} \sum_{k \in K} x_{it}^{ju}(k)$$

subject to

$$\sum_{(j,u) \in G(i,t)} x_{it}^{ju}(k) - \sum_{(j,u) \in F(i,t)} x_{ju}^{it}(k) = r_{it}^{k} \quad \forall (i, t) \in V, k \in K$$

$$\sum_{k \in K} x_{it}^{ju}(k) \leq b_{it}^{ju} \quad \forall [(i, t) - (j, u)] \in A$$

$$x_{it}^{ju}(k) \geq 0 \quad \forall [(i, t) - (j, u)] \in A, k \in K$$

$$x_{it}^{ju}(k) \in Z_{+} \quad \forall [(i, t) - (j, u)] \in A, k \in K.$$

This is a multicommodity minimum cost network flow formulation. The solution method proposed by the author is to find the initial feasible solution by straightforwardly scheduling delay flights on the ground for the sole purpose of accommodating reduced arrival capacities and then to improve the solution step by step, such that in each step, a negative cost cycle is found and flow is added in the cycle. In the early stage of development, only fictional data were used for experimentation by using part of the improving cycle method.

Rue and Rosenshine [125] used Semi-Markov decision process models to study the optimal control of access to the landing queue of an airport. There is another school of literature which concentrates on the physical aspect of air traffic. These directly use physical parameters of aircraft, airports, and air space as input variables and give guidelines for inter-aircraft distances, landing and takeoff delays, and other air traffic control elements in various circumstances. Newell [105] gave a survey on airport capacity. He described how the capacity of a runway configuration depends upon the strategy for sequencing operations such as arriving and departing of heavy and light aircraft, the runway geometry, the instrument flight rules, etc. Andreussi et al. [6] introduced a discrete event simulation model that simulates the aircraft sequencing operations in the near terminal area. The model follows the landing procedure from common practice that all aircraft fly the "race-track" shaped paths at different altitudes before leaving from several fixes to enter the runway. Models concerning en route traffic conflict and capacity were given by Blumstein [25], Dunlay [41], Geisinger [53], Hockaday and Kanafani [67], Janic and Tosic [72], and Siddiqee [127].

7 IRREGULAR OPERATIONS CONTROL

Airlines operate in a very complex environment. Many factors are beyond human control. The one factor that affects airline operations the most is inclement weather. Severe weather situations worsen conditions at airports, hence reducing the allowed arrival and departure rates, sometimes even forcing airport closure. For each airline's every arrival at every airport in the United States, an arrival time slot is assigned to it by the ATCSCC. Under normal conditions, the slots match the flights' scheduled arrival times. In the face of bad weather or abnormal conditions, the ATCSCC allocates to airlines slots with reduced arrival rates that make delays and cancellations unavoidable. Each slot is an interval of time centered around a controlled arrival time within which the arrival has to be made. The airlines have certain freedom in making their own decisions as to which flights are to be canceled and which slot each flight will fill under constraints such as rescheduled arrival not being earlier than the scheduled arrival time. The airlines' decisions are fed back to the ATCSCC for approval. An approved slot allocation can be put into execution, while denial forces the whole process to be repeated. Very often, an allocation is denied because an airline takes too long in making the online slot allocation decision. Different slot allocations have different effects on the number of passengers who have to be transferred to other flights or compensated due to flight cancellation, on the total amount of passenger delay time, on the airlines' dependability data, and on the intangible factors such as passenger goodwill and the airlines' reputation. Therefore, a systems that effectively and efficiently allocates flights to slots is highly sought after.

An arrival slot allocation system(ASAS) was implemented at American Airlines in 1989 [150]. It consists of an algorithm that minimizes the amount of delay, taking advantage of flight cancellations and a data-processing component that cuts down the overall turnaround time of responses to ATCSCC by automating the process of sending and receiving messages. The slot allocation model is associated with the directed traveling salesman problem [82]. Each flight segment under consideration is represented as a node in a network; a salesman has to visit each node, deliver a slot in each visit, and create a tour through the network. The order of nodes in the salesman's tour corresponds to the sequence of arrival slot substitutions available for each empty slot. The algorithm to solve it is a tour-building heuristic that preserves aircraft and crew balance and gate connections among flights at hub airports. The ASAS has saved American Airlines the cost of seven human dispatchers and an additional \$5.2 million due to reduced amounts of delay time.

In order to accurately measure the perturbation's propagation effect caused by the Ground Delay Program (GDP), Luo and Yu [87] introduced and formalized the concepts of critical departure times and critical arrival times. For each flight, its 1-stage critical departure time is the latest time that the flight can depart without affecting the flight's subsequent scheduled activities; its n-stage critical departure time is the latest time that the flight can depart without affecting the scheduled activities after n subsequent flights have been executed. The critical arrival times are similarly defined. Aircraft and crew (even individual crew members) are treated as generic resources. The authors point out that, in general, the pure slot allocation problem without considering aircraft maintenance, crew legality, and the splitting of resources (crew and aircraft fly the same subsequent flights at the problem station) is an assignment problem that can be solved to optimality in polynomial time. Nevertheless, for various criteria, faster algorithms are provided by the authors. Because the airlines' operation controllers have to respond as quickly as possible to the ATCSCC, the algorithms will be helpful to the controllers. When the objective is to minimize total passenger delay, the authors' solution strategy is to assign slots chronologically, and for each slot being considered, assign to it the earlier unassigned arrival that has the largest passenger load. The complexity of the procedure is $O(n^2)$. Another similar algorithm was devised for minimizing the total delay beyond critical times. To minimize the number of delayed out-flights, i.e., the number of flights that are delayed beyond their 1-stage departure times, an algorithm adapted from Moore's algorithm [101] for solving the one machine scheduling problem $n/1//n_T$ is employed. For each in-flight, its corresponding due date is the latest slot that enables the subsequent out-flight to be before the 1-stage critical departure time. The algorithm goes through two steps. In the first step, all the in-flights are ordered non-decreasingly in their due dates and get sequence J. In the second step, the first in-flight j in J whose associated out-flight is delayed under the current landing sequence is identified and taken out of J. The process is repeated until no in-flight whose associated out-flight is late remains in J. The algorithm takes $O(n\log n)$ time.

When resources are splittable after landing (crew and aircraft follow different routes at the problem station), each out-flight will need several types of resources that are provided possibly by different in-flights. To capture the lateness of out-flights, due date q_i of an in-flight $i \in I$ has to be defined as the earliest departure time of the out-flight that uses the resources from this in-flight minus the minimum turnaround time:

$$q_i = \min\{d_k \mid \phi_r(k) = i, r \in R, k \in K\},$$

where $\phi_r(k)$ denotes the in-flight that provides out-flight k with the rth resource, R is the set of resources needed for flight operation, K the set of out-flights, and d_k is the scheduled departure time of out-flight k minus the minimum turnaround time. Luo and Yu [88] study the landing slot allocation problem caused by GDP with splittable resources using several criteria. The first objective they consider is to minimize the maximum delay L_{max} among all out-flights. When in-flight i is assigned to landing slot $\sigma(i)$, the definition of L_{max} is:

$$L_{max} = \max_{k \in K}[\max_{r \in R}\{s_{\sigma(\phi_r(k))} - d_k\}]^+ = \max_{i \in I}[s_{\sigma(i)} - q_i]^+,$$

where s_j is slot j's starting time. The authors proved that the method of assigning in-flights with earlier due dates to earlier landing slots guarantees the optimal result.

To formulate the model to minimize the total number of in-flight and out-flight delays, the following decision variables were introduced: x_{ij} indicates whether in-flight i is assigned to landing slot j and y_k indicates whether out-flight k is on-time. The formulation is as follows:

$$(GDPLT) \qquad \min \sum_i \sum_{j:s_j>a_i} x_{ij} + \sum_k (1-y_k)$$

subject to

$$\sum_{j:s_j \geq a_i - \delta} x_{ij} = 1 \qquad \forall i \in I$$

$$\sum_{i:a_i \leq s_j + \delta} x_{ij} = 1 \qquad \forall j \in J$$

$$\sum_{j:s_j > d_k} x_{\phi_r(k)j} + y_k \leq 1 \qquad \forall k \in K, r \in R$$

$$x_{ij}, y_k \in \{0,1\} \qquad \forall i, j, k.$$

In the formulation, a_i denotes the scheduled landing time for flight $i \in I$, s_j denotes the starting time of slot $j \in J$, and δ denotes the interval of slot j. The first two sets of constraints require that every in-flight is assigned to exactly one landing slot not earlier than its scheduled landing time; the third set of constraints declares the meaning of y_k's. The authors recognized that integrality constraints on x_{ij}'s can be dropped as long as they are enforced on y_k's. Also they find the following useful valid inequalities:

$$\sum_{i':a_{i'}x_{i'j} \leq a_i} \sum_{j:s_j>a_i} \geq \mid \{i' : a_{i'} \leq a_i\} \mid - \mid \{j : s_j \leq a_i\} \mid \qquad \forall i \in I,$$

which specifies the minimum necessary delays for the first i in-flights;

$$\sum_{k':d_{k'} \leq d_k} (1 - y_{k'}) \geq \mid \{k' : d_{k'} \leq d_k\} \mid - \mid \{j : s_j \leq d_k\} \mid \qquad \forall k \in K,$$

which specifies the minimum necessary delays for the first k out-flights;

$$\sum_{i=\phi_r(k),\exists r\in R} x_{ij} + y_k \leq 1 \qquad \forall j,k : s_j > d_k,$$

which states that an out-flight is late whenever an in-flight that provides resources to it is late; and

$$\sum_{j:s_j>d_{\phi_r^{-1}(i)},\forall r\in R} x_{ij} + (\frac{1}{\mid R\mid})\sum_{r\in R} y_{\phi_r^{-1}(i)} \leq 1 \qquad \forall i \in I,$$

which makes sure that if an in-flight lands later than d_k's of all the out-flights that need resources from it, then all the out-flights will be delayed. With the help of these valid inequalities and the aforementioned model simplification, optimal solutions for realistic problems were reached within seconds on microcomputers. A way to get a good upper bound was also provided by the authors: Find the level of optimal in-flight delays; then at this level, optimize the number of takeoff delays.

When the resources are split into at most two, and if between the d_k's of two out-flights requiring resources from the same in-flight there are no landing slots, the authors provided a very simple model for minimizing the number of takeoff delays. The two assumptions used are very realistic: Normally the only two resources that need consideration are crew and aircraft, and takeoffs are much closer in time than landings. When out-flights are thought of as nodes, and two out-flights that need resources from the same in-flight are linked by an arc representing the in-flight, the relationship of resource sharing and flowing can be represented by clusters of cycles distributed in many time intervals. Between the beginning of time and the first interval, and between these intervals, there are landing slots. An out-flight arc is on time if and only if it is assigned to a landing slot before it. An out-flight node is on time if and only if the two in-flight arcs that link it are all on time. The problem is that of maximizing the number of on-time nodes with the restriction of limited landing slots. The authors provide an $O(n)$ time heuristic that guarantees that the solution will be less than 1 delay away from the optimal one. The heuristic tries to assign arcs that complete cycles, then at last leaves a chain which has 1 node less than arcs that are on time. This is a greedy complete-cycle algorithm. The exact solution procedure provided by the authors tries to find the cycles that accommodate the maximum number of on-time arcs; the slots left are assigned to arcs that form a chain. The effort to find the cycles are proven to be of $O(n^2)$ complexity.

Other major factors that affect airlines' regular operations are mechanical failures and crew absence which cause aircraft shortages. The shortage of an aircraft for one flight will propagate to downline operations and bring great inconvenience to passengers and significant damage to the airline's current operations and long-term profitability. Although reallocating slots can reduce the effects of abnormal conditions at the current station, the downline effects often are still enormous and require proper treatment. To mitigate the rippling effect of flight delays and cancellations caused by a few flights, airlines can opt to reschedule later flights and make appropriate adjustments to aircraft, crew, and gate assignments. Many authors have considered the aircraft routing problem that deals with unplanned operational changes. Because the flight schedule of an airline is always very tight, locally adjusting the schedule and aircraft routing for full recovery is often impossible. A scheme for global recovery is needed to reduce the irregular operations cost to the airline. To estimate how the recovery decisions perform, costs of delaying and canceling flights must be specified. Also, minimum turnaround time for each aircraft is imposed, so that an aircraft can take off again only after it has been on the ground for that amount of time. For each station, there are curfew times after when no takeoff and landing can take place. For the purpose of continuous operation and maintenance, aircraft balance must be kept. In other words, for each station, a prescribed number of aircraft must overnight there.

Teodorovic and Guberinic [142] proposed a graphical model that does not consider flight cancellation and curfew constraints. In the model, all the flights to be rescheduled are nodes x_i's, and all the aircraft to be re-routed are nodes y_j's. If flight x_i lands at the station that x_j starts from, there is a directed arc pointing from x_i to x_j. Also, all the aircraft nodes are connected to all the flight nodes with undirected arcs. A rescheduled routing for the entire planning horizon is represented by a chain

$$q = (x_a, x_b), (x_b, x_c), ..., (x_r, y_1), (y_1, x_h), ..., (x_q, y_j),$$

such that $x_a, x_b, ..., x_r$ is the sequence of flights scheduled to be flown by aircraft y_1, and so on and so forth. The authors supposed that the shortage of aircraft would occur at the beginning of the day, so that a flight's delay cost, which is proportional to the total passenger number multiplied by the amount of the delay, could be calculated by comparing its position in the sequence of the aircraft it is assigned to and the scheduled departure time, and could be expressed as the length of the arc between two flight nodes. The model seeks to minimize the total cost caused by all the rescheduled flights. The problem is transformed into finding the shortest path of the graph that traverses all of its nodes. The authors apply branch-and-bound methodology to solve it. At each branching node, the partial sum of total arc length is a straightforward lower bound for the partial setting – a partial sequence of nodes for q. If the partial sum is greater than the most recent upper bound, i.e., the shortest total length ever obtained in the process from a complete path, then all the paths with the first part identical to the partial setting are abandoned and thus a cut is made. The authors gave numerical examples to illustrate the effectiveness of the model and solution strategy. No practical problem was tested.

Teodorovic and Stojkovic [144] proposed a lexicographical model meant to minimize the number of cancellations first and, at the level of minimum cancellations, minimize total passenger delays. The model takes into account airport curfew times. The authors devised a heuristic that sequentially decides the chain of flights to be flown by each aircraft. For each aircraft i, the chain is allowed to grow in a multistage network, which is comprised of one stage 0 of a starting node, one stage 1 of nodes representing flights departing from the station where the aircraft is, and stages 2 to n of nodes representing the n flights left unassigned to aircraft $1, ..., i-1$, and directed arcs from nodes in stage i to nodes in stage $i+1$, each of which reflects that the flight in stage $i+1$ departs from the airport where the flight in stage i lands. A chain is a sequence of connected nodes and arcs extending from stage 0 to the stage where it can grow no more because the station-curfew constraint represents a feasible assignment of the aircraft. The heuristic lets the chain grow stage by stage. When the chain up to stage $i-1$ is found, the arc (and thus the node) in stage i to be added to the chain is chosen such that the incurred delay taking into consideration the part of sequence of flights already decided is minimized. After the chain for each aircraft is found, the flights left unassigned are canceled. The model again assumes a common starting time of aircraft operation time. In addition, the aircraft balance problem essential to maintenance is not addressed.

Jarrah et al.'s [73] models use network flows to trace the effect of aircraft shortage. In one of their models dealing only with flight delays, a network was built for the station where there is a shortage of aircraft. There are aircraft nodes at the times of scheduled availability and flight nodes at the times of scheduled takeoff. Also, there are supply nodes and recovery nodes representing spare aircraft and aircraft recovered from previous unavailability at the times of their readiness. A directed forward arc is linked from each aircraft node to its assigned flight node on the schedule; and a backward arc is linked from each flight node to every aircraft node, every supply node, and every recovery node that is early enough so that the corresponding delay is still acceptable. On each backward arc, there is a delay cost incurred by delay of the potential assignment. Sources of unit flows are provided at the aircraft nodes where the aircraft are unavailable; supply and recovery nodes are treated as sinks where no more than unit flow can be absorbed. In a path of unit flow from a source node to a sink node, the backward arcs represent the reassignment of flights to aircraft, and the sum of their costs corresponds to the marginal operation cost. The minimum-cost network flow algorithm helps find the optimal reassignments.

In Jarrah et al.'s model dealing with flight cancellations, the aforementioned networks for all the stations involved become parts of a complete network. These parts are linked by additional arcs from flight nodes at their departure stations to aircraft nodes corresponding to the assigned aircraft on the schedule at the arrival stations. Cancellation costs are put onto these arcs. Because delays due to upstream delays are not captured in this model, the backward arcs considered in this model do not represent real delays. In fact, they always go from flights to aircraft that are ready before their scheduled takeoff times. In a path of unit flow, the backward arcs still represent reassignments, or aircraft swaps, while the additional arcs represent flight cancellations. Again, a minimum-cost network flow algorithm serves the purpose of making the optimal decision. The models were tested on real data from United Airlines. They generated effective solutions fast, and, therefore are, suitable for real-time implementation. The values and impact generated from the system and some of its implementation issues were reported in Rakshit et al.[114]. The drawback of the models is that they only consider limited alternatives for tackling the aircraft shortage problem. Real delays and cancellations are not considered simultaneously and the rippling effects of delays are not studied.

Arguello et al. [8] proposed models that deal with the problem more thoroughly and give approximate solution techniques. The authors proposed two exact models for the problem. The first one assumes that all the feasible flight paths that can be flown by one aircraft continuously are generated, and the cost of assigning every aircraft to every feasible flight path is calculated. Also known is each flight's cancellation cost. Feasibility means that a flight path observes the constraints of minimum turnaround time, station curfew requirements, and path continuity requirements. The model finds the least cost assignment of every aircraft to a feasible path so that each flight is either covered by exactly one flight path or canceled, and the aircraft balance is preserved. The number of feasible flight paths is of exponential order; thus, the model is intractable. The authors did not attempt to solve it. The second model considers aircraft and cancellations as two types of commodities that flow in a network. Every flight's departure or arrival corresponds to a node in the network. Departure nodes point arcs to their corresponding arrival nodes. For each station, there is an aircraft source node whose supply is the number of aircraft available at the location, a cancellation source node, and a station sink node whose demand is the number of required overnight aircraft. All the source nodes and arrival nodes at one station point arcs to the station's sink and departure nodes. The model makes the approximation that delay costs are stored on arcs that enter departure nodes and are accrued for an aircraft's assignment. The cancellation cost for a flight is allocated to the arc representing that flight. The problem becomes a two-commodity minimum-cost binary flow problem with homologous arcs. The homologous arcs are the flight arcs that must have flows of one. This is an NP-hard problem as well.

The third model Arguello et al. proposed is an approximate one based on the time-space network. The time horizon is discretized to have a finite number of bands. For every station there are a number of station-time nodes, each in the time band when there can be arrival or departure events. For every flight departing from a station, there is an arc going from each station-time node that belongs to the station-time node that corresponds to the flight's arrival time and station, if the flight implied by the arc satisfies the curfew requirement. If it does not, the arc goes to the station-sink node designated for the departing station. So for each flight, there are a variety of arcs that represent it, and each is executed at a different time and has a different delay cost. The problem can again be thought of as a minimum cost network flow problem with side constraints for flight coverage. We make the following definitions before we present the formulation:

Indices:

$i, j =$ node indices

$k =$ flight index

Sets:

$F =$ set of flights

$G(i) =$ set of flights originating at station-time node i

$H(k, i) =$ set of destination station-time nodes for flight k that originates from station-time node i

$I =$ set of station-time nodes

$J =$ set of station-sink nodes

$L(i) =$ set of flights terminating at node i

$M(k, i) =$ set of origination station-time nodes for flight k terminating at node i

$P(k) =$ set of station-time nodes at the station from which flight k originates

$Q(i) =$ set of station-time nodes at station that contains station-sink node i

Parameters:

$a_i =$ number of aircraft that become available at station-time node i

$c_k =$ cost of canceling flight k

$d_{ij}^k =$ delay cost of flight k from station-time node i to station-time node j

$h_i =$ number of aircraft required to terminate at station-sink node i

Variables:

$x_{ij}^k =$ amount of aircraft flow for flight k from station-time node i to node j

$y_k =$ cancellation indicator for flight k

$z_i =$ amount of aircraft flow from station-time node i to station-sink node at same station

With the above definitions, the formulation is written as:

$$\min \sum_{k \in F} \sum_{i \in P(k)} \sum_{j \in H(k,i)} d_{ij}^k x_{ij}^k + \sum_{k \in F} c_k y_k$$

subject to

$$\sum_{i\in P(k)}\sum_{j\in H(k,i)} x_{ij}^k + y_k = 1 \qquad \forall k \in F$$

$$\sum_{k\in G(i)}\sum_{j\in H(k,i)} x_{ij}^k + z_i = \sum_{k\in L(i)}\sum_{j\in M(k,i)} x_{ij}^k + a_i \qquad \forall i \in I$$

$$\sum_{k\in L(i)}\sum_{j\in M(k,i)} x_{ji}^k + \sum_{j\in Q(i)} z_j = h_i \qquad \forall i \in J$$

$$x_{ij}^k \in \{0,1\} \qquad \forall k \in F, i \in I, j \in H(k,i)$$

$$y_k \in \{0,1\} \qquad \forall k \in F$$

$$z_i \in Z_+ = \{0,1,2,...\} \qquad \forall i \in I.$$

In the model, the first set of constraints specifies that a flight is either carried out at certain time or canceled. The second set balances inflows and outflows of aircraft at all station-time nodes. The third set enforces aircraft balance at the end of day at all the stations. The rest of the constraints define the decision variables' domains.

Due to the discretization, all the aircraft are assumed to be able to service a flight at the starting time of the band where its available time is. d_{ij}^k underestimates the real delay. Therefore the solution of the model is a lower bound of the actual optimal solution. The linear relaxation of the above model will give a lower bound to the model itself, thus a lower bound to the optimal value.

Arguello et al. [9] adopted a greedy randomized adaptive search procedure (GRASP) to get a feasible suboptimal solution to the problem. The scheme of GRASP is: Starting with an incumbent solution, study all its neighboring solutions and store the ones considered good compared to the incumbent on the restricted candidate list. In the next iteration, randomly pick one solution from the restricted candidate list as the new incumbent. The procedure continues until the stopping criterion is met. In the GRASP for this problem, the initial incumbent solution is simply the one that cancels all the subsequent flights of the aircraft that is causing the shortages. All the neighbors are found by doing certain operations to all the pairs of two flight paths and flight-cancellation paths in the incumbent solution. The operations done on a pair append a part of one path to another or exchange portions of paths between each other, so that the aircraft balance is preserved. Delay cost can be estimated on each path, and curfew time violations can be found. Finding all the neighboring solutions in one iteration needs polynomial time. The procedure does not involve spare aircraft when starting from the default initial incumbent solution.

Wei et al. [157] [132] addressed the real-time crew recovery problem. The crew schedule disruption problem is often caused by delays, cancellations, diversions, crew sickness, missed connections, and/or legality conditions. The alternatives for resolving these problems include swapping crews, using crew reserves, deadheading crews, and any combination of these. Crew legality, seniority, qualification, pay protection, and returning to base after service are the main issues that need to be taken into account. A network model is constructed for the crew pairing repair problem. The main components of the network include:

- Crew nodes: The crew nodes represent either arriving crew or crew originating from the problem station. They are placed at the time of availability.
- Flight nodes: The flight nodes represent departure flights, and they are placed at the scheduled time of departure.
- Reserve nodes: The reserve nodes represent the availability of crew reserves, and they are placed at their available station and time.
- Return nodes: These nodes are used to force the crews to return to their original schedule after the recovery time.
- Scheduled arcs: These arcs emanate from crew nodes to their originally scheduled flight nodes to represent the original schedule.
- Swap arcs: These arcs emanate from crew nodes to flight nodes which are not their originally assigned flight and whose departure time is later than the crew availability time.
- Flight arcs: These arcs represent the flight from one airport to another. They originate from flight nodes at departure airports and end at the corresponding crew nodes at their destination airports.

- Reserve arcs: These arcs emanate from reserve nodes to those flight nodes at the same airport which can be served by the reserve crew.
- Return arcs: These arcs emanate from crew nodes at the airport where their corresponding return nodes are placed to their return nodes.

Costs are assigned to the arcs to account for crew swap, deadhead, use of reserves, etc. Based on the network, a multi-commodity integer network flow model and a heuristic search algorithm were presented and discussed. Their computational example showed that the model and algorithm are effective in solving the crew recovery problem.

A quadratic 0-1 programming model simultaneously dealing with flight delays and cancellations during irregular operations was given by Cao and Kanafani [29][30]. A solution algorithm was provided and tested to be effective and efficient enough to carry out real time missions. Yan and Yang [160] formulate the schedule perturbation problem caused by aircraft shortage as network flow problems and network flow problem with side constraints, while Yan and Lin [159] formulated the schedule perturbation problem caused by temporary airport closure as network flow problems and network flow problems with side constraints. In both applications, the network flow problems are solved by network simplex methods and the network flow problems with side constraints are solved by applying subgradient methods on the problems' Lagrangian relaxations. To fit the fleet assignment into the fluctuating passenger demand pattern, Klincewicz and Rosenwein [75] also formulated and solve a network flow problem. They treated a daily schedule varying from day to day as the combination of a repetitive "skeleton" schedule and daily changes.

Mathaisel [95] reported on the development of the Airline Schedule Control (ASC) system which provides a systematic interaction between the system and humans, centralizes the database across all functions of the airline, and distributes the decision support and optimization to the respective groups. The system was built on a network of workstations. One of them works as a server, and all the others work as clients. The common database resides on the server, and all the schedule retrievals and updates are through the server. The ASC display on each client workstation is controlled by a local copy. Human controllers interact with the client workstations to accomplish the various decision making processes. The optimization engine integrated into the ASC environment for irregular operations is a space-time network-based model and the out-of-kilter network flow algorithm. The space-time network consists of nodes representing the arrival and departure times and the station of flights, arcs linking the origin and destination nodes of flights, ground arcs that link all the nodes in one station, and overnight (for a longer schedule cycle, it could be over weekend) arcs that for each station link the last node in the time horizon to the first node. When perturbation occurs, alternative options are evaluated, and the most cost-effective action is chosen by adding arcs and changing flow bounds that reflect delaying or canceling of flights, designating to them the estimated cost of carrying them out, and solving the minimum cost network flow problem by the out-of-kilter method. The determination of the arc costs depends on many factors, including flight duration, amount of delaying, the number of passengers on board, etc. The ASC system provides a framework on which continuing independent expansions and improvements on sub-problem solving for dealing with various operational problems including irregular operations are possible without inconveniencing human controllers with disparate and ever-changing interfaces.

In [162], Yu addressed various issues encountered in implementing real-time, mission-critical decision support systems for aircraft and crew recovery. The following issues were discussed:

- Solution time is important due to the real-time nature of the irregular operations control problem. The complexity of the underline optimization problem demands simplification of the problem and effective heuristics.

- Keeping the optimization model always up-to-the-minute is a challenging task. The real-time data needs to be received and processed on-line. This issue was resolved by keeping the model in memory and using a messaging system to update the model in real-time.

- Coping with crew legalities requires embedding some basic legality checking in the optimization engine and complete checking when solutions are constructed in the search engine.

- In a multiple user environment, who is responsible for which resource usage is a practical issue. In such situations, a resource locking mechanism needs to be provided so that the resources engaged by one user are not used by a different user, thus avoiding conflict.

- Multiple solutions are desirable to support the decision makers' choice. This is particularly important since there are many soft constraints that are difficult to incorporated into the optimization engine.

- Partial solutions are also desirable in order to resolve immediate problems and postpone unavoidable resource shortages to a later stage.

- What-if capabilities are a must for successful implementation of such systems. These enable users to solve immediate as well as anticipated problems.

- Since a real-time control system needs to interface with many existing systems, the network communication and database integrity problem should not be overlooked.

- To facilitate users' acceptance, a user-friendly graphical interface is of great importance.

8 CONCLUDING REMARKS

In this paper, we intended to cover some major lines of optimization research and applications in the airline industry. Due to the technology-driven nature of the airline industry, optimization can have a great impact on almost every part of the operational process ranging from planning, routing, and scheduling to real-time control of all resources involved. Our task is by no means complete. There are still many topics. To name but a few: they include maintenance scheduling and routing, manpower planning, crew training scheduling, gate assignment, aircraft load balance, baggage routing and tracking, airport facility management, aircraft procurement, aircraft parts inventory management, and food service and supply management. Some of these can be found in Yu [164] and in Yu [163].

Currently, most airlines store data, make plans, generate schedules, manage resources, and control operations using fragmented systems. These systems do not communicate seamlessly with each other. Complete real-time data is lacking. Multiple data instances reside in different systems which leads to pitfalls in data integrity and synchronization. Graphical user interfaces do not have the same look and feel across systems. Many managers still rely on computer printouts and manually-generated charts to make their planning and daily operational decisions.

More importantly, even though some companies deploy isolated decision support systems, these systems generate sub-optimal, localized, and uncoordinated solutions. For example, the system which generates aircraft routes does not take into account the difficulty in compiling corresponding crew pairings; the system which generates flight schedules disregards the complexity of its consequent aircraft routes; the system for manpower planning decisions does not consider subsequent training scheduling and training resource usage; the systems for planning decisions do not offer robust solutions that can be effectively recovered and remedied in the case of schedule disruptions during plan execution; etc. The lack of integration among decision support systems leads to inferior solutions in terms of enterprise-wide cost and responsiveness to change. We would like to see more research, development, and implementation effort on integrated decision support systems.

Reaching and staying at the leading edge in the competitive air transportation market, managing operations efficiently, and responding to customer needs effectively are among the challenges facing top management at every commercial airline. The key to meeting these challenges is the successful deployment of sophisticated and integrated, optimization-based decision support systems utilizing state-of-the-art computer and optimization technology. As successfully implemented decision support systems at several major airlines start to demonstrate their tremendous value and impact, we anticipate an overwhelming acceptance of the optimization concept and a prosper future for optimization applications in the airline industry.

References

[1] Abara, J., "Applying Integer Linear Programming to the Fleet Assignment Problem," *Interfaces*, 19:4, 1989, pp. 20-28.

[2] Alstrup, J., Boas, S., Madsen, O.B.G., and Vidal, R.V.V., "Booking Policy for Flights with Two Types of Passengers," *European Journal of Operational Research*, 27, 1986, pp. 274-288.

[3] Anbil, R., Gelman, E., Patty, B., and Tanga, R., "Recent Advances in Crew-Pairing Optimization at American Airlines," *Interfaces*, 21:1, 1991, pp. 62-74.

[4] Anbil, R., Tanga, R., and Johnson, E.L., "A Global Approach to Crew-Pairing Optimization," *IBM Systems Journal*, 31:1, 1992, pp. 71-78.

[5] Andreatta, G. and Romanin-Jacur, G., "Aircraft Flow Management under Congestion," *Transportation Science*, 21:4, 1987, pp. 249-253.

[6] Andreussi. A., Bianco, L., and Ricciar-Delli, S., "A Simulation Model for Aircraft Sequencing in the Near Terminal Area," *European Journal of Operational Research*, 8, 1981, pp. 345-354.

[7] Arabeyre, J.P., Fearnley, J., Steiger, F.C., and Teather, W., "The Airline Crew Scheduling Problem: A Survey," *Transportation Science*, 3, 1969, pp. 140-163.

[8] Arguello, M.F., Bard, J.F., and Yu, G., "Models and Methods for Managing Airline Irregular Operations Aircraft Routing," in G. Yu eds. *Operations Research in the Airline Industry*, 1997, pp. 1-45.

[9] Arguello, M.F., Bard, J.F., and Yu, G., "Bounding Procedures and a GRASP Heuristic for the Aircraft Routing Problem," *Journal of Combinatorial Optimization*, 1, 3, 1997, pp. 211-228.

[10] Aykin, T., "On the Location of Hub Facilities," *Transportation Science*, 22:2, 1988, pp. 155-157.

[11] Aykin, T., "Networking Policies for the Hub-and-Spoke Systems with Application to the Air Transportation System," *Transportation Science,* 29, 1995, pp. 201-221.

[12] Balas, E. and Ho, A., "Set Covering Algorithms Using Cutting Planes, Heuristics and Subgradient Optimization: A Computational Study," *Mathematical Program Study*, 12, 1980, pp. 37-60.

[13] Balas, E. and Padberg, M.W., "Set Partitioning: A Survey," *SIAM Review*, 18:4, 1976, pp. 710-760.

[14] Bailey, E.E., Graham, D.R., and Kaplan, P.D., "Deregulating the Airlines," The MIT Press, Cambridge, MA, 1985.

[15] Ball, M., Bodin, L., and Dial, R., "A Matching Based Heuristic for Scheduling Mass Transit Crews and Vehicles," *Transportation Science*, 17, 1983, pp. 4-31.

[16] Ball, M. and Roberts, A., "A Graph Partitioning Approach to Airline Crew Scheduling," *Transportation Science*, 19:2, 1985, pp. 107-126.

[17] Bauer, P.W., "Airline Hubs: A Study of Determining Factors and Effects," *Economic Review: Federal Reserve Bank of Cleveland*, Forth Quarter, 1987, pp. 13-19.

[18] Beckmann, M.J., "Decision and Team Problems in Airline Reservations," *Econometrica*, 26, 1958, pp. 134-145.

[19] Belobaba, P.P., "Airline Yield Management An Overview of Seat Inventory Control," *Transportation Science*, 21:2, 1987, pp. 63-73.

[20] Belobaba, P.P., "Application of A Probabilistic Decision Model to Airline Seat Management Control," *Operations Research* , 37, 1989, pp. 183-197.

[21] Benchakroun, A., Ferland, J.A., and Cleeroux, R., "Distribution System Planning Through a Generalized Benders Decomposition Approach," *European Journal of Operational Research*, 62, 1991, pp. 149-162.

[22] Benders, J.F., "Partitioning procedure for solving mixed-variables programming problem," *Rand Symp. Math. Programming*, March 16-20, 1959, Santa Monica, California.

[23] Berge, M.A. and Hopperstad, C.A., "Demand Driven Dispatch: A Method for Dynamic Aircraft Capacity Assignment, Models and Algorithms," *Operations Research*, 41, 1993, pp. 153-168.

[24] Bixby, R.E., "Implementing the Simplex Methods, Part I, Introduction; Part II, The Initial Basis," *TR 90-32 Mathematical Science*, Rice University, Houston, TX, 1990.

[25] Blumstein, A., "An Analytical Investigation of Airport Capacity," Cornell Aeronautical Laboratory, Report TA-1358-6-1, June 1960.

[26] Brown, J.H., "An Economic Model of Airline Hubbing-and-Spoking," *Logistics and Transportation Review*, 27:3, 1991, pp. 225-239.

[27] Brumelle, S.L., McGill, J.I., Oum, T.H., Sawaki, K., and Tretheway, M.W., "Allocation of Airline Seats between Stochastically Dependent Demands," *Transportation Science*, 24:3, 1990, pp. 183-192.

[28] Camerini, P.K., Fratta, L., and Maffioli, F., "On Improving Relaxation Methods by Modified Gradient Techniques," *Mathematical Programming Studies*, 3, 1975, pp. 6-25.

[29] Cao, J. and Kanafani, A., "Real-Time Decision Support for Integration of Airline Flight Cancellations and Delays Part I: Mathematical Formulation," *Transportation Planning and Technology*, 20, 1997, pp. 183-199.

[30] Cao, J. and Kanafani, A., "Real-Time Decision Support for Integration of Airline Flight Cancellations and Delays Part II: Algorithm and Computational Experiments," *Transportation Planning and Technology*, 20, 1997, pp. 201-217.

[31] Christofides, N. and Korman, S., "A Computational Survey of Methods for the Set Covering Problem," *Management Science*, 21:5, 1975, pp. 591-599.

[32] Conley, W., *Computer Optimization Techniques*, New York, Petrocelli Books, 1980.

[33] Cornuejols, G., Fisher, M., and Nemhauser, G., "Location of Bank Accounts to Optimize Float: An Analytic Study of Exact and Approximate Algorithms," *Management Science*, 23:6, 1977, pp. 789-810.

[34] Crowder, H., Johnson, E.L., and Padberg, M.W., "Solving Large Scale Zero-one Linear Programming Problems," *Operations Research*, 31, 1983, pp. 803-834.

[35] Curry, R.E., "Optimal Airline Seat Allocation with Fare Classes Nested by Origins and Destinations," *Transportation Science*, 24:3, 1990, pp. 193-204.

[36] Dantzig, G.B., *Linear Programming and Extension*, Princeton University Press, Princeton, 1963.

[37] Daskin, M.S. and Panayotopoulos, N.D., "A Lagrangian Relaxation Approach to Assigning Aircraft to Routes in Hub and Spoke Networks," *Transportation Science*, 23:2, 1989, pp. 91-99.

[38] Desaulniers, G., Desrosiers, J., Dumas, Y., Solomon, M.M., and Soumis, F., "Daily Aircraft Routing and Scheduling," *Management Science*, 43:6, 1997, pp. 841-855.

[39] Druckerman, J., Silverman, D., and Viaropulos, K., "IBM Optimization Subroutine Library, Guide and Reference, Release 2," Document Number SC230519-02, IBM, Kinston, New York, 1991.

[40] Du, D.-Z. and Hwang, F.K., *Combinatorial Group Testing and Its Applications*, World Scientific Corp., Inc., 1993.

[41] Dunlay, J.W., "Analytical Models of Perceived Air Traffic Control Conflicts," *Transportation Science*, 9, 1987, pp. 149-164.

[42] Elce, I., "The Development and Implementation of Air Canada's Long Range Planning Model," *AGIFORS Symposium Proceedings*, 10, 1970.

[43] Etschmaier, M.M., "Schedule Construction and Evaluation for Short and Medium Range Corporate Planning," *AGIFORS Symposium Proceedings*, 10, 1970.

[44] Etschmaier, M.M., "A Survey of the Scheduling Methodology Used in Air Transportation," in R. Genser, M. Strobel, and M.M. Etschmaier (eds.) *Optimization Applied to Transportation Systems*, Vienna, IIASA, 1977.

[45] Etschmaier, M.M. and Mathaisel, D.F.X., "Airline Scheduling: An Overview," *Transportation Science*, 19:2, 1985, pp. 127-138.

[46] Etschmaier, M.M. and Rothstein, M., "Operations Research in the Management of the Airlines," OMEGA 2, 1974, pp. 157-179.

[47] Fisher, M.L., "The Lagrangian Relaxation Method for Solving Integer Programming Problems," *Management Science*, 27, 1981, pp. 1-18.

[48] Florian, M., Guerin, G.G., and Bushel, G., "The Engine Scheduling Problem on a Railway Network," *INFORJ.*, 14, 1976, pp. 121-128.

[49] Gagnon, G., "A Model for Flowing Passengers Over Airline Networks," *AGIFORS Symposium Proceedings*, 7, 1967.

[50] Garfinkel, R.S. and Nemhauser, C.L., "The Set Partitioning Problem: Set Covering with Equality Constraints," *Operations Research*, 17, 1969, pp. 848-856.

[51] Fitzpatrick, G.L. and Modlin, M.J., "*Direct-Line Distances, United States Edition*," The Scarecrow Press, 1986.

[52] Gasco, J.L., "Reservations and Booking Control," *AGIFORS Symposium Proceedings*, 17, 1977.

[53] Geisinger, K.E., "Airspace Conflict Equations," *Transportation Science*, 19:2, 1985, pp. 139-153.

[54] Geoffrion, A.M. and Graves, G.W., "Multicommodity Distribution System Design by Benders Decomposition," *Management Science*, 20:5, 1974, pp. 822-844.

[55] Gershkoff, I., "Optimizing Flight Crew Schedules," *Interfaces*, 19:4, 1989, pp. 29-43.

[56] Ghobrial, A. and Kanafani, A., "Airline Hubbing: Some Implications for Airport Economics," *Transportation Research*, 19A, 1985, pp. 15-27.

[57] Glover, F., Glover, R., Lorenzo, J., and McMillan, C., "The Passenger-Mix Problem in the Scheduled Airlines," *Interfaces* , 12, 1982, pp. 73-79.

[58] Gomory, R., "An Algorithm for Integer Solution to Linear Programs," in G. Graves and P. Wolfe (eds.) *Recent Advances in Mathematical Programming*, New York, McGraw-Hill, 1963.

[59] Graves,G., McBride, R., Gershkoff, I., Anderson, D., and Mahidhara, D., "Flight Crew Scheduling," *Management Science*, 39:6, 1993, pp. 736-745.

[60] Hane, C.A., Barnhart, C., Johnson, E.L., Marsten., R.E., Nemhauser, G.L., and Sigismondi, G., "The Fleet Assignment Problem: Solving a Large-Scale Integer Program," *Mathematical Programming*, 70, 1995, pp. 211-232.

[61] Hansen, M., "A Model of Airline Hub Competition," University of California, Berkeley, Institute of Transportation Studies, Dissertation Series 88-2, Berkeley CA.

[62] Hansen, M. and Kanafani, A., "International Airline Hubbing in a Competitive Environment," *Transportation Planning and Technology*, 13, 1988, pp. 3-18.

[63] Hansen, M. and Kanafani, A., "Airline Hubbing and Airport Economics in the Pacific Market," *Transportation Research*, 24A:3, 1990, pp. 217-239.

[64] Helme, M.P., "A Selective Multicommodity Network Flow Algorithm for Air Traffic Control," in G. Yu eds. *Operations Research in the Airline Industry*, 1997, pp. 101-121.

[65] Hersh, M. and Ladany, S.P., "Optimal Seat Allocation for Flights with One Intermediate Stop," *Computers and Operations Research*, 5, 1978, pp. 31-37.

[66] Hoang, H.H., "Topological Optimization of Networks: A Nonlinear Mixed Integer Model Employing Generalized Benders Decomposition," *IEEE Trans. Automatic Control*, AC-27, 1982, pp. 164-169.

[67] Hockaday, S.L.M. and Kanafani, A.K., "Developments in Airport Capacity Analysis," *Transportation Research* , 6, 1974, pp. 171-180.

[68] Hoffman, K.L. and Padberg, M.W., "Solving Airline Crew Scheduling Problems by Branch-and-Cut," *Management Science* , 39:6, 1993, pp. 657-682.

[69] Holloway, C., "A Generalized Approach to Dantzig-Wolfe Decomposition for Concave Programs," *Operations Research*, 21, 1973, pp. 210-220.

[70] Jaillet, P., Song, G., and Yu, G., "Airline Network Design and Hub Location Problems," *Location Science*, 4, 3, 1997, pp. 195-212.

[71] Jaillet, P., Song, G., and Yu, G., "Networking Design Problems with Applications to the Airline Industry," *Proceedings for TRISTAN II*, Capri, Italy, 1994.

[72] Janic, M. and Tosic, V., "En Route Sector Capacity Model," *Transportation Science*, 25:4, 1991, pp. 299-307.

[73] Jarrah, A.I.Z., Yu, G., Krishnamurthy, N., and Rakshit, A., "A Decision Support Framework for Airline Flight Cancellations and Delays," *Transportation Science*, 27:3, 1993, pp. 266-280.

[74] Kanafani, A., "Transportation Demand Analysis," McGraw-Hill, New York, 1983, pp. 256-258.

[75] Klincewicz, J.G. and Rosenwein, M.B., "The Airline Exception Scheduling Problem," *Transportation Science*, 29:1, 1995, pp. 4-16.

[76] Kushige, T., "A Solution of Most Profitable Aircraft Routing," *AGIFORS Symposium Proceedings*, 3, 1963.

[77] Labombarda, P. and Nicoletti, B., "Aircraft Rotations by Computer," *AGIFORS Symposium Proceedings*, 11, 1971.

[78] Ladany, S.P. and Bedi, D.N., "Dynamic Booking Rules for Flights with an Intermediate Stop," OMEGA, 5:6, 1977, pp. 721-730.

[79] Ladany, S.P. and Hersh, M., "Non-Stop vs. One-Stop Flights," *Transportation Research*, 11:3, 1977, pp. 155-159.

[80] Land, A. H. and Doig, A., "An Automatic Method of Solving Discrete Programming Problems," *Economics*, 28, 1960, pp. 497-520.

[81] Lavoie, S., Minoux, M., and Odier, E., "A New Approach for Crew Pairing Problems by Column Generation with an Application to Air Transportation," *European Journal of Operations Research*, 35, 1988, pp. 45-58.

[82] Lawler, E.L., Lenstra, J.K., and Rinnooy Kan, A.H.G., "Recent Developments in Deterministic Sequencing and Scheduling: A Survey," in M.A.H. Dempster, J.K. Lenstra, and A.H.G. Rinnooy Kan (eds.) *Deterministic and Stochastic Scheduling*, The Netherlands, Reidel/Kluwer Dordrecht, 1982.

[83] Levin, A., "Scheduling and Fleet Routing Models for Transportation Systems," *Transportation Systems*, 5, 1971, pp. 232-255.

[84] Littlewood, K., "Forecasting and Control of Passenger Bookings," *AGIFORS Symposium Proceedings*, 12, 1972, pp. 95-117.

[85] "Yield Managers Now Control Tactical Marketing," *Lloyd's Aviation Economist*, 12-13, 1985.

[86] Loughran, B.P., "An Airline Schedule Construction Model," *AGIFORS Symposium Proceedings*, 12, 1972.

[87] Luo, S. and Yu, G., "Airline Schedule Perturbation Problem: Landing and Takeoff With Nonsplittable Resource for the Ground Delay

Program," in G. Yu eds. *Operations Research in the Airline Industry*, 1997, pp. 404-431.

[88] Luo, S. and Yu, G., "On the Airline Schedule Perturbation Problem Caused by the Ground Delay Program," *Transportation Science*, 31, 4, 1997, pp. 298-311.

[89] Lustig, I.J., Marsten, R.E., and Shannon, D.F., "Computational Experience with a Primal-Dual Interior Point Method for Linear Programming," *Linear Algebra and Its Applications*, 152, 1991, pp. 191-222.

[90] Magnanti, T.L., Mireault, P., and Wong, R.T., "Tailoring Benders Decomposition for Uncapacitated Network Design," *Mathematical Programming Study*, 26, 1986, pp. 112-154.

[91] Magnanti, T.L. and Wong, R.T., "Accelerating Benders Decomposition: Algorithmic Enhancement and Model Selection," *Operations Research*, 29, 1981, pp. 465-484.

[92] *Marketing News*, June 20, 1986.

[93] Marsten, R.E., "An Algorithm for Large Set Partitioning Problems," *Management Science*, 20:5, 1974, pp. 774-787.

[94] Marsten, R.E. and Shepardson, F., "Exact Solution of Crew Problems using the Set Partitioning Mode: Recent Successful Applications," *Networks*, 11, 1981, pp. 165-177.

[95] Mathaisel D.F.X., "Decision Support for Airline System Operations Control and Irregular Operations," *Computers and Operations Research*, 23:11, 1996, pp. 1083-1098.

[96] Mayer, M., "Seat Allocation, or a Simple Model of Seat Allocation via Sophisticated Ones," *AGIFORS Symposium Proceedings*, 16, 1976.

[97] McShan, S. and Windle, R., "The Implication of Hub-and-Spoke Routing for Airline Costs and Competitiveness," *Logistics and Transportation Review*, 25:3, 1989, pp. 209-229.

[98] Mevert, P., "Fixed Charge Network Flow Problems: Applications and Methods of Solution," Presented at *Large Scale and Hierarchical Systems Workshop*, May 1977, Brussels.

[99] Miller, R., "An Optimization Model for Transportation Planning," *Transportation Research*, 1, 1967, pp. 271-286.

[100] Mirchandani, P., "Polyhedral Structure of the Capacitated Network Design Problem with an Application to the Telecommunication Industry," Unpublished Ph.D dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1989.

[101] Moore, J.M., "An n Job, One Machine Sequencing Algorithm for Minimizing the Number of Late Jobs," *Management Science*, 15:1, 1968, pp. 102-109.

[102] Morrison, S. and Winston, C., "The Economic Effects of Airline Deregulation," Washington, D.C., Brookings Institution, 1986.

[103] Nakayama, H. and Sawaragi, Y., "Satisfying Tradeoff Method for Multiobjective Programming," in M. Graver and A.P. Wierzbicki (eds.) *Interactive Decision Analysis*, (Proceedings, Laxenburg, Australia), Berlin, Springer-Verlag, 1984, pp. 113-122.

[104] Nemhauster, G.L. and Widhelm,W.B., "A Modified Linear Program for Columnar Methods in Mathematical Programming," *Operations Research*, 19, 1971, pp. 1051-1060.

[105] Newell, G.F., "Airport Capacity and Delays," *Transportation Science*, 13:3, 1979, pp. 201-241.

[106] Odoni, A.R., "The Flow Management Problem in Air Traffic Control," in A.R. Odoni, L. Bianco, and G. Szego (eds.) *Flow Control of Congested Networks*, Berlin, Springer-Verlag, 1987.

[107] O'Kelly M.E., "The Location of Interacting Hub Facilities," *Transportation Science*, 20:2, 1986, pp. 92-106.

[108] O'Kelly M.E., "A Quadratic Integer Program for the Location of Interacting Hub Facilities," *European Journal of Operational Research*, 32, 1987, pp. 393-404.

[109] Orchard-Hays, W., "*Advanced Linear Programming Computing Techniques*," McGraw-Hill, New York, 1986.

[110] Ostresh, L., Rushton, G. and Goodchild, M.F., "TWAIN-Exact Solution to the Two Source Location-Allocation Problem," Computer

Programs for Location-Allocation Problem, University of Iowa, Monograph 6, 1973.

[111] Padberg, M.W., "On the Facial Structure of Set Packing Polyhedra," *Mathematical Programming*, 5, 1973, pp. 199-215.

[112] Padberg, M.W. and Rinaldi, G., "A Branch-and-Cut Algorithm for the Solution of Large-scale Traveling Salesman Problems," *SIAM Review*, 33, 1991, pp. 60-100.

[113] Powell, W.B., "Analysis of Airline Operating Strategies under Stochastic Demand," *Transportation Research*, 168, 1982, pp. 31-43.

[114] Rakshit, A., Krishnamurthy, N., and Yu, G., "A Real Time Decision Support System for Managing Airline Operations at United Airlines," *Interfaces*, 26, 2, 1996, pp. 50-58.

[115] Reynolds-Feighan, A.J., "*The Effects of Deregulation on U.S. Air Networks*," Berlin, Heidelberg, Springer-Verlag, 1992.

[116] Richardson, R., "An Optimization Approach to Routing Aircraft," *Transportation Research*, 13B, 1979, pp. 49-63.

[117] Richardson, R., "An Optimization Approach to Routing Aircraft," *Transportation Science*, 10, 1976, pp. 52-71.

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

[119] Richter, R.J., "Optimal Aircraft Rotations Based on Optimal Flight Timing," *AGIFORS Symposium Proceedings*, 8, 1968.

[120] Richter, R.J., "Experience with the Aircraft Rotation Model," *AGIFORS Symposium Proceedings*, 10, 1970.

[121] Rothstein, M., "An Airline Overbooking Model," *Transportation Science*, 5, 1971, pp. 180-192.

[122] Rothstein, M., "OR and Airline Overbooking Problem," *Operations Research*, 33:2, 1985, pp. 237-248.

[123] Rothstein, M. and Stone, A.W., "Passenger Booking Levels," *AGIFORS Symposium Proceedings*, 7, 1967.

[124] Rubin, J., “A Technique for the Solution of Massive Set Covering Problems, with Application to Airline Crew Scheduling,” *Transportation Science*, 7, 1973, pp. 34-48.

[125] Rue, R.C. and Rosenshine, M., “The Application of Semi-Markov Decision Processes to Queueing of Aircraft for Landing at an Airport,” *Transportation Science*, 19:2, 1985, pp. 154-172.

[126] Sassano, A., “On the Facial Structure of the Set Covering Polytope,” *Mathematical Programming*, 44, 1989, pp. 181-202.

[127] Siddiqee, W., “A Mathematical Model for Predicting the Number of Potential Conflict Situations at Intersecting Air Routes,” *Transportation Science*, 2, 1973, pp. 158-167.

[128] Simpson, R.W., “A Review of Scheduling and Routing Models for Airline Scheduling,” paper presented at the AGIFORS Meeting, Broadway, England, October 1969.

[129] Shlifer, E. and Vardi, Y., “An Airline Overbooking Policy,” *Transportation Science*, 9:2, 1975, pp. 101-114.

[130] Smith, B.C., Leimkuhler J.F., and Darrow R.M., “Yield Management at American Airlines,” *Interfaces*, 22:1-2, 1992, pp. 8-31.

[131] Song, G., “Integer Programming Models for Airline Network Design Problems,” Ph.D. dissertation, Graduate School of Business, The University of Texas at Austin, 1995.

[132] Song, G., Wei, G., and Yu, G., “A Decision Support Framework for Crew Management During Irregular Operations,” in G. Yu eds. *Operations Research in the Airline Industry*, 1997, pp. 259-286.

[133] Soudarovich, J., “Routing Selection and Aircraft Allocation,” *AGIFORS Symposium Proceedings*, 11, 1971.

[134] Struve, D.L., “Intercity Transportation Effectiveness Model,” *AGIFORS Symposium Proceedings*, 10, 1970.

[135] Subramanian, R., Scheff, R.P. Jr., Quillinan, J.D., Wiper, D.S., and Marsten, R.E., “Coldstart: Fleet Assignment at Delta Air Lines,” *Interfaces*, 24:1-2, 1994, pp. 104-120.

[136] Sun, X., Brauner, E., and Homby, S., "A Large-Scale Neural Network for Airline Forecasting in Revenue Management," in G. Yu (eds.) *Operations Research in the Airline Industry*, Boston, Kluwer Academic Publishers, 1997, pp. 46-65.

[137] Sussner, P., Pardalos, P.M., and Ritter, G.X., "On Integer Programming Approaches for Morphological Template Decomposition," *Journal of Combinatorial Optimization*, 1:2, 1997, pp. 177-188.

[138] Taha, H.A., *Integer Programming Theory, Applications, and Computations*, Academic Press, 1975, pp. 126-128.

[139] Talluri, K.T., "Swapping Applications in a Daily Airline Fleet Assignment," *Transportation Science*, 30:3, 1996, pp. 237-248.

[140] Taylor, C.J., "The Determination of Passenger Booking Levels," *AGIFORS Symposium Proceedings*, 2, 1962, pp. 93-116.

[141] Teodorovi, D., "Matching of Transportation, Capacities and Passenger Demands in Air Transportation," *Civil Engineering Practice*, Lancaster, Technomic Publishing, 1988, pp. 365-392.

[142] Teodorovic, D. and Guberinic, S., "Optimal Dispatching Strategy on an Airline Network After a Schedule Perturbation," *European Journal of Operational Research*, 15, 1984, pp. 178-182.

[143] Teodorovic, D. and Krcmar-Nozic, E., "Multicriteria Model to Determine Flight Frequencies on an Airline Network under Competitive Conditions," *Transportation Science*, 23:1, 1989, pp. 14-25.

[144] Teodorovic, D. and Stojkovic, G., "Model for Operational Daily Airline Scheduling," *Transportation Planning and Technology*, 14, 1990, pp. 273-285.

[145] Terrab, M., "Ground Holding Strategies for Air Traffic Control," Ph.D. thesis, Massachusetts Institute of Technology, February 1990.

[146] Terrab, M. and Odoni, A.R., "Strategic Flow Management for Air Traffic Control," *Operations Research*, 41:1, 1993, pp. 138-152.

[147] Tewinkel, D., "An Algorithm for Aircraft Scheduling in a Radial Network," *AGIFORS Symposium Proceedings*, 9, 1969.

[148] Thompson, H.R., "Statistical Problems in Airline Reservation Control," *Operational Research Quarterly*, 12, 1961, pp. 167-185.

[149] Titze, B. and Griesshaber, R., "Realistic Passenger Booking Behavior and Simple Low-Fare/High-Fare Seat Allotment Model," *AGIFORS Symposium Proceedings*, 23, 1983.

[150] Vasquez-Marquez, A., "American Airlines Arrival Slot Allocation System (ASAS)," *Interfaces*, 21:1, 1991, pp. 42-61.

[151] Verleger, P.K., "Model of the Demand for Air Transportation," *Bell J. Econ. Management Science*, 3:2, 1972.

[152] Vranas, P.B., Bertsimas, D.J., and Odoni, A.R., "The Multi-Airport Ground-Holding Problem in Air Traffic Control," *Operations Research*, 42:2, 1994, pp. 249-261.

[153] Vranas, P.B., Bertsimas, D.J., and Odoni, A.R., "Dynamic Ground-Holding Policies for a Network of Airports," *Transportation Science*, 28:4, 1994, pp. 275-291.

[154] Wang, K., "Optimum Seat Allocation for Multi-Leg Flights with Multiple Fare Types," *AGIFORS Symposium Proceedings*, 23, 1983, pp. 225-237.

[155] Wang, H., "A Dynamic Programming Framework for the Global Flow Control Problem in Air Traffic Management," *Transportation Science* , 25:4, 1991, pp. 308-313.

[156] Weatherford, L.R., "Using Prices More Realistically as Decision Variables in Perishable-Asset Revenue Management Problems," *Journal of Combinatorial Optimization*, 1, 3, 1997, pp. 277-304.

[157] Wei, G., Song, G., and Yu, G., "Model and Algorithm for Crew Management During Airline Irregular Operations," *Journal of Combinatorial Optimization*, 1, 3, 1997, pp. 305-321.

[158] Wollmer, R.D., "An Airline Seat Management Model for a Single Leg Route When Lower Fare Classes Book First," *Operations Research*, 40, 1992, pp. 26-37.

[159] Yan, S. and Lin, C., "Airline Scheduling for the Temporary Closure of Airports," *Transportation Science*, 31:1, 1997, pp. 72-82.

[160] Yan, S. and Yang, D., "A Decision Support Framework for Handling Scheduling Perturbation," *Transportation Research*, 30B:6, 1996, pp. 405-419.

[161] Yan, S. and Young, H., "A Decision Support Framework for Multi-Fleet Routing and Multi-Stop Flight Scheduling," *Transportation Research* 30A:5, 1996, pp. 379-398.

[162] Yu, G., "Real-Time, Mission-Critical Decision Support Systems for Managing and Controlling Airlines' Operations," *Proceedings at the International Conference on Management Science and Economic Development*, Hong Kong, 1996.

[163] Yu, G., *Operations Research in the Airline Industry*, Kluwer Academic Publishers, Boston, 1997.

[164] Yu, G. ed., "Recent Advances of Optimization Applications in the Airline Industry," Special Issue of *Journal of Combinatorial Optimization*, 1, 3, 1997.