Titre: Title:	Airline crew scheduling: models, algorithms, and data sets
Auteurs: Authors:	Atoosa Kasirzadeh, Mohammed Saddoune, & François Soumis
Date:	2017
Type:	Article de revue / Article
Référence: Citation:	Kasirzadeh, A., Saddoune, M., & Soumis, F. (2017). Airline crew scheduling: models, algorithms, and data sets. EURO Journal on Transportation and Logistics, 6(2), 111-137. https://doi.org/10.1007/s13676-015-0080-x

CrossMark

EU OR

Airline crew scheduling: models, algorithms, and data sets

Atoosa Kasirzadeh · Mohammed Saddoune · François Soumis

Abstract The airline crew scheduling problem has received extensive attention, particularly in the last 60 years. This problem is frequently divided into crew pairing and crew assignment because of its large size and the complex safety agreements and contractual rules. Several solution methodologies have been developed, but many objectives and constraints are treated approximately and research is ongoing. In this paper, we present a comprehensive problem definition for the airline crew scheduling problem, and we review existing problem formulations and solution methodologies. In addition, we formulate the personalized cockpit crew scheduling problem as a set covering problem and we solve it using column generation. We present computational results for real data from a major US carrier, and we describe the data sets (available on the internet) in detail to establish a basis for future research.

A. Kasirzadeh (✉) · M. Saddoune · F. Soumis
Department of Mathematics and Industrial Engineering, GERAD and École Polytechnique de Montréal, Succ. Centre-Ville, C.P. 6079, Montreal, QC H3C 3A7, Canada
e-mail: atoosa.kasirzadeh@gerad.ca

M. Saddoune
e-mail: mohammed.saddoune@polymtl.ca

F. Soumis
e-mail: francois.soumis@gerad.ca

M. Saddoune
Université Hassan II, Faculté des Sciences et Techniques de Mohammedia, BP 146, 20650 Mohammedia, Morocco

 Springer

1 Introduction

The airline industry and its operations have been a major focus of operation researchers, especially since the advent of the jet age in the late 1950s, which was followed by major technological advances. The industry has become a significant economic force from two perspectives: its own operations and its impact on related industries such as tourism and aircraft manufacturing. The revenue mainly comes from passenger tickets, while the costs include airplane expenses, fuel, crew, and equipment. The total profit is a complicated function of all of the operations. Data from the Air Transport Association (2008) indicate that the largest administrative cost is fuel expenses, and the second largest is labor costs (23.4 %) (Belobaba et al. 2009). Minimizing the crew costs is therefore an essential task in today's competitive airline industry, and even a small reduction can lead to significant savings. In addition, the recent appearance of low-fare airlines has increased the pressure to provide affordable tickets and reemphasized the importance of minimizing expenses. As a result, the airline crew scheduling problem has received much attention in both industry and academia.

Airline crew scheduling is the problem of assigning a group of crew members to a set of scheduled flights such that all the scheduled flights are covered while the rules and collective agreements, imposed mainly by safety and labor organizations, are respected. The complex restrictions make this one of the most difficult crew scheduling problems in the transportation industry.

We present an extensive review of research into the airline crew scheduling problem, and we observe that authors do not compare their methods on the same data. We present a new model and solution approach for the personalized pilot assignment problem with pre-assigned activities and preferences that have not yet been introduced in the literature. Our model is a set partitioning model that we solve via column generation (CG). The remainder of this paper is structured as follows. In Sect. 2, we define the airline crew scheduling terminology that we use. Section 3 explains the different decision problems faced by airlines. In Sect. 4, we present an extensive review of airline crew scheduling. Section 5 discusses the different rules and agreements that generally apply. In Sect. 6, we give a detailed mathematical description of our problem, personalized pilot assignment, and Sect. 7 describes our solution approach in detail. In Sect. 8, we present numerical results for seven data sets. Our clear description of the problems (the rules and agreements) and the available data sets will allow other researchers to test their methods on the same data sets. In Sect. 9, we discuss the data sets and potential directions for future research. Section 10 presents concluding remarks.

2 Airline crew scheduling terminology

In this section, we define the terminology that we use in our discussion.

– *Air leg* A nonstop flight segment. Each air leg is characterized by five features: the flight number, the origin airport, the destination airport, the departure time, and the arrival time.

- *Deadhead* An air leg in which a crew member flies as a passenger for relocation purposes.
- *Duty* A sequence of consecutive air legs (and/or deadheads) comprising a working day for a single crew member. Two consecutive duties should begin and end at the same airport. Duties are separated by layovers.
- *Layover* A rest period (an overnight stop) between duties that typically lasts for at least 10 h.
- *Pairing* A sequence of duties and layovers for an unspecified crew member that starts and ends at a base. In short- and medium-haul problems, pairings typically last 1–5 days; in long-haul problems, longer pairings are allowed.
- *Base* A large airport. Each crew member is associated with a base, which means that all his/her associated pairings must begin and end at that airport.
- *Elapsed time* A period of time in which a crew member is away from the base, referred to as Time Away From Base (TAFB).
- *Credited flying time in a duty* The active flying time plus a specific percentage of deadhead flying time (typically 50 %).
- *Monthly schedule (schedule)* A sequence of pairings separated by time off that covers a given time horizon (usually a standardized planning month). In this paper, the term *schedule* refers to a monthly schedule.
- *Briefing time* A period of time before the start of each duty that is spent on instructions and crew discussions with the goal of transforming a group of individuals into an effective team.
- *Debriefing time* A period of time at the end of each duty that gives the crew members an understanding of the events that occurred and their implications.
- *Crew members* Generally divided into two groups based on their role: the *cockpit crew members* are the pilot (captain), copilot (first officer), and flight engineer, all of whom are qualified to fly one or more aircraft types. The *cabin crew members* are the cabin captain and the flight attendants.
- *Post-pairing rest* A rest period between two consecutive pairings that respects a minimum and a maximum duration.
- *Post-pairing* A rest period between two consecutive pairings that contains a complete day off (from midnight to midnight).
- *Aircraft route* A sequence of air legs flown by a specific aircraft.

3 Airline decision process

Because of its complexity and the potential perturbations, most major airlines divide the overall decision problem into two closely related procedures: planning and recovery. Each procedure is then divided into several steps that are often treated separately. The planning procedure consists of flight scheduling, fleet assignment, aircraft maintenance and routing, and crew scheduling. The recovery procedure has three steps that adjust the plans to take into account unexpected perturbations: aircraft recovery, crew recovery, and passenger recovery (see Fig. 1).
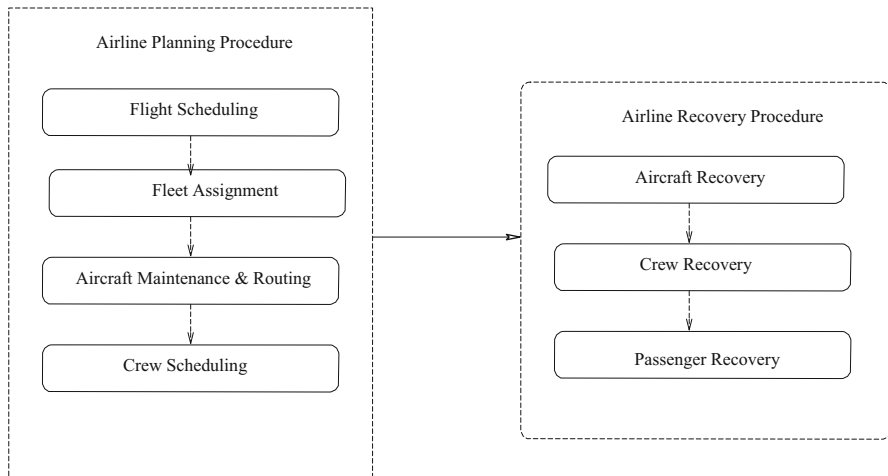
Fig. 1 Airline decision process

3.1 Airline planning procedure

The steps of this procedure are interrelated, but because of their complexity they are often solved sequentially. The solution of each step becomes input data for the next step. Recently, some researchers have integrated two or more of these steps; this is discussed in Sect. 4. The first step is the flight scheduling problem in which the air legs to be flown for a specific time horizon are scheduled with the objective of maximizing the expected profit. In the second step, fleet assignment, the various aircraft types such as the Boeing 707 and the Airbus A318 are assigned to the flights, taking into account the estimated passenger demand, the aircraft availability, and aircraft flow conservation. In the third step, aircraft maintenance and routing, individual aircraft are assigned to each scheduled flight, and the solution ensures that each aircraft spends adequate time at specific airports for routine maintenance. The fourth step, crew scheduling, is separable by crew category and aircraft type or family. It finds crew schedules that cover all the scheduled flights and satisfy the constraints. As mentioned in Sect. 2, the *cockpit crew* fly the aircraft and the *cabin crew* are responsible for passenger services and safety. A member of one group cannot normally be substituted for a member of the other group. The two groups are scheduled separately for three reasons. First, each cockpit crew is qualified to fly a specific aircraft type or family, whereas cabin crews can be assigned to multiple aircraft types. Second, the number of cabin crew required depends on the number of passengers, whereas the size of the cockpit crew is fixed. Third, cockpit crews are paid substantially more than cabin crews because of their level of expertise. As a result, most of the research into crew cost optimization has focused on cockpit crew scheduling. Hereinafter, we refer to *cockpit crew members* simply as *crew members*.

Because of its complexity and size, the crew scheduling problem is usually separated into two steps: *crew pairing* and *crew assignment*. Traditionally, the steps

have been treated sequentially. The *crew pairing* forms a minimum-cost set of anonymous feasible pairings from the scheduled flights such that all flights are covered exactly once and all the pairing regulations and contractual rules are respected. Different airlines have different rules, but the main characteristics of the anonymous pairings are common to all airlines. *Crew assignment* combines the anonymous pairings with rest periods, vacations, pre-assigned activities such as training, and other breaks over a standardized month to produce a set of individual schedules for the crew members. The schedules must satisfy all the safety regulations and contractual rules. In contrast to the crew pairing problem, the crew assignment problem is separable by crew base and fleet type, and one of two general approaches is used:

1. *Bidline schedules* are constructed anonymously and the airline then announces them to the crew members. The crew members bid on these schedules, and their bids are used to complete the schedule allocation.
2. *Personalized schedules* take into account the crew members' preferred tasks and their needs for special activities such as vacations and training periods. The pairings are combined to give monthly schedules respecting airline objectives and providing a certain level of crew satisfaction. Two types of personalized schedules are considered: *rostering* and *seniority-based*. *Rostering* aims to maximize global satisfaction and may consider a second objective of fairness, measured in terms of the number of satisfied preferences. *Seniority-based* personalized schedules give priority to the satisfaction of the more senior crew members.

Historically, bidline scheduling has been the most common approach in the US whereas personalized scheduling is more common in the rest of the world. However, personalized schedules are now increasingly accepted at North American airlines because they offer advantages for both crew members and airlines. From the crew member's perspective, this approach considers the employee's requests during the construction of the schedule. From the airline's perspective, this approach considers the predefined employee activities: vacations, training, and unfinished pairings from the previous month. This reduces the number of schedule adjustments and increases productivity.

3.2 Airline recovery procedure

In real life, there are many unpredicted disruptions that arise because of weather conditions, aircraft maintenance issues, crew problems, and other unplanned events. These can lead to delayed or canceled flights, and they perturb the crew schedules.

When such disruptions occur, the airline recovery procedure updates the scheduled flights, aircraft routes, crew schedules, and passenger itineraries. The time horizon is short (usually 1–4 days). The recovery procedure is not the focus of this paper, and we do not discuss it further.

4 Review of crew scheduling

We now provide a comprehensive review of models and algorithms for airline crew scheduling. We discuss the different objectives such as cost-minimization and employee satisfaction. As mentioned in Sect. 3.1, the crew scheduling problem is usually separated into two steps: crew pairing and crew assignment. The steps are usually solved sequentially. However, recently several researchers have explored joint optimization approaches. We do not discuss the heuristic acceleration techniques that are used in industry to solve large problems because there are many parameters that must be carefully adjusted, and this topic is beyond the scope of our paper.

Mathematically, crew pairing and assignment problems are usually modeled via the set partitioning problem (SPP) or the set covering problem (SCP) with additional constraints. The problems are difficult because of the number of constraints and variables; they are large-scale mixed integer problems. The three most common solution methodologies are Lagrangian relaxation (Geoffrion 1974; Fisher 1981, 1988; Martin 1999), Benders decomposition (Benders 1962; Minoux and Vajda 1986) and branch-and-price (Desaulniers et al. 1998; Desrosiers et al. 1995; Barnhart et al. 1996). Since the 1990s, the most popular approach has been the SCP with CG embedded in branch-and-bound (see Desrosiers and Lübbecke 2005; Desrosiers et al. 1995; Barnhart et al. 1996). This method will be discussed in Sect. 7.

Barnhart et al. (2003) provide an excellent literature review and a detailed survey of crew pairing problems. Gopalakrishnan and Johnson (2005) give a survey of different approaches and solution methodologies for airline crew scheduling problems. Klabjan (2005) surveys the models and algorithms for large-scale airline planning and operational problems, giving extensive references. More recently, researchers have explored the monthly pairing problem, the personalized assignment problem, and integrated crew scheduling. In this paper, we survey the state of the art of studied approaches and solution methodologies for the airline crew scheduling problem.

In Sect. 4.1, we provide a comprehensive review of the sequential crew planning problem, and in Sect. 4.2 we describe the models and algorithms for the integrated crew planning problem.

4.1 Sequential crew planning

Section 4.1.1 discusses the crew pairing problem and Sect. 4.1.2 discusses the crew assignment problem.

4.1.1 Crew pairing

For a given crew category and fleet type, the crew pairing problem finds a set of minimum-cost pairings such that each scheduled flight over the time horizon is included in exactly one pairing. The solution approach depends on the airline's size, network structure (e.g., hub-and-spoke), rules and regulations, and cost structure.

There are three traditional approaches. The *daily* problem assumes that the air legs are identical for all the days of the planning horizon, and the minimum-cost pairings are generated based on the scheduled legs for a single day. The *weekly* problem assumes that the air legs are repeated every week, and the pairing problem is solved based on the scheduled air legs for 1 week. The *monthly* problem has a time horizon of a full month. Recent research has focused on the weekly and monthly problems. Because of vacation periods and variations in the flight schedules, the monthly time horizon is the most realistic.

The crew pairing problem is typically formulated as an SPP or SCP in which each task (air leg) is a constraint and each feasible pairing is a variable. There are usually additional constraints that enforce the various restrictions, safety rules, and regulations such as the maximum flying time for each base. The number of feasible pairings is extremely large, so it is often impossible to consider all of them. The problem has therefore traditionally been treated in two steps: the first step generates a subset of good pairings by enumeration, and the second uses the SPP to select the best pairings of this subset. Initially, heuristic local-search algorithms were often used; a typical example is presented by Marsten and Shepardson (1981). They present an SPP model and a solution technique based on Lagrangian relaxation and subgradient optimization. They report tests on data sets from the Flying Tiger Line, Pacific Southwest Airways, Continental Airlines, and Helsinki City Transport. Gershkoff (1989) presents an SPP that minimizes the cost for the daily pairing problem. Gershkoff introduces a heuristic algorithm in which possible pairings are constructed at each iteration for a subset of the scheduled air legs. The heuristic continues until no further improvement is possible or a stopping criterion such as a time restriction is satisfied. Results are presented for American Airlines. Anbil et al. (1991) use this approach in software for American Airlines called the Trip Reevaluation and Improvement Program (TRIP). They explain the advances that allow the software to solve large problems more quickly. This software was sold to ten other airlines.

Anbil et al. (1992) introduce a global approach with a cost-minimization objective. In this joint study by IBM and American Airlines Decision Technologies, millions of feasible pairings (columns) are enumerated a priori and several thousands are provided to the LP solver. At each iteration, most of the nonbasic columns are discarded and new columns are added. The process continues until all the columns have been taken into account. Bixby et al. (1992) combine the interior point and simplex methods to find the solution of the LP relaxation for very large problems. The experiments show that the hybrid approach is more efficient than applying either method separately. Hoffman and Padberg (1993) propose a branch-and-cut approach for an SPP in which pairings are generated heuristically a priori, and cuts are used to find an integer solution. They present results for 68 data sets from four major airlines. The integrality gaps are large (up to 5 %) and much effort is necessary to obtain a good integer solution. Klabjan et al. (2001) improve on the approach of Hoffman and Padberg (1993). They enumerate millions of random pairings. The relaxation is first solved and millions of columns are selected based on their reduced costs. The number is then reduced by a heuristic based on LP, and integer solutions are obtained using a commercial IP solver. The branching rule is

enhanced by combining strong branching with a specialized branching rule. Experiments are reported for a large US airline.

Heuristic approaches have three major problems. First, they do not consider all the scheduled flights at once, and they usually perform several iterations before finding a reasonable solution. Second, they do not consider all the possible pairings. Third, they provide no information on how far the solution is from the optimal solution. Therefore, more sophisticated approaches for pairing generation have been proposed. Lavoie et al. (1988) present an SCP and propose an algorithm for the continuous relaxation of the problem based on generalized LP, generating columns via shortest-path subproblems. Experiments for instances with up to 329 air legs give good results. Desrosiers et al. (1991), Barnhart et al. (1994), and Desaulniers et al. (1997, 1998) use CG to consider all the pairings instead of a subset of a priori-generated pairings. They propose a dynamic CG approach (and a branch-and-price algorithm) that implicitly considers all possible pairings when solving the LP relaxation; we explain this approach in more detail later. The integrality gaps become smaller (less than 1 %) when all the feasible pairings are considered.

Wedelin (1995) introduces a dual coordinate search together with an approximation algorithm for large-scale 0–1 integer problems. The approximation scheme adjusts the costs as little as possible such that the new problem has an integer solution. The degree of approximation is determined by a parameter, and for different values of this parameter the algorithm can be interpreted in terms of LP, dynamic programming, or a greedy approach. It is applied to large-scale data sets extracted from the Carmen system (the model is the SCP). The results show that the algorithm compares well with CPLEX in terms of both computational time and solution quality. Andersson et al. (1998) give a general overview of the Carmen pairing-construction system that is used at most major European airlines. This system has been successful because it integrates manual and automatic approaches to scheduling, provides high-quality and robust optimization, and is user friendly.

Vance et al. (1997) consider a model for a two-stage decision making process for the crew scheduling problem. In the first stage, they select a set of duty periods that cover the scheduled flights. They then construct the pairings based on these duty periods. They use a decomposition approach based on dynamic CG. Experiments are presented for a major US domestic airline: the new approach provides tighter LP bounds, but the solution process is more difficult. Barnhart and Shenoi (1998) solve an approximate model of the crew pairing problem and use the solution as the initial solution for conventional approaches. Promising results are presented for a long-haul airline. Hu and Johnson (1999) propose a primal-dual subproblem simplex algorithm to speed up the solution of the LP relaxation. Experiments are reported for instances with up to 930 air legs.

Hjorring and Hansen (1999) propose a black-box rule system to simplify the implementation of the various rules and regulations. They integrate CG with a pricing subproblem, based on a duty network and a kth shortest path algorithm. Results are provided for some realistic data sets. Subramanian and Sherali (2008) propose an effective deflected subgradient optimization scheme for generating good dual solutions for the LP problems. This approach, used together with CG, is embedded in the crew pairing solver at United Airlines. Tests using historical data

sets show that significant benefits can be obtained using this approach instead of a standard solver for the intermediate LPs. AhmadBeygi et al. (2009) consider a new IP approach that is easy to implement, facilitating prototyping and the testing of new ideas. Their proposed model uses connection variables and market variables to capture the nonlinear cost function and constraints. Results for data sets from a major US hub-and-spoke carrier demonstrate the performance of the approach. Dück et al. (2011) present a column- and cut-generation algorithm. The objective is to minimize the number of pairings to cover a set of scheduled flights with respect to the regulations, ignoring the duty duration. The problem is formulated as an SPP with shortest-path resource-constrained subproblems. This algorithm has been applied to some small and medium instances from a European airline. Saddoune et al. (2013) introduce a rolling-horizon approach to find the minimal-cost pairings for the SPP formulation of the monthly pairing problem. They discuss the weaknesses of the traditional approach in which daily, weekly, and monthly problems are solved sequentially. Experiments for a major short- and medium-haul US airline give good results.

4.1.2 Crew assignment

There are several possible objectives for the problem of constructing monthly schedules (the crew assignment problem). Compared to crew pairing, crew assignment has received less attention.

In the context of the bidline assignment problem, Beasley and Cao (1996) present an IP formulation; they use Lagrangian relaxation and subgradient optimization. This approach is embedded into a tree search to find the optimal solution. Results are provided for randomly generated test instances with up to 204 crew members and 500 tasks. Campbell et al. (1997) describe a bidline generator system for a US express transportation company (FedEx). The goal is to minimize the number of bidlines and the amount of flying time not assigned to bidlines. They use a metaheuristic algorithm based on simulated annealing. Jarrah and Diamond (1997) propose a heuristic SPP-based approach for the bidline assignment problem using a priori CG. The objective is to maximize the covered credit time while minimizing the number of bidlines. The system is semi-automatic: the user influences the subset of columns generated. This system is implemented in a major US airline, and good solutions are reported. Christou et al. (1999) introduce a two-phase approach based on genetic algorithms for bidline generation at Delta Air Lines. The objective is to maximize the average total value and the quality of the bidlines. The first phase of the algorithm aims to construct good bidlines, while the second phase completes the assignment by constructing valid bidlines, taking into account the pairings not covered at the first phase. The results, for up to 320 crew members, show that the algorithm provides significant savings compared to the semi-automated approach to bidline generation used at Delta Air Lines.

Weir and Johnson (2004) propose a three-phase approach for bidline generation. In the first phase, a mixed integer problem is solved to provide tentative bidlines (patterns). The second phase, based on an SPP, uses these bidlines to find final schedules that cover all the pairings. If phase two is not successful, phase three

integrates the uncovered pairings into the schedules. Good results for up to 150 crew members are presented. Boubaker et al. (2010) describe two heuristic algorithms for solving the SPP-based bidline scheduling problem where, as far as possible, each bidline should have the same number of days off and paid hours (bidline with equity). The first heuristic is a standard branch-and-price algorithm that relies on a rounding procedure to obtain integer solutions. The second algorithm combines dynamic constraint aggregation (Elhallaoui et al. 2005) with the first heuristic. The results show that for the largest instances (up to 564 pilots and 2924 pairings), the dynamic constraint aggregation heuristic gives solutions better than those of the standard branch-and-price heuristic.

For the rostering problem, Day and Ryan (1997) consider Air New Zealand's short-haul operations. In their approach, using integer programming, the days off are first allocated and then the pairings and other activities are assigned. The method leads to the efficient construction of good-quality schedules since most of the pairings are 1-day assignments, and it has been used since 1993 for all short-haul flight attendant rosters at Air New Zealand. Gamache et al. (1999) describe a generalized SPP and a heuristic approach based on CG to find good integer solutions for the problem of constructing personalized schedules while maximizing satisfaction and considering pre-assigned activities. They use control strategies in the CG to reduce the computational time for large problems. Results for medium instances from Air France demonstrate the quality of the solutions, in terms of both cost and computational time. The acceleration strategies reduce the computational times by a factor of more than 1000. The resulting schedules are compared with schedules constructed by the CADET program (then in use at Air France), and the new schedules had fewer uncovered duties. El Moudani et al. (2001) propose a heuristic bi-criterion approach that takes into account the satisfaction of the crew members. This approach is combined with a genetic algorithm to produce minimum-cost schedules that achieve a specific level of crew satisfaction. Results for data from a medium-haul airline are given. König and Strauss (2000a, b) introduce a heuristic that implicitly enumerates schedules using propagation techniques. This approach is implemented in the SWIFTROSTER algorithm, and good results are achieved for medium and large European airlines. Fahle et al. (2002), Kohl and Karisch (2000), and Sellmann et al. (2002) describe the Parrot project (1997), which is based on CG and constraint programming. The master problem (the selection of schedules) is solved as an LP, and constraint programming is used to prune the search. Kohl and Karisch (2004) provide a comprehensive study of the Carmen crew rostering system that has been used at KLM since 1995. They highlight practical considerations relating to the production settings of crew scheduling systems. Maenhout and Vanhoucke (2010) use Dantzig–Wolfe decomposition and propose a metaheuristic scatter search algorithm to assign personalized rosters to each crew member while minimizing the total operational cost and achieving a required schedule quality. They compare their method with an exact solution approach based on branch-and-price and steepest-descent variable neighborhood search. Results are given for instances with up to 150 pilots and 800 pairings.

For the seniority-based personalized assignment problem, Gamache et al. (1998) study the preferential bidding problem, considering seniority for the assignment of

personalized schedules to pilots and officers. At Air Canada about 75 different bids, e.g., for a weekend off, are available. Each employee associates a weight with each bid and these weights are used to calculate a score for the potential schedules. A *residual* problem is then solved for each employee, from the most senior to the most junior. This determines the maximum-score schedule for the current employee, taking into account the other employees and the set of unassigned pairings. The problem is modeled as an IP and is solved by CG embedded in a branch-and-bound tree. Results are presented for twenty-four data sets from Air Canada. This research resulted in a preferential bidding system (PBS) that has been used at Air Canada since 1995. Achour et al. (2007) introduce the first exact solution approach based on CG for the PBS. They solve a sequence of LPs in seniority order, and the schedules of the employees are fixed as the algorithm progresses. When a tentative maximum score for a crew has been established, they explicitly enumerate all the feasible schedules with that score for that crew member. Results for real data sets with up to 91 pilots indicate a substantial improvement in the solution quality.

4.2 Integrated crew planning

As explained in Sect. 3.1, solving the airline planning problem sequentially does not guarantee an optimal result for the overall planning problem. Recently, researchers have investigated integrating some of the steps. A summary of research on the integration of crew pairing with aircraft routing and/or fleet assignment is provided in Sect. 4.2.1. Section 4.2.2 discusses the integration of crew pairing and assignment.

4.2.1 Integrated crew pairing, aircraft routing, and fleet assignment

Cordeau et al. (2001), Cohn and Barnhart (2003), Mercier et al. (2005), and Chen et al. (2012) address the integration of aircraft routing and crew pairing. Sandhu and Klabjan (2007) and Gao et al. (2009) consider the integration of fleet assignment and crew pairing. Klabjan et al. (2002) consider the integration of flight scheduling, aircraft routing, and crew pairing. Mercier and Soumis (2007) integrate the aircraft routing, crew scheduling, and flight retiming problems. Papadakos (2009) proposes various integration models for the fleet assignment, maintenance routing, and crew pairing problems. Ruther (2010) studies the integration of aircraft routing, crew pairing, and tail assignment.

4.2.2 Integrated crew scheduling

Guo et al. (2006) introduce a heuristic algorithm for partially integrating crew pairing and crew assignment. It is based on aggregated time-space networks where the crew members are unevenly stationed among bases, and their availability changes dynamically during the planning period. The algorithm first generates a group of pairings that are separated by weekly rests. Some parts of these pairings are then rearranged into individual crew schedules. Results for a European airline indicate a significant reduction in the cost of the schedules. Zeghal and Minoux

(2006) integrate crew pairing and crew scheduling for technical crew members and for airlines with short- and medium-haul air legs. They replace a large number of binary constraints with a smaller number of stronger constraints (clique constraints), which improves the computational time and solution quality. To improve the efficiency of the exact methods, a heuristic method based on a rounding strategy is embedded in a partial tree-search procedure. Results for TunisAir, where it is possible to enumerate all the feasible duties, indicate that good solutions can be obtained in a reasonable computational time. Souai and Teghem (2009) propose a hybrid genetic algorithm to integrate crew pairing and crew assignment. They use three heuristics to deal with the regulations. Results for three data sets demonstrate the success of the approach.

Deng and Lin (2011) use the ant colony optimization algorithm to solve the crew scheduling problem. They formulate the problem as a traveling salesman problem on a weighted and constrained graph. Results for small real instances indicate that good solutions can be found. Saddoune et al. (2012) integrate the crew pairing and bidline crew assignment problems for pilots; the objective is to minimize the total cost and the number of pilots. They combine the dynamic constraint aggregation of Elhallaoui et al. (2005) with CG. Results for a major short- and medium-haul US airline show that the approach gives significant savings, but the computational times are higher than in the sequential approach. Saddoune et al. (2011) reduce the computational time of the algorithm of Saddoune et al. (2012). They introduce a bi-dynamic constraint aggregation method that uses the neighborhood structure when generating columns for the CG. The results confirm the reduction in the computational time. Azadeh et al. (2013) introduce a hybrid metaheuristic for the nonlinear optimization of a crew scheduling problem in which the objective is to minimize the total crew cost while respecting the regulations. They also propose two hybrid algorithms based on the genetic algorithm and ant colony optimization. They perform experiments on twenty randomly generated data sets of various sizes.

5 Regulations for airline crew scheduling

Airlines must respect many regulations during the crew scheduling process. These regulations have three main sources (Barnhart et al. 2003). Many are imposed by governing agencies (e.g., FAA in the US) to ensure safety. Labor unions often influence the working conditions of the crew members. Finally, airlines impose some conditions; for example, they may restrict the set of feasible solutions to obtain more robust schedules. We now discuss typical regulations on the duties, pairings, and schedules. In Sect. 7, we discuss the regulations imposed in our crew scheduling model.

5.1 Duties

There must be idle time between any two sequential air legs to allow for connections. There are lower and upper bounds on this interval. Briefing and debriefing times are often required at the beginning and end of each duty. There are

also strict constraints on the total number of flying hours in a duty, and there is usually a maximum number of landings per duty.

5.2 Pairings

There is typically a maximum number of duties in a pairing, a minimum and maximum duration of the layovers between duties, a maximum TAFB, and a maximum pairing duration. In addition, all pairings should begin and end at a base. FAA imposes an 8-in-24 rule, which means that extra rest is required if a pairing contains more than eight flying hours in any 24-h period.

5.3 Monthly schedules

There are typically restrictions on the maximum number of flying hours per month, the minimum and maximum number of working days, the minimum number of days off, etc. There may be additional constraints relating to the needs and preferences of the crew members.

6 Description of personalized pilot assignment problem

The problem of constructing pairings and assigning personalized schedules to the crew members varies from one airline to another, depending on the regulations, the pre-assigned activities, and the extent to which the preferences of the crew members are taken into account. The pre-assigned activities are training periods, annual leave, medical appointments, and pre-assigned vacations. An example of an assignment problem for which the pre-assigned activities are taken into account is given by Gamache et al. (1999) for cabin crew members. The crew preferences may contain a large number of factors. The factors considered may vary from one company to another, and the objectives may also vary. The objective can be a weighted sum of factors or may be based on equity between crew members with priority given to the most senior employees (Gamache et al. 1998). We will present a simple case and a basic algorithm to solve large-scale problems with heuristics to speed up the solution process.

We solve the personalized monthly assignment problem for a fixed number of pilots based on a set of anonymous pairings. These pairings are the results of the experiments of Saddoune et al. (2013). The number of pilots comes from the solution of the bidline assignment problem in the sequential context by Saddoune et al. (2012). We assume that the airline allows two types of preferences: *vacations* and *preferred flights*. Each month, some of the pilots are asked to specify their preferred vacation period (if they wish to take a vacation in that month). The other pilots will be invited to enter vacation requests in other months. Each pilot also has the option of choosing a set of preferred air legs from the scheduled flights corresponding to the pairings associated with his/her base. The details of these two categories of preferences are given in Sect. 8.2.

We use the sequential approach that is common in the airline industry. Our approach differs from that of Saddoune et al. (2012) because we consider crew preferences. The problem of constructing anonymous pairings using a rolling-horizon approach has been solved by Saddoune et al. (2013). They also consider the problem of sequential bidline scheduling (Saddoune et al. 2012) in which they construct anonymous schedules for pilots that minimize the total cost and the number of pilots per base. They provide their results for the set of instances that we consider, and we use their results for the anonymous set of pairings (Saddoune et al. 2013) and the number of pilots per base obtained from solving the bidline assignment problem (Saddoune et al. 2012) for our personalized assignment problem. Furthermore, we use the same solution methodology based on CG, but we solve an enhanced mathematical model for the personalized monthly assignment problem. We assume a fixed number of pilots per base. Given the pilot preferences and the anonymous pairings, we construct personalized schedules that cover the pairings, minimize the total crew cost, and satisfy a minimum number of the preferences (preferred air legs and vacations).

Instead of considering the preferences and minimizing the cost in two separate steps, we combine the two factors into a single objective function to provide a unique SCP formulation for our problem. Section 6.1 introduces the notation that we use and Sect. 6.2 discusses the personalized assignment problem.

6.1 Notation

The notation is as follows:

Sets

F set of all scheduled flights to be covered;
L set of all pilots;
V_l set of preferred vacations for pilot $l \in L$;
P set of pairings;
S_l set of all feasible personalized schedules for pilot $l \in L$;
B_l set of preferred flights for pilot $l \in L$;

Parameters

C_p cost of pairing $p \in P$;
$\bar{C}_f$ penalty cost for not covering flight $f \in F$;
C_s^l cost of schedule $s \in S_l$ for pilot $l \in L$;
n_s^l number of preferred flights in schedule $s \in S_l$ for pilot $l \in L$;
c_f^l bonus cost for covering preferred flight $f \in B_l$;
c_v^l penalty cost for not covering preferred vacation $v \in V_l$;
u minimum number of preferred flights in schedules;
w minimum number of preferred vacations to be covered;

$$e_p^{s,l} = \begin{cases} 1 & \text{if} \quad \text{pairing } p \in P \text{ is covered by pilot } l \in L \text{ in personalized schedule } s \in S_l \\ 0 & \text{otherwise;} \end{cases}$$

$$\bar{e}_p = \begin{cases} 1 & \text{if} \quad \text{flight } f \in P \text{ is not covered} \\ 0 & \text{otherwise;} \end{cases}$$

$$e_f^p = \begin{cases} 1 & \text{if} \quad \text{flight } f \in F \text{ is covered by pairing } p \in P \\ 0 & \text{otherwise;} \end{cases}$$

$$v_v^{s,l} = \begin{cases} 1 & \text{if} \quad \text{vacation } v \in V_l \text{ for pilot } l \in L \text{ is covered by schedule } s \in S_l \\ 0 & \text{otherwise;} \end{cases}$$

Variables

$$x_l^s = \begin{cases} 1 & \text{if} \quad \text{schedule } s \in S_l \text{ is chosen for pilot } l \in L \\ 0 & \text{otherwise;} \end{cases}$$

6.2 Personalized assignment problem

Given a minimum-cost set of pairings that covers all the scheduled flights in the planning horizon, the personalized pilot assignment problem finds minimum-cost schedules such that the pairings are covered exactly once and at least given numbers of the flight and vacation preferences are satisfied.

The cost of each schedule has fixed and variable components. The variable costs are the costs of covering the pairings, bonuses for satisfying the preferred flights, and penalties for not covering the preferred vacations; the fixed costs are the fees paid to the pilots. The costs of the pairings are given by Saddoune et al. (2012).

A global constraint is defined to ensure that a minimum number of the vacation preferences are satisfied. We add a bonus factor (a negative cost) for each flight preference that a schedule includes. The cost of covering the air legs is set to 0. The cost of personalized schedule s for pilot l is

$$C_s^l = \sum_{p \in P} e_p^{s,l} C_p + n_s^l . c_f^l + \sum_{v \in V_l} (1 - v_v^{s,l}) . c_v^l.$$

The SCP formulation is

$$\text{Minimize} \sum_{l \in L} \sum_{s \in S_l} C_s^l x_l^s + \sum_{f \in P} \bar{e}_p \bar{C}_f \tag{1}$$

$$\text{Subject to:} \sum_{l \in L} \sum_{s \in S_l} e_p^{s,l} x_l^s + \bar{e}_p = 1, \quad \forall p \in P \tag{2}$$

$$\sum_{l \in L} \sum_{f \in B_l} \sum_{s \in S_l} \sum_{p \in P} e_f^p e_p^{s,l} x_l^s \geq u \tag{3}$$

$$\sum_{l \in L} \sum_{s \in S_l} \sum_{v \in V_l} v_v^{s,l} x_l^s \geq w \tag{4}$$

$$\sum_{s \in S_l} x_l^s \leq 1, \quad \forall l \in L \tag{5}$$

$$x_l^s \in \{0,1\}, \forall l \in L, \quad \forall s \in S_l \tag{6}$$

The objective function (1) minimizes the total cost of the schedules and penalty costs for uncovered flights. Constraint (2) ensures that each pairing is covered exactly once. Constraint (3) is a global constraint on the minimum number of preferred flights. Constraint (4) is a global constraint on the minimum number of satisfied vacation preferences. Constraint (5) ensures that at most one schedule is chosen for each pilot, and constraint (6) is the integrality condition.

7 Algorithm

Because of the large number of variables, we apply CG, which is embedded in a branch-and-bound scheme (branch-and-price) to obtain integer solutions. This method was pioneered by Desrosiers et al. (1995), Barnhart et al. (1996), and Desrosiers and Lübbecke (2005). Optimality can be achieved for small problems, and a near-optimal solution (usually within 1 % of optimality) can be found for large problems. The rules and regulations for constructing monthly crew schedules (as explained in detail in Sect. 5) are simplified to make it possible to obtain optimal solutions for the LP relaxations (for instance the 8-in-24 rule is not implemented). The computational time to optimally solve the LP relaxation of the simplified problem is similar to the time used in industry to approximately solve the complete problem. This is a good choice for benchmark because it removes the violations on results due to approximate solutions and permits to have clear view of the affects of algorithms and their parameters.

In Sect. 7.1, we describe the CG, and in Sect. 7.3 we discuss the approach used to find integer solutions.

7.1 Column generation

CG is considered one of the most significant advances in the solution of large-scale linear mixed integer models (Klabjan 2005). It is an optimal iterative method for LPs with a large number of variables.

The linear relaxation of the personalized assignment problem (1)–(6) is called master problem (MP). At each iteration, we consider a restricted master problem (RMP) that contains a subset of the variables (columns). The RMP is solved by a standard LP algorithm such as the simplex method and finds an optimal objective-function value and a pair of primal and dual solutions. Given the optimal dual solution from the RMP, the current subproblem tries to find columns with negative reduced costs. If such columns are found, they are added to the RMP for the next

iteration. Each subproblem corresponds to a resource-constrained shortest path problem and is usually solved by dynamic programming. When no variable with a negative reduced cost can be found, the optimal solution for the RMP is optimal for the MP. In practice, the CG is often stopped before optimality is attained because of slow convergence (the tailing-off effect).

In Sect. 7.2, we describe the subproblems for the personalized assignment problem. Section 7.3 discusses our method for finding integer solutions.

7.2 Personalized assignment subproblem

There is one CG subproblem for each pilot in the personalized assignment problem. The subproblem is defined on a directed acyclic time-space network $G^l = (N^l, A^l)$, where N^l and A^l represent the node and arc sets for pilot l, respectively. Each network corresponds to a shortest path problem with resource constraints, and the goal is to find schedules with negative reduced costs. These subproblems are solved using a label-setting algorithm (Irnich and Desaulniers 2005). The network G^l is an enhancement of the bidline assignment network proposed by Saddoune et al. (2012) in the sequential context. It contains the arcs for preferences. Figure 2 gives a partial illustration of the network (some of the nodes and arcs have been omitted).

The network has five node types. The unique *source node* and *sink node* represent the start and end of the schedules, respectively. Each pairing is specified by a *pairing start node* and a *pairing end node*. *Midnight nodes* specify the midnights of the planning horizon. The horizon starts and ends at midnight.

There are eight arc types. A single *schedule start arc* connects the source node to the first midnight node of the horizon and represents the start of the schedule; its cost is 0. *Pairing start arcs* link each potential midnight node to the pairing start node at the base station; their cost is 0. For each pairing, a *pairing arc* links the node at the beginning of the pairing to the node at the end of the pairing. Its cost is calculated based on the pairing cost function of Saddoune et al. (2012). A *preferred vacation arc* links two midnight nodes in the pilot's network if he/she has a vacation preference; its cost is 0. We assume that the vacations start and end at the pilot's base station and that vacations occur between two midnights (e.g., a vacation may start at 00:00 on the second day of the month and finish at 00:00 on the tenth day). A *post-pairing rest arc* links the end pairing node at the base station to the start node of a subsequent pairing provided the intervening time is greater than the post-pairing rest. A *post-pairing arc* links the end node of the pairing to the first midnight node at the base station that permits a complete day off. A *day-off arc* links a pair of consecutive midnight nodes at the base station. The cost of these three arcs is 0. Finally, a *schedule end arc* connects the last midnight node in the horizon to the sink node and represents the end of the schedule.

In addition to constraints (2)–(6), the local restrictions are enforced via the resource constraints. There is a lower bound on the minimum number of days off in a schedule and a maximum number of consecutive working days. There is also a maximum credited flying time per schedule (85 h for our tests).
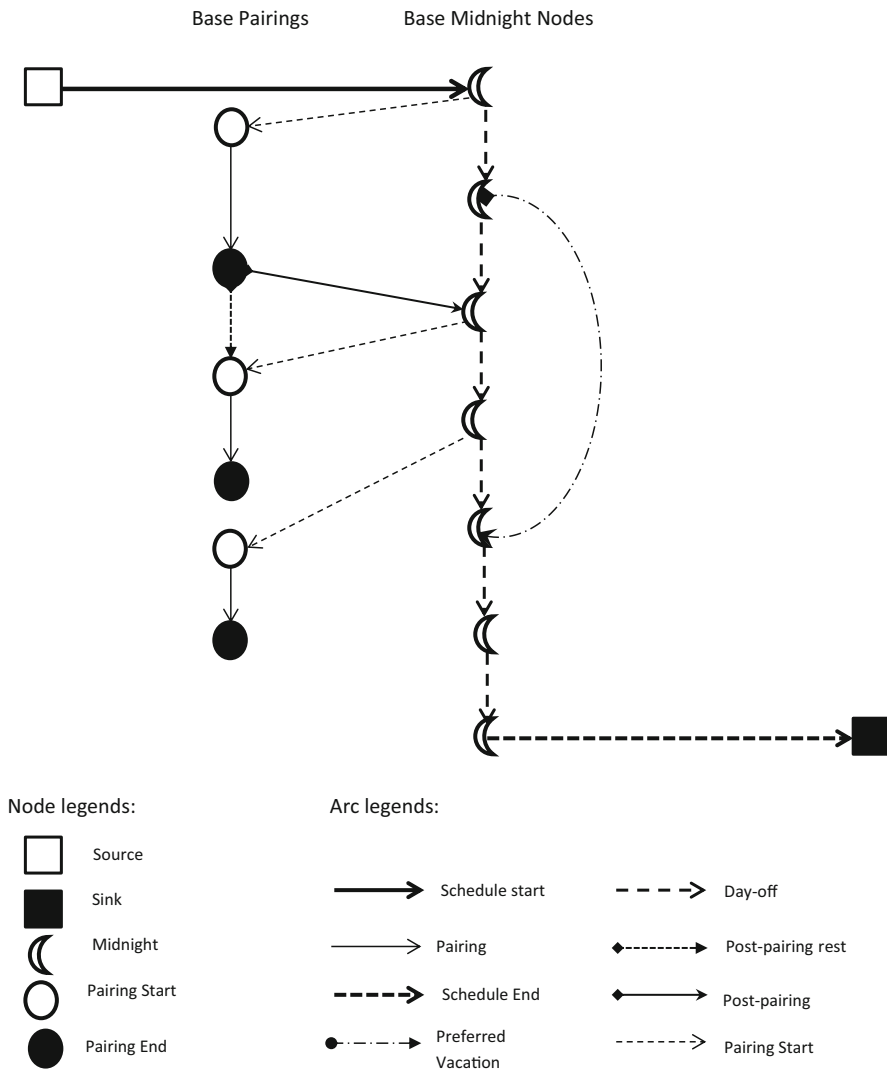
Fig. 2 Directed acyclic time-space network for personalized assignment subproblem

The cost of a feasible path starting from the source node and ending at the sink node corresponds to the cost of a schedule and is equal to the sum of the arc costs on the path. However, the subproblem tries to find paths with a negative reduced cost so the arc costs should sum to the reduced cost of the associated pairing. Therefore, the arc costs are updated based on the dual variables and the coefficients of the variables in the MP constraints. The pricing in the personalized assignment subproblems is carried out via a multi-label shortest path algorithm (Desrosiers et al. 1995; Irnich and Desaulniers 2005). A label is associated with each partial

path originating from the source node. This label has a component for each resource and one for the reduced cost. The resource components give the value of the resource at the last node of the partial path, and the reduced-cost component indicates the reduced cost of this path. The initial labels are set to 0 at the source node. Before adding a new arc m to a partial path, we check that the resource labels are within the upper and lower bounds of the resource windows at the start node of arc m. When adding an arc to an existing partial path, we recalculate the labels via extension functions, and the arcs with infeasible resource values are discarded. For large networks, the number of labels can be extremely large, and this can lead to long computational times at each CG iteration. To avoid this, we use a label dominance rule to discard suboptimal paths: label L_1 is dominated by label L_2 if both the resource and reduced-cost components of L_1 are less than or equal to those of L_2 (at least one inequality should be strict).

7.3 Integer solution

After solving the linear relaxation of (1)–(6), we must find an integer solution. We embed the CG in a heuristic branch-and-bound scheme: at each branch-and-bound node, the lower bounds are calculated by CG. A fixing procedure is used to impose permanent decisions. Of the five branching strategies implemented in the GENCOL software that we use (two optimal strategies and three heuristics), we apply the two heuristic strategies: column fixing and heuristic inter-task fixing. Column fixing simply fixes a subset of the variables (columns) of the MP to 1 when their score is greater than or equal to a predetermined threshold. At each branching node, a score is calculated for each potential branching strategy. Through two parameters, the maximum and minimum scores, we map these scores to a scaled score. We compare the scaled scores and select the branching strategy with the better score. We use 0.85 as the threshold. If no suitable variable exists, two tasks (air legs for the pairing problem or pairings for the assignment problem) are assigned to the same pairing/ schedule and are performed consecutively.

8 Computational results

In this section, we present results for the personalized pilot scheduling problem based on monthly flight schedules operated by short- or medium-haul aircraft. Seven data sets are provided by a major North American airline. In Sect. 8.1, we describe the seven instances. Section 8.2 explains how we generate the preferences, and Sect. 8.3 gives the results for the personalized assignment problem.

The anonymous pairings are extracted from the results of the rolling-horizon/CG approach of Saddoune et al. (2013). They consider a 3-day time slice and a 1.5-day overlap between two consecutive slices. We determine the number of pilots for each instance from the results of Saddoune et al. (2012) for the sequential bidline scheduling problem.

8.1 Description of data sets

All 7 instances have 3 crew bases, and the number of flights ranges from 1013 to 7765. The number of stations varies between 26 and 54. Table 1 shows the characteristics of these instances. For the feasibility of the pairings, Saddoune et al. (2012) use the parameter values of Mercier et al. (2005), and for the feasibility of the schedules, they use the parameter values of Boubaker et al. (2010). We use the same values for our tests. In addition, for the personalized assignment problem, we define some new parameters including a bonus for covering preferred flights and a penalty for failing to cover the preferred vacations. The costs of failing to cover flights, the preferred vacations, the preferred flights, and the cost of covering flights are all related. To show the relationship between covering flights and the preferences, we use the following example. We assume that it is more important to cover the scheduled flights than to satisfy the vacation and flight preferences. We assume that the cost of covering a scheduled flight is 0, and we set the cost of failing to cover a scheduled flight to 10,000. The results show that this cost is large enough to ensure that the percentage of uncovered flights is very small. The cost of failing to satisfy a vacation preference is set to 1000. This cost is large enough to ensure appropriate satisfaction of the requested vacations and to avoid a large number of uncovered flights. We apply a negative cost (a bonus) of -100 for covering a flight preference to help satisfy the global constraint on the minimum number of preferred flights. It is important to mention that the costs of preferences are not very sensitive parameters; constraints (3) and (4) ensure a minimum number of preferred vacations and preferred flights.

We did not have access to real preference data for these data sets, so we randomly generated simple pilot preferences with parameters based on the expertise of analysts who have worked on data from more than 20 airlines. We now discuss the random generators.

8.2 Random generators

We developed a random generator for the vacation preferences and another for the flight preferences, assuming that the scheduling month is a typical month outside of the high season.

Table 1 Characteristics of instances

	No. of scheduled flights	No. of stations
$I1 - 727$	1013	26
$I2 - DC9$	1500	35
$I3 - D94$	1854	41
$I4 - D95$	5613	49
$I5 - 757$	5743	34
$I6 - 319$	5886	52
$I7 - 320$	7765	54

8.2.1 Vacation generator

In this random generator, the vacations are uniformly distributed during the month. Since the scheduling month is not a month with special events (e.g., Thanksgiving, Christmas), this assumption is reasonable. Each month, about 30 % of the pilots, with variations due to integer rounding, request a vacation, and the duration ranges from 2 to 15 days. These numbers correspond to values observed in the industry.

Our experiments show that we can satisfy at least 38 % of these requests; a penalty is associated with each unsatisfied request. The vacation days are consecutive, and the vacations start and end at the pilot's base.

8.2.2 Flight preference generator

In this random generator, each pilot selects a specific percentage of preferred flights from the set of scheduled flights included in the pairings corresponding to his/her base. The selection follows a uniform distribution. For our experiments, we assume that each pilot prefers 10 % of the scheduled flights.

We ensure that at least 20 % of the flights contained in the pilots' schedules are preferred flights. We do not specify a minimum acceptable percentage of preferred flights for each pilot, because the pairings are constructed anonymously and the personalized schedules are constructed based on these pairings.

8.3 Summary of results

Section 8.3.1 presents a summary of the results of Saddoune et al. (2012) for the pairing and bidline assignment problems. Section 8.3.2 discusses the results of solving the personalized assignment problem.

8.3.1 Pairing and bidline assignment problem

Table 2 presents the results of Saddoune et al. (2012) for the pairing and bidline problems. They conducted their experiments on a Linux PC equipped with an Intel Xeon processor clocked at 2.8 GHz, using version 4.5 of GENCOL and version 10.1 of CPLEX. We use the pairing results to clarify the set of pairings for each instance

Table 2 Results for pairing and bidline assignment problems

	No. of pairings	No. of pilots	CPU time (pairing)	CPU time (bidline)
$I1 - 727$	172	33	2.50	1.50
$I2 - DC9$	303	34	4.34	1.46
$I3 - D94$	274	47	9.14	2.26
$I4 - D95$	1079	145	393.58	129.02
$I5 - 757$	1497	247	67.80	164.10
$I6 - 319$	1187	223	154.75	105.25
$I7 - 320$	1648	305	289.22	218.38

when constructing monthly personalized schedules. We also use the bidline results on the total number of pilots for each instance to approximate the number of pilots per base. In contrast to the model of Saddoune et al. (2012), our model does not minimize the number of pilots. Instead, we assume a fixed number of pilots per base. Given a list of preferences for each pilot, we want to satisfy their preferences while covering the pairings. The CPU times are given in minutes.

8.3.2 Personalized assignment problem

We conducted our tests on a Linux PC with an Intel(R) Core(TM) processor clocked at 3.40 GHz. All of our implementations are coded in C++, and version 4.5 of GENCOL is used. The RMPs are solved using CPLEX 12.4. Table 3 gives the number of pilots per base. As mentioned, we did not minimize the number of pilots and sometimes the total number of pilot schedules is slightly lower than the number of pilots assigned to a base; the pilots without schedules become reserve crew members. Table 4 summarizes the solution process. *Number of pilots* indicates the number of subproblems for each instance. We have many more subproblems (one for each pilot) than Saddoune et al. (2012) deal with in the context of sequential bidline scheduling (three subproblems, one for each base). *Number of CG iterations* indicates the total number of CG iterations. *Number of branching nodes* indicates the total number of branching nodes in the branch-and-price scheme. *Gap* gives the percentage difference between the lower bounds (LP solutions) and upper bounds (integer solutions). *CPU time (in minutes)* indicates the total CPU time.

Table 3 Number of pilots per base

	B1	B2	B3	Total
I1 − 727	7	20	6	33
I2 − DC9	10	9	15	34
I3 − D94	10	30	7	47
I4 − D95	42	78	25	145
I5 − 757	141	101	6	247
I6 − 319	117	66	40	223
I7 − 320	158	96	51	305

Table 4 Summary of solution process for personalized assignment problem

	No. of pilots	No. of CG iterations	No. of branching nodes	Gap (%)	CPU time (min)
I1 − 727	33	239	7	0.00	0.16
I2 − DC9	34	1968	30	0.18	0.45
I3 − D94	47	466	14	0.02	1.81
I4 − D95	145	2417	129	0.38	47.59
I5 − 757	247	4531	172	2.91	149.58
I6 − 319	223	2975	168	0.49	75.31
I7 − 320	305	4011	195	0.37	184.76

Table 5 Quality of Solutions

	Uncovered flights (%)	ACFT	Preferred vacations (%)	Preferred flights (%)
$I1 - 727$	0.00	68.75	15.15	25.90
$I2 - DC9$	0.00	75.76	14.71	24.27
$I3 - D94$	0.00	72.70	14.89	26.66
$I4 - D95$	0.04	74.70	14.48	36.37
$I5 - 757$	1.81	81.12	12.55	25.00
$I6 - 319$	0.14	76.75	14.35	39.13
$I7 - 320$	0.86	81.65	10.16	26.32
Average	0.48	75.92	13.76	29.09

Table 5 reports the quality of the solutions. *Percentage of uncovered flights* indicates the percentage of scheduled flights that are not covered despite the large penalty that we impose. Average Credited Flying Time (ACFT) shows how many hours each pilot works on average. The credited time consists of the duration of the air legs, half of the duration of deadheads, a debriefing time for each post pairing, and a briefing and debriefing time for each post-pairing rest. *Preferred vacations* shows the percentages of pilots with satisfied vacation requests. *Percentage of preferred flights* shows the percentages of the preferred flights in schedules. This is a measure of the quality of the personalization.

Although the number of subproblems is high (between 33 and 305) the solutions are found in a reasonable time (0.16 min for the smallest instance and 184.76 min for the largest). The CPU times are lower than those for bidline assignment. This reduction in time is due to the faster CPU and/or the improved version of CPLEX we use. On average, 29.09 % of the flights are preferred flights, and 13.76 % of the pilots have preferred vacations in their schedules. On average, 0.41 % of the air legs are not covered; this is an acceptable figure. Reserve crew members can be assigned to the uncovered flights. Except for $I5 - 757$, the gap is smaller than 1 %. For $I5 - 757$, the number of CG iterations and the CPU time indicates that this is a difficult instance. The gap of 2.91 % confirms that it is difficult to find integer solutions for this instance; this is because the branching strategies do not fully explore the branch-and-bound tree. In industry, when users are not satisfied with the quality of a solution (e.g., large gaps), they modify some of the parameters and resolve the problem; we did not do this.

9 Future directions for airline crew scheduling

Airline crew scheduling remains an active and interesting area with several unexplored avenues for research. Pairing construction for a monthly planning horizon, rather than daily or weekly problems, is still a challenge. In addition, some difficult constraints are currently treated heuristically, and research could take into account more sophisticated industrial regulations such as the 8-in-24 rule. It would also be interesting to introduce more sophisticated types of preferences. Research

into the integration of the crew pairing and crew assignment problems is ongoing. Robust scheduling in the presence of uncertainty and possible perturbations is another challenging problem. The scheduling of cabin crew has received less attention. Each flight requires many cabin crew members, and some may need specific qualifications (administrative level, language, safety). As a result, each cabin crew member must be considered individually. Another potential research area is the simultaneous construction of schedules for cabin and cockpit crews that include common duties and pairings for different types of crew members (pilots, copilots, flight attendants).

We have decided to make available our data sets, preference generators, and the two types of supplementary constraints. These data sets will permit the research community to compare different algorithms. Adding supplementary constraints is one way to improve the link between the crew pairing and the crew assignment problem. These constraints lead to better pairings and more productive crew schedules. Our collaborators at AD OPT have suggested two types of supplementary constraints, and we have developed two generators to produce these constraints. The first group of constraints restricts the number of credited flying hours per base during the pairing construction. Using the unconstrained pairings, we compile statistics on the total credited hours. Two parameters are considered. The first adds an allowance to the total credited hours; the second controls the percentage of credited hours per base. We associate a maximum percentage of the total credited flying hours with the main base and evenly divide the remaining time between the other bases. The second group of constraints is stronger than the first group and controls the crew availability per base per day. Given a planning horizon and an unconstrained solution, we count the number of duties per day. A parameter adds an allowance to the total number of duties for each day of the horizon. Another parameter controls the number of crew members per base per day. If these values are not integer, we round them up while preserving the total number of crew members. All the data sets and generators are available at www.gerad.ca/en/papers/G-2014-22.

10 Conclusion

We have provided an extensive review of the airline crew scheduling problem. We have also proposed a mathematical formulation for the personalized crew assignment problem in the context of a sequential approach (crew pairing followed by crew assignment). We constructed personalized schedules by associating a subproblem with each crew member. In our tests, the number of subproblems varies between 33 and 305. In the bidline problem for the same set of instances, a subproblem is associated with each crew base (giving a total of 3 subproblems), so the personalized assignment problem is more challenging.

Taking the crew preferences into account is common in European airlines and becoming more frequent in North American airlines. We considered two types of preferences, for flights and vacations, in the context of short- and medium-haul flights. The results show that an acceptable level of crew satisfaction can be

achieved when monthly schedules are constructed for pilots via a sequential approach based on branch and price.

References

Achour H, Gamache M, Soumis F, Desaulniers G (2007) An exact solution approach for the preferential bidding system problem in the airline industry. Transp Sci 41(3):354–365

AhmadBeygi S, Cohn AM, Weir M (2009) An integer programming approach to generating airline crew pairings. Comput Oper Res 36(4):1284–1298

Anbil R, Gelman E, Patty B, Tanga R (1991) Recent advances in crew-pairing optimization at American Airlines. Interfaces 21(1):62–74

Anbil R, Tanga R, Johnson EL (1992) A global approach to crew-pairing optimization. IBM Syst J 31(1):71–78

Andersson E, Housos E, Kohl N, Wedelin D (1998) Operations Research in the airline industry. Springer, US, pp 228–258

Azadeh A, Farahani M, Eivazy H, Nazari-Shirkouhi S, Asadipour G (2013) A hybrid meta-heuristic algorithm for optimization of crew scheduling. Appl Soft Comput 13(1):158–164

Barnhart C, Shenoi R (1998) An approximate model and solution approach for the long-haul crew pairing problem. Transp Sci 32(3):221–231

Barnhart C, Johnson EL, Anbil R, Hatay L (1994) A column-generation technique for the long-haul crew-assignment problem. In: Optimization in Industry 2, John Wiley & Sons Inc, pp 7–24

Barnhart C, Johnson EL, Nemhauser GL, Savelsbergh MWP, Vance PH (1996) Branch-and-price: Column generation for solving huge integer programs. Oper Res 46:316–329

Barnhart C, Cohn AM, Johnson E, Klabjan D, Nemhauser G, Vance P (2003) Airline crew scheduling. In: Hall R (ed) Handbook of Transportation Science, International Series in Operations Research & Management Science. Springer, US, pp 517–560

Beasley J, Cao B (1996) A tree search algorithm for the crew scheduling problem. Eur J Oper Res 94(3):517–526

Belobaba P, Odoni A, Barnhart C (2009) The global airline industry. Wiley, Chichester

Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. Numer Math 4(1):238–252

Bixby RE, Gregory JW, Lustig IJ, Marsten RE, Shanno DF (1992) Very large-scale linear programming: A case study in combining interior point and simplex methods. Oper Res 40(5):885–897

Boubaker K, Desaulniers G, Elhallaoui I (2010) Bidline scheduling with equity by heuristic dynamic constraint aggregation. Transp Res Part B Methodol 44(1):50–61

Campbell KW, Durfee RB, Hines GS (1997) FedEx generates bid lines using simulated annealing. Interfaces 27(2):1–16

Chen C, Liu T, Chou J, Wang C (2012) Multiobjective airline scheduling: An integrated approach. In: Proceedings of, IEEE, SICE Annual Conference (SICE), 2012, pp 1266–1270

Christou IT, Zakarian A, Liu JM, Carter H (1999) A two-phase genetic algorithm for large-scale bidline-generation problems at Delta Air Lines. Interfaces 29(5):51–65

Cohn AM, Barnhart C (2003) Improving crew scheduling by incorporating key maintenance routing decisions. Oper Res 51(3):387–396

Cordeau J, Stojković G, Soumis F, Desrosiers J (2001) Benders decomposition for simultaneous aircraft routing and crew scheduling. Transp Sci 35(4):375–388

Day PR, Ryan DM (1997) Flight attendant rostering for short-haul airline operations. Oper Res 45(5):649–661

Deng G, Lin W (2011) Ant colony optimization-based algorithm for airline crew scheduling problem. Expert Syst Appl 38(5):5787–5793

Desaulniers G, Desrosiers J, Dumas Y, Marc S, Rioux B, Solomon M, Soumis F (1997) Crew pairing at Air France. Eur J Oper Res 97(2):245–259

Desaulniers G, Desrosiers J, Ioachim I, Solomon MM, Soumis F, Villeneuve D (1998) A unified framework for deterministic time constrained vehicle routing and crew scheduling problems. In: Crainic TG, Laporte G (eds) Fleet Management and Logistics, Springer, New York, pp 57–93

Desrosiers J, Lübbecke ME (2005) A primer in column generation. Springer, US

Desrosiers J, Dumas Y, Desrochers M, Soumis F, Sanso B, Turdeau P (1991) A breakthrough in airline crew scheduling. Technical Report G-91-11, Cahiers du GERAD

Desrosiers J, Dumas Y, Solomon MM, Soumis F (1995) Time constrained routing and scheduling. In: Ball MO, Magnanti TL, Monma CL, Nemhauser GL (eds) Network Routing, Handbooks in Operations Research and Management Science, vol 8. Elsevier, pp 35–139

Dück V, Wesselmann F, Suhl L (2011) Implementing a branch and price and cut method for the airline crew pairing optimization problem. Public Transp 3(1):43–64

El Moudani W, Cosenza CAN, de Coligny M, Mora-Camino F (2001) A bi-criterion approach for the airlines crew rostering problem. In: Evolutionary Multi-Criterion Optimization. Springer, pp 486–500

Elhallaoui I, Villeneuve D, Soumis F, Desaulniers G (2005) Dynamic aggregation of set-partitioning constraints in column generation. Oper Res 53(4):632–645

Fahle T, Junker U, Karisch SE, Kohl N, Sellmann M, Vaaben B (2002) Constraint programming based column generation for crew assignment. J Heuristics 8(1):59–81

Fisher ML (1981) The Lagrangian relaxation method for solving integer programming problems. Manag Sci 27(1):1–18

Fisher ML (1985) An applications oriented guide to Lagrangian relaxation. Interfaces 15(2):10–21

Gamache M, Soumis F, Villeneuve D, Desrosiers J, Gélinas E (1998) The preferential bidding system at Air Canada. Transp Sci 32(3):246–255

Gamache M, Soumis F, Marquis G, Desrosiers J (1999) A column generation approach for large-scale aircrew rostering problems. Oper Res 47(2):247–263

Gao C, Johnson E, Smith B (2009) Integrated airline fleet and crew robust planning. Transp Sci 43(1):2–16

Geoffrion AM (1974) Lagrangean relaxation for integer programming. Springer

Gershkoff I (1989) Optimizing flight crew schedules. Interfaces 19(4):29–43

Gopalakrishnan B, Johnson EL (2005) Airline crew scheduling: State-of-the-art. Ann Oper Res 140(1):305–337

Guo Y, Mellouli T, Suhl L, Thiel M (2006) A partially integrated airline crew scheduling approach with time-dependent crew capacities and multiple home bases. Eur J Oper Res 171(3):1169–1181

Hjorring CA, Hansen J (1999) Column generation with a rule modelling language for airline crew pairing. In: Proceedings of the 34th Annual Conference of the Operational Research Society of New Zealand, pp 133–142

Hoffman K, Padberg M (1993) Solving airline crew scheduling problems by branch-and-cut. Manag Sci 39(6):657–682

Hu J, Johnson E (1999) Computational results with a primal-dual subproblem simplex method. Oper Res Lett 25(4):149–157

Irnich S, Desaulniers G (2005) Shortest Path Problems with Resource Constraints. Springer, US

Jarrah AIZ, Diamond JT (1997) The problem of generating crew bidlines. Interfaces 27(4):49–64

Klabjan D (2005) Large-scale models in the airline industry. In: Desaulniers G, Desrosiers J, Solomon M (eds) Column Generation. Springer, US, pp 163–195

Klabjan D, Johnson E, Nemhauser G, Gelman E, Ramaswamy S (2001) Solving large airline crew scheduling problems: Random pairing generation and strong branching. Comput Optim Appl 20(1):73–91

Klabjan D, Johnson EL, Nemhauser G, Gelman E, Ramaswamy S (2002) Airline crew scheduling with time windows and plane-count constraints. Transp Sci 36(3):334–337

Kohl N, Karisch S (2000) Integrating operations research and constraint programming techniques in crew scheduling. In: Proceedings of the 40th Annual AGIFORS Symposium, pp 20–25

Kohl N, Karisch S (2004) Airline crew rostering: Problem types, modeling, and optimization. Ann Oper Res 127(1):223–257

König J, Strauss C (2000a) Rostering-integrated services and crew efficiency. J IT Tour 3(1):27–39

König J, Strauss C (2000b) Supplements in airline cabin service. In: Information and Communication Technologies in Tourism 2000. Springer, pp 365–374

Lavoie S, Minoux M, Odier E (1988) A new approach for crew pairing problems by column generation with an application to air transportation. Eur J Oper Res 35(1):45–58

Maenhout B, Vanhoucke M (2010) A hybrid scatter search heuristic for personalized crew rostering in the airline industry. Eur J Oper Res 206(1):155–167

Marsten R, Shepardson F (1981) Exact solution of crew scheduling problems using the set partitioning model: Recent successful applications. Networks 11(2):165–177

Martin RK (1999) Large Scale Linear and Integer Optimization: A Unified Approach. Springer

Mercier A, Soumis F (2007) An integrated aircraft routing, crew scheduling and flight retiming model. Comput Oper Res 34(8):2251–2265

Mercier A, Cordeau J, Soumis F (2005) A computational study of Benders decomposition for the integrated aircraft routing and crew scheduling problem. Comput Oper Res 32(6):1451–1476

Minoux M, Vajda V (1986) Mathematical programming: theory and algorithms. Wiley, New York

Papadakos N (2009) Integrated airline scheduling. Comput Oper Res 36(1):176–195

Ruther S (2010) A multi-commodity flow formulation for the integrated aircraft routing, crew pairing, and tail assignment problem. In: Proceedings of the 45th Annual Conference of the Operational Research Society of New Zealand

Saddoune M, Desaulniers G, Elhallaoui I, Soumis F (2011) Integrated airline crew scheduling: a bi-dynamic constraint aggregation method using neighborhoods. Eur J Oper Res 212(3):445–454

Saddoune M, Desaulniers G, Elhallaoui I, Soumis F (2012) Integrated airline crew pairing and crew assignment by dynamic constraint aggregation. Transp Sci 46(1):39–55

Saddoune M, Desaulniers G, Soumis F (2013) Aircrew pairings with possible repetitions of the same flight number. Comput Oper Res 40(3):805–814

Sandhu R, Klabjan D (2007) Integrated airline fleeting and crew-pairing decisions. Oper Res 55(3):439–456

Sellmann M, Zervoudakis K, Stamatopoulos P, Fahle T (2002) Crew assignment via constraint programming: Integrating column generation and heuristic tree search. Ann Oper Res 115(1–4):207–225

Souai N, Teghem J (2009) Genetic algorithm based approach for the integrated airline crew-pairing and rostering problem. Eur J Oper Res 199(3):674–683

Subramanian S, Sherali H (2008) An effective deflected subgradient optimization scheme for implementing column generation for large-scale airline crew scheduling problems. INFORMS J Comput 20(4):565–578

Vance PH, Barnhart C, Johnson EL, Nemhauser GL (1997) Airline crew scheduling: a new formulation and decomposition algorithm. Oper Res 45(2):188–200

Wedelin D (1995) An algorithm for large scale 0–1 integer programming with application to airline crew scheduling. Ann Oper Res 57(1):283–301

Weir JD, Johnson EL (2004) A three-phase approach to solving the bidline problem. Ann Oper Res 127(1):283–308

Zeghal F, Minoux M (2006) Modeling and solving a crew assignment problem in air transportation. Eur J Oper Res 175(1):187–209