TRANSPORTATION SCIENCE
Vol. 39, No. 3, August 2005, pp. 340–348
ISSN 0041-1655 | EISSN 1526-5447 | 05 | 3903 | 0340

DOI 10.1287/trsc.1040.0091

Airline Crew Scheduling Under Uncertainty

Andrew J. Schaefer
Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, schaefer@ie.pitt.edu

Ellis L. Johnson, Anton J. Kleywegt, George L. Nemhauser
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332
{ellis.johnson@isye.gatech.edu, anton.kleywegt@isye.gatech.edu, george.nemhauser@isye.gatech.edu}

Airline crew scheduling algorithms widely used in practice assume no disruptions. Because disruptions often occur, the actual cost of the resulting crew schedules is often greater. We consider algorithms for finding crew schedules that perform well in practice. The deterministic crew scheduling model is an approximation of crew scheduling under uncertainty with the assumption that all pairings will operate as planned. We seek better approximate solution methods for crew scheduling under uncertainty that still remain tractable. We give computational results from three fleets that indicate that the crew schedules obtained from our method perform better in a model of operations with disruptions than the crew schedules found via deterministic methods. Under mild assumptions we provide a lower bound on the cost of an optimal crew schedule in operations, and we demonstrate that some of the crew schedules found using our method perform very well relative to this lower bound.

Key words: airline planning; crew scheduling; recovery; schedule disruption; on-time performance
History: Received: May 2001; revisions received: April 2003, August 2003; accepted: August 2003.

1. Introduction

For major domestic carriers, crew costs are second only to fuel costs, and can exceed $1 billion annually. Hoffman and Padberg (1993) report that total pilot compensation exceeded $1.4 billion annually at the largest domestic airlines in the early 1990s, and a senior pilot earned up to $250,000 annually. These figures are larger now. Therefore, airlines devote great effort to planning good crew schedules. The planning problem can be very difficult to solve, however, because there are many governmental and contractual regulations concerning pilots, and problems found in practice often have billions of possible solutions.

There is currently a great deal of concern about air traffic congestion. In June 2000, flight delays were up over 16% from June 1999 (Phillips and Irwin 2000). Moreover, air traffic in the United States and Europe is expected to double in the next 10–15 years. If airport capacity remains constant, it is estimated that each 1% increase in airport traffic will bring about a 5% increase in delays (*The Economist*, 2000).

Disruptions are becoming more frequent and more severe. The Air Transport Association (ATA) reports that the average daily number of delays of more than 15 minutes tied to air traffic control increased to 2,149 in 1999, up from 1,807 in 1998 and 1,416 in 1997. The ATA also estimates that delays cost consumers and airlines about $5.2 billion in 1999, up from $5 billion the year before (Mathews 2000). The Federal Aviation Administration (FAA) reports a 58% increase in delays from 1995 to 1999, and a 68% increase in flight cancellations over the same period. Atlanta's Hartsfield International Airport had a 138% increase in flight cancellations over that period. The total cost to the airlines due to cancellations at Hartsfield alone was estimated at $250.9 million in 1999 (O'Dell 2000).

Recovery is the process of reacting to a disruption. Optimal recovery decisions are hard to determine. The future is uncertain, and cancelling a leg or rerouting a crew or a plane can have costly consequences throughout the airline's system. In practice, airlines make recovery decisions manually with little decision support (Lettovský 1997). This makes airline recovery difficult to model because airline decision makers use intuition and subjective judgment. Most optimization research done on airline operations has been on crew recovery, but these models assume all legs will be flown according to their new scheduled times. We are unaware of any research on dynamic and stochastic airline recovery models.

Crew planning affects airline operations. For example, airlines may delay flights if a crew is unavailable. Recently, pilots for some major domestic carriers refused to work overtime to protest the pace of negotiations of a new contract. The resulting shortage of pilots forced numerous cancellations.

Although airlines operate in a highly uncertain environment, very few airline planning models consider uncertainty in operations. Some of this is due to the structure of airline management. A plan is typically evaluated not by operational performance, but by the quality of the plan, assuming it may be implemented in operations. With the exception of yield management, we know of no airline planning models that measure the quality of a plan by its performance in operations with disruptions. The integration of airline planning and operations is a fertile area of great practical and theoretical interest.

We define two classes of airline disruptions based on the length of the disruption. A *frictional* disruption is of limited duration. Examples include delays due to connecting passengers, airport congestion, brief and unscheduled maintenance incidents, and localized, short, weather disruptions. The other class of disruptions is *severe*, which includes lengthy, unscheduled, maintenance disruptions and large-scale, severe, weather disruptions. This classification is not a strict dichotomy; disruptions may have aspects of both frictional and severe disruptions. We limit this study to planning and operations under frictional disruptions. We show that under frictional delays the operational performance of deterministic crew scheduling models can be improved.

1.1. Crew Scheduling

Typically, pilots may only fly one type of aircraft. Therefore the crew scheduling problem is separable by fleet type. When a crew is on duty, it flies a set of consecutive flight legs that follow certain regulations and contractual restrictions. Such a set of legs is called a *duty*. The *sit time* is the time between two consecutive legs within a duty. The number of minutes that elapse between the beginning of a duty and the end of the duty is the *elapsed time*. The elapsed time includes a briefing period before the first leg of the duty and a debriefing period after the last leg of the duty.

A *pairing* or crew trip is a set of duties. Consecutive duties must be separated by a rest period. A pairing must begin and end at a specified station; such stations are called *crew bases*. Pairings flown within the United States must adhere to certain FAA as well as contractual rules. For instance, one rule requires that a crew that flies more than 8 hours within a 24-hour period must receive compensatory rest (FAA 1999). The *time away from base* (TAFB) of a pairing is the number of minutes that elapse between the beginning of the pairing and the end of the pairing. In many instances, crews are paid based on the amount of time they fly in their pairing. However, there is a minimum guaranteed pay for any pairing, and there is additional compensation for the crew if the TAFB of the pairing or the elapsed time of one or more of the duties is large enough. We describe the details of calculating crew cost in §2.

Because a crew can fly only one fleet type, the fleet assignment problem and the aircraft rotation problem are typically solved before the crew scheduling problem. If a crew flies two consecutive legs on different planes, the scheduled connection time between these legs must exceed a minimum connection time. If the crew remains on the same plane for two consecutive flights, however, there is no minimum connection time.

A *crew schedule* is a set of pairings that partitions the legs to be flown by a single fleet. Crew scheduling problems are solved by generating pairings and solving an integer program. The daily crew scheduling problem is solved under the assumption that each leg is flown every day. The crew scheduling problem is usually modeled as a set partitioning problem

$$\{\min cx \colon Ax = 1,\ x \text{ binary}\}, \tag{1}$$

where a_{ij}, the ijth entry of the matrix A, is 1 if pairing j flies leg i, and 0 otherwise.

There may be a large number of pairings for a relatively small fleet. Larger fleets have billions of legal pairings. The enormous number of pairings is a major difficulty in solving airline crew scheduling problems exactly.

There has been a great deal of research on deterministic airline crew scheduling. A recent and comprehensive survey is provided in Barnhart et al. (2002).

Yen (2000) also considers the problem of crew scheduling under uncertainty. She formulates the problem as a two-stage stochastic program, where the first stage is the crew scheduling problem and the second stage involves penalties for delays rather than operational costs. She provides computational results for small problems.

Ehrgott and Ryan (2002) seek pairings that have a low planning cost and are "robust" according to several criteria. These criteria include sufficient time between consecutive legs in each duty or pairing. These criteria are weighted according to some predetermined arrangement, and an additive penalty function is constructed. Ehrgott and Ryan investigate several multicriteria optimization frameworks, including finding Pareto-optimal solutions, weighting the two objectives of planning cost and a chosen robustness criterion, and maximizing robustness subject to the planned cost remaining below a threshold. Their method produces crew schedules that are more robust than deterministic solutions. They are restricted to additive penalty methods so that column generation techniques may be employed.

1.2. Contributions

This paper makes several contributions to the integration of airline planning and operations with disruptions.

- It introduces the problem of designing airline crew schedules that perform well in operations with disruptions.

- It describes an easily implemented procedure for finding an approximate solution to the problem of minimizing expected crew costs.
- It provides a lower bound on the expected cost of any crew schedule in operations with disruptions.
- It introduces a measure for evaluating the performance of crew schedules in practice and shows that, in terms of this measure, solutions produced by an expected cost procedure perform better than solutions produced by a procedure based on a deterministic model.
- It analyzes the crew schedules found by the expected cost and deterministic methods, and it provides insight into what type of pairings perform well in operations with disruptions.

In §2 we discuss methods of evaluating the quality of a crew schedule. In §3 we give two algorithms for finding crew schedules that may perform well in operations with disruptions. We also provide a method of finding a lower bound on the expected operational cost of a crew schedule. In §4 we provide computational results for three fleets from a major domestic carrier. We present conclusions in §5.

2. Evaluating a Crew Schedule

Although finding good crew schedules is critical for airlines, an important question is what is meant by a "good" schedule. Airlines have traditionally evaluated a crew schedule by its planned cost. This implicitly assumes that every leg will be operated as planned, but evidence suggests that this rarely happens. We propose the evaluation of a crew schedule by its performance in operations with disruptions. To evaluate a crew schedule's performance in operations, we must first specify mechanisms and probabilities of disruptions, as well as a recovery policy. To find a best crew schedule, we must prescribe a method of comparing different crew schedules. We have used SimAir, a simulation of airline operations with disruptions, for our experiments.

Two ways of measuring pilot compensation, planned cost of a crew schedule, and operational cost of a crew schedule, are discussed in §§2.1 and 2.2, respectively. The planned cost is deterministic and is the traditional method used by airlines to evaluate crew schedules. The operational cost is a random variable and is unknown in the planning stage. Therefore, to use the concept of operational cost in planning, we consider expected cost, which is very difficult to calculate.

We discuss the operational cost of a crew schedule in §2.2. We describe methods for comparing different crew schedules in operations in §2.3.

2.1. The Planned Cost of a Crew Schedule

Airline planners do not measure the cost of a crew schedule in monetary terms. Rather, they express it in terms of minutes of pay and credit. The flight-time-credit (FTC) of a duty is the difference between its total cost in minutes of pay and credit and the total block time expressed as a percentage of the total block time of the duty. A similar measure exists for pairings and crew schedules. We will let $\underline{FTC}(\cdot)$ denote the planned FTC of any duty, pairing, or crew schedule. We give an example of one method for calculating the planned cost of a crew schedule.

Let q be any pairing consisting of duties $d_1, \ldots, d_k$. For $1 \le i \le k$, duty d_i consists of legs $l_{i,1}, \ldots, l_{i,m(i)}$. For $1 \le j \le m(i)$, let $\underline{dep}(l_{i,j})$ be the scheduled departure of leg $l_{i,j}$ in minutes and let $\underline{arr}(l_{i,j})$ be its scheduled arrival time in minutes. These times are relative to the start of the pairing, so that for $1 \le i \le k-1$, $\underline{dep}(l_{i+1,1}) > \underline{arr}(l_{i,m(i)})$. Let $\underline{block}(l_{i,j})$ be the planned block time of leg $l_{i,j}$ in minutes, defined by $\underline{block}(l_{i,j}) = \underline{arr}(l_{i,j}) - \underline{dep}(l_{i,j})$. Let *brief* be the length of the pilot briefing period prior to every duty. Let *debrief* be the length of the pilot debriefing period after every duty. The parameters *brief* and *debrief* are constants and are in minutes. For each duty d_i, $1 \le i \le k$, define the planned elapsed time of the duty as $\underline{elapse}(d_i) = \underline{arr}(l_{i,m(i)}) - \underline{dep}(l_{i,1}) + brief + debrief$. Let $r_e < 1$ be a fraction representing the rate of pay for elapsed time in terms of minutes of pay and credit. Let mg_d be the minimum guarantee for a duty, which is given in minutes of pay and credit. The planned duty cost of duty d_i is expressed in minutes of pay and credit and is given by

$$\underline{b}(d_i) = \max\left\{\sum_{j=1}^{m(i)} \underline{block}(l_{i,j}),\, r_e \times \underline{elapse}(d_i),\, mg_d\right\}. \quad (2)$$

The planned FTC of duty d_i is given by

$$\underline{FTC}(d_i) = \frac{\underline{b}(d_i) - \sum_{j=1}^{m(i)} \underline{block}(l_{i,j})}{\sum_{j=1}^{m(i)} \underline{block}(l_{i,j})}. \quad (3)$$

The planned TAFB of pairing q is the total number of minutes that elapse during the pairing given by $\underline{TAFB}(q) = \underline{arr}(l_{k,m(k)}) - \underline{dep}(l_{1,1}) + brief + debrief$. Let $r_t < 1$ be a fraction representing the rate of pay of TAFB. Let mg_p be a minimum guarantee per duty in a pairing. Then the planned pairing cost of pairing q is given by

$$\underline{c}_q = \max\left\{\sum_{i=1}^{k} \underline{b}(d_i),\, r_t \times \underline{TAFB}(q),\, mg_p \times k\right\}. \quad (4)$$

We use values of $r_e = \frac{4}{7}$, $mg_d = 0$, $r_t = \frac{2}{7}$, and $mg_p = 300$, following Vance et al. (1997).

The planned FTC of pairing q is defined by

$$\underline{FTC}(q) = \frac{\underline{c}_q - \sum_{i=1}^{k}\sum_{j=1}^{m(i)} \underline{block}(l_{i,j})}{\sum_{i=1}^{k}\sum_{j=1}^{m(i)} \underline{block}(l_{i,j})}. \quad (5)$$

Let $\underline{c}(C)$ be the planned cost of a crew schedule C consisting of pairings $q_1, \ldots, q_{|C|}$, given by $\underline{c}(C) = \sum_{q\in C} \underline{c}_q$. Let $\underline{block}(C)$ be the total scheduled block time of all legs in the flight schedule. The planned FTC of crew schedule C is

$$\underline{FTC}(C) = \frac{\underline{c}(C) - \underline{block}(C)}{\underline{block}(C)}. \tag{6}$$

2.2. The Operational Cost of a Crew Schedule

The operational cost of a crew schedule is the sum of the operational costs of the pairings that compose it. We give an example of how one major domestic carrier calculates the operational cost of a pairing. Other airlines may have slightly different methods of determining the operational cost of a pairing.

The operational block time of each leg is the difference between its actual arrival time and actual departure time. The operational cost of each duty is the same as its planned cost, except that the planned elapsed time and block times are replaced by their operational counterparts. The calculation is similar for pairing costs, with the operational TAFB replacing the planned TAFB, except that the operational cost of a pairing must be at least its planned cost. The operational cost of a crew schedule is the sum of the operational costs of the pairings that compose it.

In this paper we assume that no flights are cancelled and that each duty and pairing fly the legs originally assigned to them, although possibly with different departure and arrival times. These assumptions are discussed further in §3.

2.3. SimAir—A Simulation of Airline Operations

We use SimAir, a Monte Carlo simulation of airline operations with disruptions, to evaluate a crew schedule's performance. Descriptions of SimAir and the underlying stochastic model are given in Rosenberger et al. (2000, 2002). SimAir permits the study of a crew schedule under a recovery method and delay distribution. SimAir explicitly models crews, planes, and passengers. SimAir is currently being used by major U.S. carriers, airline software firms, and academic researchers.

Although SimAir can use any recovery method, in this paper we focus on the "push-back" recovery heuristic introduced in Rosenberger et al. (2002) and Schaefer (2000). This recovery method delays the departure of each flight until the crew and the plane are available. We have focused on push-back because it is easier to analyze than other recovery methods and in cases of delays of short duration it is used in practice.

3. Methodology

Exact formulations for crew scheduling under uncertainty are intractable due to the enormous state space, action space, and number of time periods required. The deterministic crew scheduling model is an approximation of crew scheduling under uncertainty under the assumption that all pairings will operate as planned. We seek better approximate solution methods for crew scheduling under uncertainty that still remain tractable.

Airline crew scheduling under uncertainty could be formulated as a Markov decision process (MDP). However, such a model would be intractable. The state of the system describes every aspect of the system that is relevant for operational decisions. The state must contain information about the current status and history of every crew member and plane, as well as a description of the current operating environment including weather, congestion, and so on. The first stage consists of the planning period, where a flight schedule, fleet assignment, routing, and initial crew schedule are found. Operational decisions are made in subsequent stages. The number of stages could be quite large, because the state of the system can change within minutes. The action space consists of all possible feasible decisions. These include cancelling flights, rerouting planes and passengers, rescheduling crews, and so on. These operational decisions may have a profound impact on the crew legality of subsequent flights due to complicated regulations, such as the 8-in-24 rule.

We introduce two methods for finding crew schedules that may perform well in operations. These methods seek pairing costs that more accurately reflect the cost of a given pairing in operations with disruptions. After these costs are found, a set partitioning model is solved using an algorithm developed by Klabjan, Johnson, and Nemhauser (2001).

One approach is to find a linear approximation of the expected crew cost. For any crew schedule C, let $\bar{c}(C)$ be its expected crew cost in operations with disruptions. If pairing costs χ_q exist such that

$$\bar{c}(C) = \sum_{q\in C} \chi_q \tag{7}$$

for all crew schedules C, then an optimal solution to the stochastic crew scheduling problem can be found by solving the set partitioning problem using such pairing costs.

In general, such costs χ_q that satisfy (7) do not exist. (See Schaefer 2000 for an example. This example exploits the fact that the expected operational cost of a pairing may depend on the other pairings in the crew schedule.) Therefore, we seek costs χ_q such that (7) is satisfied approximately.

3.1. The Expected Cost of a Pairing

Pairings interact when the cost of a given pairing depends on other pairings in the schedule. Interactions occur because pairings share resources such as planes, gates, flight attendants, and passengers. However, the only ways in which pairings may interact directly in our model is through shared planes and recovery. We make the following assumptions to find pairing costs that satisfy (7).

Assumption A1. *The planes are always available.*

Assumption A2. *The recovery method is push-back, so that the departure of each flight is delayed until the crew is available and the scheduled departure time has passed.*

These assumptions do not hold in practice. Crews often must wait for late planes, and airlines often use recovery policies other than push-back. We ran an experiment to check the impact of Assumption A1 within our model of airline operations by considering a set of 136 crew schedules and simulating each for 10,000 days of airline operations in SimAir. For Experiment A we used the planned routing for this fleet. If the plane was delayed from a previous flight, it may not be available for its next flight, even if the crew is available. For Experiment B we used Assumption A1, so that the planes were always available. It appears from these experiments that Assumption A1 is reasonable for measuring FTC. Between Experiments A and B the average operational FTC decreased by an average of 0.0986. However, the variance was very small: 9.76×10^{-5}. This indicates that although Assumption A1 does not hold in practice the reduction in FTC is nearly constant across crew schedules.

Assumption A2 does not capture all recovery options at a hub. At hubs, airlines have many more options, and allow crews to fly pairings other than the ones to which they were assigned in planning. At spokes, there may be no reserve crews available or crews available for swaps, so Assumption A2 may reflect the only option available to airlines.

Under Assumptions A1 and A2, we define the *expected operational cost in isolation* of any pairing q, $\bar{c}_q$, to be the expected operational cost under the push-back recovery heuristic of a crew schedule that consists only of pairing q.

Theorem 1. *Under Assumptions* A1 *and* A2, $\bar{c}(C) = \sum_{q\in C} \bar{c}_q$ *for all C, that is, pairing costs* $\bar{c}_q$ *and crew schedule costs* $\bar{c}(C)$ *satisfy* (7).

We give a sketch of the full proof given in Schaefer (2000). The proof considers any pairing q for two different cases. The first assumes that the schedule consists only of pairing q and the other assumes Assumptions A1 and A2. The proof shows by induction that for any sample path of delays the operational departure and arrival time of every leg in q is the same in both cases. This implies that the pairing costs are the same in both cases, and hence the operational cost of both crew schedules is the same. Because this holds for any sample path, Theorem 1 holds.

Because there is unlikely to be a simple formula for $\bar{c}_q$, we use a Monte Carlo simulation to estimate $\bar{c}_q$. This simulation is similar to SimAir, except that while SimAir simulates an entire fleet, this method simulates one pairing for a number of days. In our experiments, the sample size consisted of at least 50 days and no more than 500 days. We used a 99% confidence level for the termination criterion, so that once the 99% confidence interval of the sample mean for the expected pairing cost was sufficiently small we stopped the simulation for that pairing. The time required to estimate the expected cost of each pairing was very small.

3.2. A Penalty Method

Certain attributes of a pairing may lead it to perform poorly in operations with disruptions. A pairing may be close to operational limits. A pairing may contain a duty that is close to operational limits on flying time or elapsed time. A pairing containing such a duty may become illegal if it is subjected to delays.

A pairing may remain legal in operations, but still perform poorly. A pairing with short sits may not be able to absorb delays without undertaking some recovery action. Similarly, a pairing with short rests may not be able to start the subsequent duty without delay.

One approach is to penalize certain attributes of pairings that may lead to poor performance in operations with disruptions. The optimal crew schedule for a given set of penalties can be found by solving a set partitioning problem (Klabjan, Johnson, and Nemhauser 2001). We give a local search method that attempts to find a good set of penalties. The hope is that the crew schedule resulting from the best set of penalties will perform well in operations with disruptions.

Consider any k attributes of pairings. Examples of the attributes we considered include the number of sit minutes when the crew changes planes, the number of minutes of rest, the number of minutes of elapsed duty time, and the number of minutes of duty flying time. Let the penalty space, Υ, be a subset of $\Re^{2k}$. For any attribute $1 \le i \le k$, let $n(i, q)$ be the number of attributes of type i that pairing q has. For instance, if attribute i is the number of minutes for a sit where the crew changes planes, $n(i, q)$ is the number of sits where the crew changes planes. Let $a_q^{(i,j)}$ be the value of attribute i for pairing q, where $1 \le j \le n(i, q)$ and $1 \le i \le k$. For instance, if pairing q has five sits where the crew changes planes, with lengths 55, 72, 63, 58, and 67 minutes, respectively, then $n(i, q) = 5$, $a_q^{(i,1)} = 55$, $a_q^{(i,2)} = 72$, $a_q^{(i,3)} = 63$, $a_q^{(i,4)} = 58$, and $a_q^{(i,5)} = 67$.

For an attribute i, $1 \le i \le k$, let ϑ_i be the largest or smallest amount that is acceptable. For the attribute corresponding to the elapsed time of a duty, ϑ is the maximum elapsed time permitted for a duty, but for sits, ϑ is the length of the minimum legal sit. For instance, if the sit attribute has a ϑ value of 45 minutes, sits shorter than 45 minutes are not permitted. Let α_i and γ_i be positive real numbers. We interpret α_i as the maximum penalty for factor i, and γ_i as the slope of the penalty function. Consider any penalty combination $(\alpha, \gamma) = (\alpha_1, \ldots, \alpha_k, \gamma_1, \ldots, \gamma_k) \in Y$. Then the function $f_i(\cdot)$ is defined as

$$f^i(\alpha_i, \gamma_i, q) = \sum_{j=1}^{n(i,q)} \max(\alpha_i - \gamma_i |a_q^{(i,j)} - \vartheta_i|, 0). \tag{8}$$

As $a_q^{(i,j)}$ approaches ϑ_i, the resulting schedule may be more likely to have disruptions due to that attribute. For example, as the sit time decreases, the more likely it will be that a flight must be delayed because the crew is unavailable. Whenever $|a_q^{(i,j)} - \vartheta_i| \geq \alpha_i/\gamma_i$, the penalized amount is 0 for that particular j.

For any $(\alpha, \gamma) \in Y$, let $x^*(\alpha, \gamma)$ be the optimal solution to the deterministic crew scheduling problem with pairing costs

$$c_q = \underline{c}_q + \sum_{i=1}^{k} f^i(\alpha_i, \gamma_i, q).$$

Let $s(\alpha, \gamma)$ be the expected cost of crew schedule $x^*(\alpha, \gamma)$ as estimated through SimAir. To determine $s(\alpha, \gamma)$, crew schedule $x^*(\alpha, \gamma)$ is simulated by SimAir. We would like to solve

$$\min_{\alpha, \gamma \in Y} s(\alpha, \gamma). \tag{9}$$

Unfortunately, this problem is very difficult to solve. In general, $s(\alpha, \gamma)$ is not continuous, and for any given $(\alpha, \gamma) \in Y$, estimating $s(\alpha, \gamma)$ is quite expensive, because finding $x^*(\alpha, \gamma)$ requires solving a deterministic crew scheduling problem and simulating operations with disruptions under the resulting schedule. The computational results given in Schaefer (2000) demonstrate that $s(\alpha, \gamma)$ is neither convex nor concave. It seems unlikely that a global optimum to problem (9) can be found. To find schedules that perform well in operations with disruptions, we propose a local search of the penalty space to find an approximate solution to problem (9).

For our experiments we used the four attributes listed below.

Attribute 1. The number of minutes of scheduled sit when the crew is scheduled to change planes, hereafter referred to as *swap time*. The parameter ϑ_1 is set to 45 minutes.

Attribute 2. The number of minutes of scheduled rest time between duties. The parameter ϑ_2 is set to 615 minutes, or 10 hours and 15 minutes.

Attribute 3. The number of minutes of flying in a duty. The parameter ϑ_3 is set to 480 minutes, or 8 hours.

Attribute 4. The number of minutes of elapsed time in a duty. The parameter ϑ_4 is set to 810 minutes, or 13 hours and 30 minutes.

The local search procedure starts with the deterministic or planning solution, with all penalty levels set to zero. It then varies the penalty parameters for Attributes 1 through 4 sequentially. After the penalty parameters for the fourth factor has been considered, the penalty parameters for the first factor are again varied to see if any improvement is possible. If no improvement has been found while varying the penalty parameters for the first attribute, the algorithm terminates and the incumbent solution is returned. Otherwise, if an improved solution is found, we reexamine the other factors.

This penalty method differs from the methods often used by some airlines, which rely on penalties that may vary by time of day and departure and arrival airports, and are determined in an ad hoc manner. However, such a large number of different penalties would make a systematic search for optimal penalties prohibitive.

3.3. A Lower Bound on the Expected Cost of an Optimal Crew Schedule

In this section we give a method that finds a lower bound on the optimal objective function value for the problem of crew scheduling under uncertainty if no 8-in-24 regulations are considered in operations with disruptions. Let q be any pairing, and let $\bar{o}_q$ be its expected cost in isolation while ignoring operational 8-in-24 rules. For any crew schedule C, define

$$\bar{o}(C) = \sum_{q \in C} \bar{o}_q, \tag{10}$$

and let $\hat{o}(C)$ be the expected cost of crew schedule C ignoring planning 8-in-24 rules as measured by SimAir.

Theorem 2. *Under the push-back recovery heuristic with no* 8-*in*-24 *rules, for any crew schedule* C,

$$\bar{o}(C) \leq \hat{o}(C). \tag{11}$$

The proof is similar to that of Theorem 1, and appears in Schaefer (2000). Notice that the difference between $\bar{o}_q$ and $\hat{o}_q$ is the interaction among pairings. The proof considers a sample path of delays and shows that by ignoring the interactions among pairings the operational crew costs are no greater. The reason for this is that when pairings interact, a crew may need to wait for a plane before flying a given leg. Although the operational flying times are not affected, operational duty elapsed time and TAFB can increase when pairings interact. Because this is true for any sample path of delays, Theorem 2 holds.

The assumption in Theorem 2 of no 8-in-24 rules is needed because it is possible to have less operational cost with 8-in-24 rules than without such a rule. The reason is that 8-in-24 rules could trigger compensatory rest, which in turn could shorten the elapsed time of the subsequent duty.

It is highly unlikely that these lower bounds are tight. If these bounds were tight it would imply that late planes do not affect crew costs in operations with disruptions, which seems dubious.

4. Computational Results

We estimated four probability distributions for random disruptions from real operational data. Three of these distributions were for the flight time disruptions, and one was for delays that occur at the airports. These flight time probability distributions are relative to the planned block time, so a negative observation would correspond to a shorter flight than planned. The flight time distribution depended on the scheduled flight time, because a long flight is more likely to be significantly early than a short flight. We divided flights into three categories: (1) Short flights are those with scheduled flight times of two hours or less. The probability distribution for the flight delays of short flights is given in Figure 1. (2) Medium flights are those with scheduled flight times between two and four hours. (3) Long flights are those with scheduled flight times greater than four hours. Ground delays are delays that may contribute to elapsed duty time or TAFB but will not count toward flying time. Because our method considers only frictional delays, we have truncated the ground time distribution and therefore our experiments did not consider ground delays exceeding five hours, although such delays do occur in practice. The ground time distribution was the same for all flights, regardless of airport, time of day, and so on, which is not true in practice but is used here because of limited data availability.

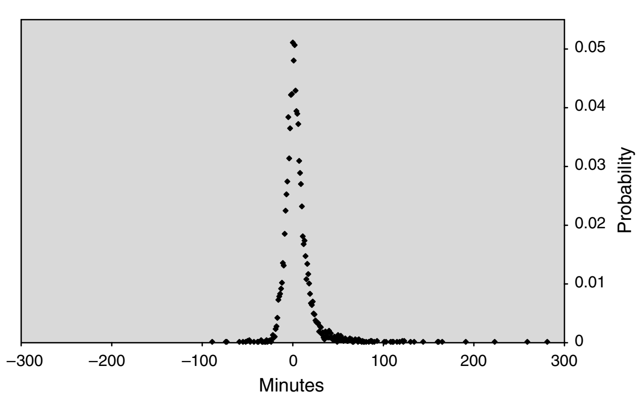

Figure 1 Probability Distribution for Delays of Short Flights

Table 1 Summary of the Four Distributions

	Distribution
Short flights	$-254 + \text{Gamma}(5.82, 5.07)$
Medium flights	$-49 + \text{Erlang}(2.81, 19)$
Long flights	$-71 + \text{Gamma}(2.96, 24.8)$
Ground time	$-0.001 + 146 * \text{Beta}(0.61, 23.6)$

We fit the various distributions to parameterized distributions using the Arena Input Analyzer from Rockwell Software. In the chi square tests, all p values were smaller than 0.005. The resulting distributions are given in Table 1. The flight time may also depend on factors such as the time of day, day of week, airport, and so on. However, these factors are considered when airlines determine the planned flight times for various flights.

4.1. Computational Results for Three Fleets

We considered three daily fleets provided to us by a major carrier. We refer to these fleets as F1, F2, and F3. For fleet $i = 1, 2, 3$, let $\underline{C}_{Fi}$ be the deterministic crew schedule, let $\bar{C}_{Fi}$ be the crew schedule found by the expected cost method, and let $\widehat{C}_{Fi}$ be the best crew schedule found by the penalty method. Table 2 summarizes the computational results. For each crew schedule, the lower bound refers to the bound provided by Theorem 2, and the "Operational FTC gap" is the difference between the operational FTC of the schedule and the lower bound on operational performance.

Although the simulation results are inconclusive because the absolute differences among the planned and operational FTCs are small for all three fleets, they do indicate that there is not a monotonic relationship between planned and operational FTCs. Schedules that anticipate disruptions do perform somewhat better in operations with disruptions, and perform only slightly worse if the planned schedule is actually flown.

For each of the three fleets the crew schedule found by the expected cost algorithm performed better than

Table 2 Summary for the Three Fleets

Fleet	Number of legs	Crew schedule	Planned FTC	Operational FTC	Lower bound	Operational FTC gap
$F1$	119	$\underline{C}_{F1}$	2.51	4.31	4.10	0.21
$F1$	119	$\bar{C}_{F1}$	2.64	4.22	4.10	0.12
$F1$	119	$\widehat{C}_{F1}$	2.51	4.29	4.10	0.19
$F2$	149	$\underline{C}_{F2}$	3.79	9.11	8.40	0.71
$F2$	149	$\bar{C}_{F2}$	3.92	8.97	8.40	0.57
$F2$	149	$\widehat{C}_{F2}$	3.79	9.11	8.40	0.71
$F3$	342	$\underline{C}_{F3}$	2.69	5.82	5.51	0.31
$F3$	342	$\bar{C}_{F3}$	2.91	5.69	5.51	0.18
$F3$	342	$\widehat{C}_{F3}$	2.74	5.75	5.51	0.24

the crew schedule found using deterministic method as well as the penalty method. Relative to the lower bounds established by ignoring operational 8-in-24 rules, the expected cost algorithm performed noticeably better than the deterministic method. The difference between the cost of the schedules found by the expected cost algorithm and the lower bound was more than 50% smaller than the difference between the cost of the deterministic schedules and the lower bound for two of the fleets.

We would anticipate that the expected cost algorithm would be even more advantageous under more severe disruptions. Furthermore, if disruptions were correlated (for instance, with weather delays) the difference in operational performance between the expected cost schedules and the deterministic schedules might also increase.

4.2. An Analysis of the Crew Schedules

To test the validity of assuming that planes are always available in the calculation of the expected pairing costs, we considered the effect of simulating the entire crew schedule compared to simulating each of the pairings in isolation. For each crew schedule C found for each of the fleets we express the difference between $\sum_{q\in C} \bar{c}_q$ and crew schedule C's planned cost as a percentage of the total increase between the operational and planned cost of crew schedule C. Mathematically, this is expressed as

$$\frac{\sum_{q\in C} \bar{c}_q - \underline{c}(C)}{\bar{c}(C) - \underline{c}(C)} \cdot 100. \tag{12}$$

The results are displayed in Table 3. The consistently large percentages indicate that possible interactions among pairings have a small impact on the total difference between a crew schedule's operational and planned cost. This provides further empirical evidence that Assumption A1 appears to be reasonable in our model of airline operations.

We now analyze the schedules to determine how often the crew follows the planes according to the actual routing, and how many pairings and duties are in both $\underline{C}_{Fi}$ and $\bar{C}_{Fi}$ for $i = 1, 2, 3$. We give the factor dominating the pairing costs, as well as the largest deterministic FTC for each fleet. The results are given in Table 4.

Table 3 The Effect of Pairing Interactions

Crew schedule	Cost increase explained by $\bar{c}$ costs (%)
$\underline{C}_{F1}$	94
$\bar{C}_{F1}$	92
$\underline{C}_{F2}$	90
$\bar{C}_{F2}$	89
$\underline{C}_{F3}$	96
$\bar{C}_{F3}$	94

Table 4 A Comparison of Crew Schedules $\underline{C}$ and $\bar{C}$ for the Three Fleets

Crew schedule	Matching pairings	Matching duties	Crew follows plane	$\sum$ duty is max	TAFB is max	MG is max	Max FTC
$\underline{C}_{F1}$	12/25	63/72	21/47	15/25	10/25	0/25	12.7
$\bar{C}_{F1}$	12/24	63/73	22/46	13/24	11/24	0/24	5.2
$\underline{C}_{F2}$	7/14	34/41	47/108	4/14	5/14	5/14	30.4
$\bar{C}_{F2}$	7/15	34/43	46/106	4/15	6/15	5/15	22.7
$\underline{C}_{F3}$	17/42	83/124	68/218	20/42	20/42	2/42	16.2
$\bar{C}_{F3}$	17/41	83/124	75/218	17/41	19/41	5/41	14.5

For these three fleets, fewer pairings determined by the expected cost algorithm had the sum of the duty costs as the dominant factor in their costs. Intuitively, this makes sense: TAFB depends largely on the end of the pairing, so in most cases it has a smaller variance than the sum of the duty costs. By choosing more pairings where TAFB is the largest factor in planning, the expected cost algorithm is able to choose from a richer set of pairings. By doing this, it is able to avoid pairings with large deterministic FTCs. It is able to recognize that even though a pairing may have TAFB or minimum guarantee dominate in planning, it does not necessarily mean that this will remain the case in operations with disruptions, as modeled by SimAir. This may be why the expected cost algorithm appears to perform better in operations with disruptions; it is able to consider a wider range of pairings that are likely to be paid for flying time in operations, rather than the smaller set that is paid for flying time in planning.

Several patterns emerged across all three fleets. Pairings with zero planned FTCs had larger differences between their planned and operational FTCs than pairings with positive planned FTCs. This has several implications. First, a pairing with a small planned FTC may be equally desirable in operations with disruptions as a pairing with zero planned FTCs. Second, given two crew schedules with equal total planned FTCs, it may be preferable to choose a crew schedule with many pairings with small planned FTCs over a crew schedule with many zero-planned-FTC pairings and a few pairings with large planned FTCs. For all three fleets, the expected cost schedule had fewer zero-planned-FTC pairings than the deterministic schedule. Because the expected cost algorithm views more pairings as acceptable, it is able to avoid using pairings with large planned FTCs. Given two crew schedules with equal planned FTCs, having many pairings with positive planned FTCs appears to be more desirable than having a few. Pairings with small planned FTCs still may have the sum of duty costs dominate in operations with disruptions; this is unlikely for pairings with large planned FTCs. It also appears that the cost due to interaction between pairings is insignificant compared with the cost arising from each pairing considered in isolation. Isolating pairings allows us to use the standard set partitioning model for solving these crew scheduling problems. This is a significant finding, because explicitly considering interactions between pairings would make solving crew scheduling problems even more difficult.

Using the pairing costs $\bar{c}$ found by the expected cost algorithm in the standard set partitioning crew scheduling model results in crew schedules that perform better than deterministic crew schedules in the model of airline operations used by SimAir. A reduction in operational crew costs may be found by considering each pairing in isolation and then using its expected operational cost in the objective function of the crew scheduling problem. One insight provided by these results is that pairings with small planned FTCs may in fact perform well in operations with disruptions, because pairings with zero planned FTCs appear to have the largest difference between operational and planned FTCs. The expected cost algorithm recognizes this, and hence it has fewer zero-planned-FTC pairings than the deterministic crew schedules.

5. Conclusions

This paper introduces a new way of evaluating the quality of a crew schedule, namely, its performance in operations with disruptions. We give a method for finding crew schedules that approximate expected cost. Our computational results indicate that the schedules found using approximate expected cost perform better in operations with disruptions than those found by using planned cost. Possible extensions of this work include incorporating recovery methods that are more realistic, considering other measures of operational performance such as on-time performance, and integrating other decisions such as fleet assignment.

Acknowledgments

The authors thank two anonymous referees and an anonymous associate editor for many helpful comments. These comments have improved the content and readability of the paper. This research was sponsored by United Airlines Research and Development grants, and National Science Foundation Grants DMI-9700285, DMI-0085723, and DMI-0121495. The authors would like to thank Srini Ramaswamy of United Airlines, Eric Gelman of American Airlines, Jay Rosenberger of the University of Texas, Arlington, and Dave Goldsman of the Georgia Institute of Technology for their suggestions and insight. The computational experiments were made using hardware donated by the Intel Corporation and optimization software provided by ILOG.

References

Barnhart, C., A. Cohn, E. L. Johnson, D. Klabjan, G. L. Nemhauser, P. Vance. 2002. Airline crew scheduling. R. W. Hall, ed. *Handbook in Transportation Science*. Kluwer Scientific Publishers, 517–560.

BTS. 1998. On-time statistics. http://www.bts.gov/ntda/oai/desc.html.

Economist, The. 2000. A jam at 32,000 feet. (February 5) 57–58.

Ehrgott, M., D. M. Ryan. 2002. Constructing robust crew schedules with bicriteria optimization. *J. Multi-Criteria Decision Anal.* **11**(3) 139–150.

FAA. 1999. Federal Aviation Regulations. http://www.faa.gov/avr/AFS/FARS/far-121.txt.

Hoffman, K. L., M. Padberg. 1993. Solving airline crew scheduling problems by branch-and-cut. *Management Sci.* **39** 657–682.

Klabjan, D., E. L. Johnson, G. L. Nemhauser. 2001. Solving large airline crew scheduling problems: Random pairing generation and strong branching. *Computational Optim. Appl.* **20** 73–91.

Lettovský, L. 1997. Airline operations recovery: An optimization approach. Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA.

Mathews, A. W. 2000. Busier skies, more stress on systems forecast by FAA for the next decade. *The Wall Street J.* (March 7) A3.

O'Dell, R. 2000. Airline delays skyrocket 58% from 1995 to '99, report says. *The Atlanta Constitution* (July 26) E1.

Phillips, D., N. Irwin. 2000. Flying into a storm of delays. *The Washington Post* **103**(July 17) A01.

Rosenberger, J. M., A. J. Schaefer, D. Goldsman, E. L. Johnson, A. J. Kleywegt, G. L. Nemhauser. 2000. Simair: A stochastic model of airline operations. J. Joines, R. R. Barton, P. Fishwick, K. Kang, eds. *Proc. 2000 Winter Simulation Conf.*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, 1118–1122.

Rosenberger, J. M., A. J. Schaefer, D. Goldsman, E. L. Johnson, A. J. Kleywegt, G. L. Nemhauser. 2002. A stochastic model of airline operations. *Transportation Sci.* **36** 357–377.

Schaefer, A. J. 2000. Airline crew scheduling under uncertainty. Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA.

Vance, P. H., A. Atamturk, C. Barnhart, E. Gelman, E. L. Johnson, A. Krishna, D. Madidhara, G. L. Nemhauser, R. Rebello. 1997. A heuristic branch-and-price approach for the airline crew pairing problem. Technical Report TLI/LEC-97-06, Georgia Institute of Technology, Atlanta, GA.

Yen, J. W. 2000. A stochastic programming formulation of the stochastic crew scheduling problem. Ph.D. thesis, University of Michigan, Ann Arbor, MI.