Generalized Algorithms for Crew Planning: Survey and Future Directions for Railways

Viswanathan Prem Kumar
Michel Bierlaire
Olivier Gallay

STRC 2010  September 2010
Generalized Algorithms for Crew Planning: Survey and Future Directions for Railways

Viswanathan Prem Kumar  
EPFL, TRANSP-OR  
Station 18, GC B3 435  
CH-1015 Lausanne  
phone: +41216932432  
fax: +41216938060  
prem.viswanathan@epfl.ch

Michel Bierlaire  
EPFL, TRANSP-OR  
Station 18, GC B3 454  
CH-1015 Lausanne  
phone: +41216932537  
fax: +41216938060  
michel.bierlaire@epfl.ch

Olivier Gallay  
EPFL, TRANSP-OR  
Station 18, GC B3 445  
CH-1015 Lausanne  
phone: +41216932532  
fax: +41216938060  
olivier.gallay@epfl.ch

September 2010

Abstract

This paper studies the crew planning problem as observed in the transportation industry. We first survey the existing literature on crew scheduling applications in railways and airlines. Next, we identify the synergies in the two domains and propose new directions for railway crew scheduling inspired from the applications in airlines.

Keywords

Crew Scheduling, Survey, Airlines, Railways.
1 Introduction

Staff scheduling, or rostering, is the process of building working timetables for manpower resources so that the provider can serve their customers effectively. The first part of this process involves determining the number of resources, with specific skills and experience, needed to meet the service demand. Individual staff members are allocated to shifts so as to meet the required staffing levels at different times, and duties are then assigned to individuals for each shift. All regulatory and labor regulations associated with the relevant workplace agreements must be observed during the process.

Staff scheduling is known as crew scheduling and rostering in the transportation services - airlines, railways, mass transit, water-ways and surface transport. The common features for all these applications are i) both temporal and spatial features are involved, because, each task is characterized by its starting time and location, as well as its finishing time and location, and ii) all tasks to be performed by crew members are determined from an existing timetable.

In transportation systems, total crew costs, which include the salaries, benefits and expenses, is second only to the fuel costs of all operational cost components. But unlike fuel costs, a large portion of crew costs are controllable. Achieving efficiency in crew hiring and crew utilization by optimal rostering could substantially improve the profits of the service provider. Owing to the span of operations, it is likely that crew members operate out of various stations, referred to as crew bases. A pairing refers to a sequence of trips starting and ending at the same crew base. A pairing that satisfies all legal, contractual and regulatory agreements is a feasible pairing. A sequence of pairings covering the entire planning horizon is called as a bid line. Each bid line is assigned to a crew member based on her preferences.

A typical crew schedule planning problem for a transportation system involves a set of trips to be covered every day in a given period by a set of crews, building a daily assignment of each trip to a crew so as to guarantee that all the trips are covered in the period and the corresponding overall cost is minimized. Due to its complexity the Crew Scheduling Problem (CSP) is typically divided into two sequential sub-problems: first, the Crew Pairing Problem (CPP) where a set of pairings is generated that minimizes operational cost in such a way that each flight belongs to exactly one pairing and so that all pairings can be operated by the minimal set of anonymous crews. Second, the Crew Assignment Problem (CAP) or Crew Rostering Problem (CRP) would assign these duties over repeating horizon period to individual crew members taking into account their preferences or at least ensuring equitable distribution of perceived prejudices.

A planned timetable operated by the service provider every day for the entire planning
horizon, a set of base and change-over stations and a set of constraints resulting from union contracts or from company regulations is given. It is extremely difficult to find good solutions to these highly constrained and complex problems and even more difficult to determine optimal solutions that minimize costs, meet crew preferences, distribute shifts equitably among crew members and satisfy all the workplace constraints. Thus, the crew scheduling and rostering problems are solved for feasibility rather than optimality as the main objective. In many situations, the analysts and supervisors involved in developing rosters need decision support tools to help provide the right crew members at the right time and at the right cost while achieving a high level of employee job satisfaction and business service continuity.

2 Literature Review

2.1 Non-Integrated Airline Crew Scheduling

Crew scheduling for commercial airlines is one of the oldest practical problems in civilian applications that has been studied. The seminal work by Arabeyre et al. (1969) considered the problem scale at several major airlines of that time and described their solution method which varied from use of heuristics to relaxations of the Mixed Integer Program (MIP) model. Marsten and Shepardson (1981) propose the set-partitioning based MIP model and a solution technique based on Lagrangian relaxation and sub-gradient optimization. Their experience on several large airlines is reported and they also make a case for heuristic decomposition of the large constraint matrix into smaller matrices to be solved separately.

Desrochers and Soumis (1989) marked a watershed in research on large-scale optimization problems by proposing a column generation approach to solve a metropolitan mass transit crew scheduling problem. The column generation approach decomposed the problem into two parts. The set covering problem chose a schedule from known feasible workdays. The second subproblem was a shortest path problem with resource constraints and was used to propose new feasible workdays to improve the current solution of the set covering problem. Anbil et al. (1991) report the implementation of a crew scheduling software called TRIP at American Airlines and sold to 10 other airlines and to a railroad company. TRIP is based on an approach where pairings are iteratively improved by generating and solving a series of sub-problems. This paper also provides a guideline to estimate crew costs.

Hoffman and Padberg (1993) propose a branch-and-cut solution for the crew pairing problem and report their findings on the real data for two US and two European airlines. Wedelin (1995) reports the crew scheduling algorithm that forms the backbone of the generic
solution offered by Carmen Systems. This paper suggests the use of dynamic programming to select the columns in MIP to improve the pairing generation. Beasley and Cao (1996) models the problem as a Lagrangian relaxation that is improved by subgradient optimization and solves it by a tree search procedure to optimality. Obviously the size of problem instances solved to optimality is fairly small.

Vance et al. (1997) present a new model for airline crew scheduling based on breaking the decision process into two stages. In the first stage a set of duty periods that cover the flights in the schedule is selected. Then, in the second stage, pairings using those duty periods are built. The paper compares, with examples, the pros and cons of generating pairing from flights vis-à-vis from duties. It also suggests a decomposition approach based on reduced costs for solving the model by branch-and-price and present computational results for a major US airline.

Klabjan et al. (2001) develop a new approach for solving the crew scheduling problem that is based on enumerating hundreds of millions random pairings for a large US airline. The linear programming relaxation is solved first and then millions of columns with best reduced cost are selected for the integer program. The number of columns is further reduced by a linear programming based heuristic. Finally an integer solution is obtained with a commercial integer programming solver. The branching rule of the solver is enhanced with a combination of solver’s strong branching option and a specialized branching rule.

Kohl and Karisch (2004) consider the crew rostering problem and describe the various types of logical rules proposed by the crew members. This is done by generating the duties by pre-processing to take care of crew and business requirements such that there is at least one duty corresponding to each crew member. Ideally all duties must be selected to obtain optimal solution, but in practice only a large number of "good" ones are fed into the set-covering model. Side constraints representing inter-crew business realities such as at least one experienced crew are added to the formulation and solved using column generation algorithm.

2.2 Non-Integrated Railway Crew Scheduling

It may be worth noting that while airline crew scheduling was pioneering techniques for solving crew scheduling with set covering formulation and column generation techniques, railway crew scheduling lagged far behind. Contemporary works such as Morgado and Martins (1992) primarily concentrated by automating manual thumb-rules and contextual heuristics.

This lag in research advances may be because of certain differences in the problem definition and terminology used for railway crew scheduling. An airline flight leg normally
consists of a non-stop, direct flight leg between an origin and a destination where the crew can either continue on the same aircraft or be flown on a different one. Drivers and conductors in a train can change over at not only origin and destination, but at any station en route. This could result in a potentially large number of trips (by definition) which can be combined to form duties. Railway crew scheduling for small networks normally requires that the crew returns to their base station at the completion of the shift while it is not the case in even domestic airline crew scheduling. There could be other minor differences in the union contracts and perceptions in the trips undertaken by a driver or conductor.

One of the seminal works in railway crew scheduling for the Italian Railway Company was reported by Caprara et al. (1997). They follow the approach used for airline crew scheduling (referred to as crew management) and decomposes the problem into two sub-problems - crew pairing (referred to as crew scheduling) and crew rostering. Solution approach followed in this paper has similarities with Vance et al. (1997) and generates pairings by using a set covering model on the maximal set of duties and uses Lagrangian and greedy heuristics to select the master problem columns.

The Dutch Railway Company has invested on the application of operations research techniques for their operations. Kroon and Fischetti (2000) describe the generation of efficient and acceptable duties with a high robustness against the transfer of delays of trains. The applied set covering model is solved by dynamic column generation techniques, Lagrangian relaxation and powerful heuristics. Freling et al. (2001) describe a heuristic branch-and-price algorithm for the assignment of train guards (also referred to as conductors) with emphasis on schedule robustness and non-uniform requirement across the day. For the dynamic programming algorithm, dominance criteria are used to reduce the state space. Abbink et al. (2004) suggested further improvements to this model by including substantial, acceptable variation in duty as desired by the drivers and, in particular, conductors. This project was appropriately termed as "sharing sweet and sour". Abbink et al. (2007) describe several methods to partition large instances into several smaller ones. These smaller instances are then solved iteratively with the existing crew scheduling software. They also compare several partitioning methods with each other.

Vaidyanathan et al. (2007) propose a network flow model that maps the assignment of crews to railroad company in USA as the flow of crews on an underlying network, where different crew types are modeled as different commodities in this network. They formulate the problem as an integer programming problem on this network and also develop several highly efficient algorithms using problem decomposition and relaxation techniques, in which the special structure of the underlying network model to obtain significant increases in speed is used.
Wang et al. (2008) propose an approach that starts with a decomposition phase to assign all trips to depots and gets a set of subproblems for the Chinese Railway Company. A column generation algorithm is embedded in a genetic algorithm to get the lower bounds of the subproblems. In the second phase, the linear programming relaxation of the models using a column generation scheme is solved using a branch-and-bound technique to get integer solutions.

### 2.3 Integrated Crew Scheduling

A recent trend, particularly in airline crew scheduling, has been to integrate different modules of the scheduling process to extract further efficiencies by optimization. A pictorial view of the various processes involved in airline industry is shown below in Fig. 1.

![Airline Scheduling Process](image)

**Figure 1: Airline Scheduling Process**

One of the first efforts towards an integrated solution for scheduling process is reported by Ernst et al. (2001) for the Australian National Rail Company. Their proposed approach involves determining the maximal clique of crew for the given network of services. Medard and Sawhney (2004) consider the operational problem of crew rescheduling (also referred to as recovery) in the event of a disruption and solve it by integrating both models to solve time critical crew recovery problems. This paper also describes how pairing construction and pairing assignment are done in a single step, while providing solution techniques based on simple tree search and more sophisticated column generation and shortest-path algorithms. Guo et al. (2006) suggest a partially integrated approach based on two tightly coupled components: the first constructs chains of crew pairings spaced by weekly rests, where crew capacities at different domiciles and time-dependent availabilities are considered. The second component rearranges parts of these pairing chains into individual crew schedules. Huisman

Mercier et al. (2005) describe the first attempt to integrate aircraft routing and crew pairing problem by determining a minimum-cost set of aircraft routes and crew pairings such that each flight leg is covered by one aircraft and one crew, and side constraints are satisfied. They propose an enhanced model incorporating robustness to handle linking constraints in the two problems and compare two Benders decomposition methods - one with the aircraft routing problem as the master problem and one with the crew pairing problem. They also study the impact of generating Pareto-optimal cuts on the speed of convergence of these methods.

Mercier and Soumis (2007) integrate aircraft routing, crew scheduling and flight retiming problem, by constructing a minimum-cost set of aircraft routes and crew pairings while choosing a departure time for each flight leg within a given time window. Linking constraints ensure that the same schedule is chosen for both the aircraft routes and the crew pairings, and impose minimum connection times for crews that depend on aircraft connections and departure times. The paper proposes a compact formulation of the problem and a Benders decomposition method with a dynamic constraint generation procedure to solve it. The results have been reported for two major airline companies.

Freling et al. (2003) deal with models, relaxations, and algorithms for an integrated approach to vehicle and crew scheduling for an urban mass transit system. They propose new mathematical formulations for integrated problem and suggest Lagrangian relaxations and Lagrangian heuristics. Column generation applied to set partitioning type of models is used to solve the Lagrangian model. Sandhu and Klabjan (2007) propose a model that completely integrates the fleeting and crew-pairing stages and guarantees feasibility of plane-count feasible aircraft routings, but neglects aircraft maintenance constraints. They design two solution methodologies to solve the model. One is based on a combination of Lagrangian relaxation and column generation, while the other one is a Benders decomposition approach.

### 2.4 Stochastic Crew Scheduling

We now touch upon some literature that deals with stochastic crew scheduling and robustness of the generated schedule. Yen and Birge (2006) consider a stochastic crew scheduling model and devise a solution methodology for integrating disruptions in the evaluation of crew schedules. They use information to find robust solutions that could better withstand disruptions.
They describe a stochastic integer programming model for the airline crew scheduling problem and develop a branching algorithm to identify expensive flight connections and find alternative solutions. The branching algorithm uses the structure of the problem to branch simultaneously on multiple variables without invalidating the optimality of the algorithm.

Schaefer et al. (2005) give computational results from three fleets that indicate that the crew schedules obtained from their methodology perform better in operations than the crew schedules found via state-of-the-art methods. They provide a lower bound on the cost of an optimal crew schedule in operations, and demonstrate that some of the crew schedules found using their methodology perform very well relative to this lower bound. Shebalov and Klabjan (2006) introduce the objective of maximizing the number of move-up crews, i.e. the crews that can potentially be swapped in operations. To solve the resulting large-scale integer program, a combination of delayed column generation and Lagrangian relaxation is used. The restricted master problem is solved by means of Lagrangian relaxation and the “duals” of the restricted master problem, which are used in delayed column generation, correspond to the Lagrangian multipliers.

3 Comparison of Crew Scheduling in Railways with Airlines and Public Transport

Substantial academic and industrial research on the crew pairing problem has focused on the problem faced by airlines. In the airline problem, the set of tasks (the smallest atomic unit that would allow for a crew changeover either at the start or at the end) is small - hundreds or a few thousands of legs. On the contrary the long haul traffic in trains, consists of more than 1’000 trains per day and many trains correspond to more than 10 legs since it is possible to change crew at several major stations in between. The problem of conductor assignment on a train has some clear similarities with cabin crew assignment in airlines because preparation times and qualification requirements depend on the position and each position must be considered individually. The problem corresponding to driver and conductor assignment for a local public transport company is much smaller in comparison to the airline or railway problem, but robust scheduling may have more relevance in that case because of traffic delays.

In railways, the set of tasks for a weekly pairing problem is generally more than 100’000. Further, the number of non-zero coefficients per variable is much higher than in the airline case. An airline trip will typically contain 2-4 legs per day and rarely have duration of more than 4 days. A railway trip is seldom more than two days long, but since the average duration of a leg is only 30 minutes for the long haul traffic and 20 minutes for the regional traffic,
pairings with 20 or more legs are not uncommon. Trips in local public transport correspond to
duties and pairing because of the smaller problem size.

It is known that the average case difficulty of a set partitioning problem increase with
the average number of non-zero coefficients per variable, so the railway set partitioning
problems are not only more difficult than their airline counterparts due to their size, but
also due to their inherently complex structure. Calculation of legality and cost of a trip is
substantially more complex for railways. It is because the train crew carries out a number
of duties in addition to driving trains. In particular train drivers have a lot of preparatory
tasks, which must be derived from the locomotive and wagon rotations. Rest calculations for
conductors are also complicated by the possibility of taking a rest while the train is running.
At certain railway companies, rest time at stations between two connections does not get paid
beyond the minimal connection time. Since waiting period is natural for changing trains, a trip
with more work can actually produce less pay than a trip with less work because the waiting
period gets construed as rest and will be unpaid.

Wagon and locomotive rotations are needed to identify preparatory and terminal activi-
ties as well as calculate connection times. Consistency with the time table must be maintained.
In some cases infrastructure data is used to determine the route a train takes between two
stations, since each route will require a specific skill and experience profile.

4 Further Directions for Railway Crew Scheduling

The topic of crew scheduling is clearly receiving increased attention as measured by the
number of contributions in the last few years. The nature and scope of the research conducted
is also gaining in diversity as nearly every domain of application - surface transport, airlines
and railways have been the object of some recent research.

There also appears to be a constant refinement and diversification of the modeling and
solution methods proposed and used for crew scheduling. Early models were usually built
to have a structure that made them solvable by thumb rules and crude approximation of the
reality. Later years witnessed a gradual introduction of integer programming formulation as a
set covering model. Most of the existing solution methodology focuses on solving the problem
by dynamic column generation technique; few others solved using meta-heuristics that have
proven to be very effective for several classes of discrete optimization problems. Of course,
this progression is also made possible by the increased power of computers and information
systems.
As already stated above, railway crew scheduling represents a computationally complex and challenging problem because of both the size of the instances to be solved and the type and number of operational constraints. That said, implementation of a clever algorithm to generate duties from individual trips is the key to obtaining a reasonable scheduling solution to the problem. On the whole, we feel that crew scheduling research over the next few years would concentrate on the following themes.

It is demonstrated that column generation based algorithms are effective to solve railway CSP. However, the efficiency of these algorithms is affected by the generation of initial solutions, the number of columns added into the RMP every time, its branching rule and the search strategy of B&B tree. Also, detection of the first feasible solution gets more difficult as the problem size gets larger.

Most of the crew scheduling applications for railways are specific to the network and the country for which they are being built. We feel that it is necessary to generate a scale-free crew scheduling algorithm irrespective of the size of the railway network and exploit the constraints to reduce the feasible solution sets.

There is little work on crew rostering algorithms, particularly where the crew preferences are considered. Since the duty cycle of a railway crew is much shorter than an airline crew, crew preferences are likely to be different and more specific about vacations, time offs and hours of work. Obviously such preferences have a cost, but research can actually attribute a quantitative merit by evaluating business performance vis-à-vis employee satisfaction.

Another important area requiring further work is generalization of models and methods. Currently, models and algorithms often require significant modification when they are to be transferred to a different application area, or to accommodate changes within an organization. In a continually changing environment it is not desirable to have organization’s internal structures, processes and work policies hard-wired into models, algorithms and software for personnel scheduling. New models need to be formulated that provide more flexibility to accommodate individual workplace practices. This can then lead to the development of more general algorithms that will be more robust to changes in the rostering requirements.

The crew planning system is a system of real-world optimization. The challenge of real world optimization is to find the best possible solution, to the right problem, as fast as possible [Kohl and Karisch (2004)]. All these three dimensions are essential to the success of a crew planning system. The academic research almost solely focuses on the first dimension in the case of difficult problems, e.g. large scale vehicle routing problems, and on the third dimension in the case of "easy" problems, e.g. linear programming.
The second dimension of the right problem is too often omitted even in scientific works supposed to be related to real-life problems. All commercial systems of today solve a problem which is not exactly the real-life problem but an approximation or a sub-problem resembling the real-life problem only to a certain extent. Even though the sub-problem may be solved optimally, the underlying real-life problem is only partially solved. In fact the crew rostering problem itself is not the right problem. The real problem is to assign individual tasks to individuals. The pairings which we have assumed to be input to the crew rostering problem are just the result of another suboptimal planning problem - the crew pairing problem. Even the crew rostering problem described in this paper is to some extent a simplification of the real problem faced after pairings have been produced. For example, complex training programs where pilots have to go through a series of phases each with different requirements on the assigned instructor, duration and activities to be carried out are not really contained within the framework outlined here.

Future research paths in railway crew scheduling and planning are oriented toward models that address the integration of various policies. Because rail activities are generally complex and involve large-scale systems, the traditional approach in the industry has been to separate planning activities into several components. This natural tendency yields more manageable sub-systems but also presents several limitations. In particular, there is a strong incentive to simultaneously treat routing, loco assignment or rolling stock assignment with crew scheduling problems because of the important interactions linking these multiple categories of decisions. Hence, models that integrate several aspects and levels of planning should be increasingly common in upcoming years.

Real-time control is at the other side of the planning spectrum. The current trend in the railway industry is a shift from "planning in detail" to "effective real-time control". Disturbances and disruptions in the railway operations are inevitable. Therefore, large parts of the operational plans are never carried out. In case of a disruption, one needs as soon as possible an alternative plan. To some extent, several potential alternative plans can be prepared already, e.g. in the form of disruption scenarios for adapting the timetable and the rolling stock circulation. The latter may be particularly affective in the case of a cyclic timetable. However, crew schedules are usually non-cyclic. Therefore, being able to quickly generate alternative crew schedules is highly important in case of a disruption of the railway system. In order to make this effective in practice, one needs i) to have detailed information on the status quo of the railway system (e.g. the positions of trains and crews), ii) to be able to quickly generate alternative crew schedules, and iii) to disseminate the alternative plans in a dependable way among all stake-holders. Although, from a mathematical point of view, these problems may seem to be similar to the corresponding operational planning problems, they are quite different,
mainly due to the dynamic character of real-time control and the high time pressure.

In the future, crew pairing and crew rostering will be integrated into one planning problem. Only a combined problem represents the correct formulation of the crew planning problem and allows the best rosters both from a cost and a quality point of view. We view the creation of rosters out of legs, i.e. skipping the intermediate step of producing a pairing solution as one of the important research areas in crew planning. Another area which currently attracts great interest is crew recovery where individual rosters are maintained and repaired after their publication until the day of operations.

Currently, crew scheduling problems in many situations are still solved manually in practice, partly due to the traditionally rather conservative character of the railway industry. The innovative possibilities provided by the effective application of mathematical models and optimization techniques in the airline industry has acted as a trigger for the railway industry, and software applications based on these techniques recently started to be implemented. Researchers in mathematical optimization should grasp the currently available momentum and opportunities in the railway industry by not focusing too much on theoretical results, but by going for real-world applications of their models and techniques. The latter will lead to a win-win situation, both for the researchers and for the railway industry.

References


