CASPT 2015

Optimizing Crew Schedules with Fairness Preferences

Silke Jütte · Daniel Müller · Ulrich W. Thonemann

Abstract Railway crew scheduling deals with generating duties for train drivers to cover all train movements of a given timetable. The objective is to minimize the overall costs associated with a crew schedule, which includes workforce costs, hotel costs, etc. A cost minimal schedule often contains duties that are unpopular to train drivers, and these unpopular duties are often unevenly distributed among crew depots. At the company that motivated our research, train drivers dislike hotel rests, for example. Currently, some crew depots operate large numbers of duties with hotel rests, while others do not have any duties with hotel rests at all. This situation is perceived as unfair. Train drivers prefer schedules with fewer and more evenly distributed unpopular duties across crew depots. In this paper, we define and measure unpopularity and (un)fairness in a railway crew scheduling context. We integrate fairness conditions into a column generation based solution algorithm and analyze the effects of increased fairness on the unpopularity of a schedule. Our method has been applied to test instances at a large European railway freight carrier. We were able to significantly improve schedule fairness, while schedule costs were only marginally affected. In most cases, increased fairness also induced a reduction in the number of unpopular duties in the schedule.

Keywords Human resource planning · Scheduling · Fairness · Large scale optimization

S. Jütte · U.W. Thonemann
Department of Supply Chain Management and Management Science, University of Cologne, Albertus-Magnus-Platz, D-50923 Cologne, Germany
Tel.: +49 221 470 1510
Fax: +49 221 470 7950
E-mail: silke.juette@uni-koeln.de, ulrich.thonemann@uni-koeln.de

D. Müller
E-mail: mueller.daniel@mail.com

1 Introduction

Railway crew scheduling deals with generating train driver duties for a given train schedule such that all work regulations are met and the resulting schedule costs are minimized. Train drivers are located at crew depots where duties start and end. The number of train drivers assigned to depots varies, and the crew schedule optimization must take the capacity limits at the depots into account.

Typically, some duties of a crew schedule have properties that are unpopular to train drivers. For example, train drivers dislike hotel rests within a duty and prefer returning to their home depot after a one-day duty. The aversion towards hotel rests is especially strong in the freight railway business, where intensified train operations at night times often require hotel rests at day times.

The second main source of unpopularity are early morning duties, i.e., duties that start in the early morning hours and require driving "into the day". Such duties are conflicting with the train driver's circadian rhythm. In the short term, this can result in increased fatigue during work and hence increased risk of mistakes and accidents as well as smaller productivity (Akerstedt, 2003; Folkard and Tucker, 2003). In the long term, working hours that conflict with the circadian rhythm can cause severe health issues (Harrington, 2001). The company that motivated this research considers early morning duties starting between midnight and 4 a.m. as most conflicting with the circadian rhythm.

In a cost-optimal schedule, unpopular duties are generally unequally distributed between the crew depots. At crew depots with a higher-than-average fraction of unpopular duties, the schedule is perceived as unfair. This decreases the job satisfaction of the train drivers and can result in lower job performance, bickering, and increased absenteeism (Bard and Purnomo, 2005). A high perception of unfairness has also been shown to increase turnover rates of employees (Smet et al, 2013) and to trigger labor strikes with severe economic impacts (Abbink et al, 2005).

In this paper, we analyze the integration of fairness considerations into a crew scheduling optimization approach that has been implemented at a major European railway freight carrier. In the following, we will refer to this company as *Freightrail*. We analyze the effects of fairness constraints on the total unpopularity of a crew schedule. To solve problems of reasonable size, we use a heuristic solution approach. The approach is based on column generation and generates solutions that are within one percent of optimality.

In Section 2, we describe the situation at Freightrail. In Section 3, we provide an overview of research on fairness that is relevant for crew scheduling. In Section 4, we describe the basic solution algorithm for solving the railway crew scheduling problem and show how we incorporated fairness in the model. In Section 5, we evaluate our approach using data from Freightrail and analyze the effect of fairness considerations on the unpopularity, fairness, and cost of the crew schedule. In Section 6, we summarize our findings and provide directions for future research.

2 Problem Description

In the freight railway business, crew scheduling is one of a series of planning steps necessary to operate trains (see Cordeau et al, 1998, for an overview of planning problems and corresponding optimization models within railway operations). Once trains have been assembled, timetables have been fixed, and engines have been assigned to the trains, the crew scheduling problem consists of generating duties for the train drivers such that all work regulations and capacities are met and the resulting schedule has minimal cost. Real-world problem instances often require generating duties for thousands of train drivers located at hundreds of depots and operating tens of thousands of trains. Due to the large problem dimension, the assignment of the generated duties to individual train drivers is performed in a separate, subsequent planning step ("crew rostering") which is out of the scope of this paper (see Caprara et al, 1998, for an example of crew rostering in the railway context).

In crew scheduling, a train denotes a transportation order between two locations. For planning purposes, a train is divided into trips, i.e., segments of scheduled train movements between two consecutive transfer points that must be serviced by the same train driver. The railway crew scheduling problem is commonly formulated as a set covering model with side constraints: Let T denote the set of trips. A duty d is a sequence of trips in T. d is feasible if it fulfills a set of legal, union, and company requirements, which regulate, e.g., minimum connection times, the length and position of breaks, maximum working times, etc. Let D denote the set of all feasible duties for the set T of trips, and let $a_{td} = 1$ if trip t is contained (covered) in d ($a_{td} = 0$ otherwise). For each feasible duty d, a parameter c_d reflects the costs associated with the duty (worktime costs, costs for using public transport or taxi connections, and hotel accommodation). Let $x_d = 1$ if duty d is part of the solution schedule and $x_d = 0$ otherwise. A duty d is assigned to a crew depot $j \in J$ from which the train driver starts and ends the duty. Let $b_{jd} = 1$ if d is assigned to depot j and $b_{jd} = 0$ otherwise. At crew depot j, k_j train drivers are available. The basic set covering formulation (SC) of the railway crew scheduling problem is

$$\text{minimize} \quad \sum_{d \in D} c_d x_d + \sum_{j \in J} \hat{c}_j \hat{y}_j \tag{1}$$

$$\text{subject to} \quad \sum_{d \in D} a_{td} x_d \geq 1 \quad \forall t \in T \tag{2}$$

$$\sum_{d \in D} b_{jd} x_d - \hat{y}_j \leq k_j \quad \forall j \in J \tag{3}$$

$$x_d \in \{0, 1\} \; \forall d \in D \tag{4}$$

$$\hat{y}_j \geq 0 \quad \forall j \in J \tag{5}$$

Constraints (2) guarantee that each trip is covered by at least one duty. Constraints (3) are the depot capacity constraints. Exceeding a depot's capacity k_j is captured in the penalty variable $\hat{y}_j$ and penalized in the objective Function (1) with a penalty cost factor of $\hat{c}_j$.

The set D can contain millions of feasible duties. Common solution approaches resort to a column generation approach, which iterates between solving the model (SC) restricted to a subset of all feasible duties and generating some duties that can improve the current schedule (see Kroon and Fischetti, 2001; Caprara et al, 2007, for examples of implementations at the Dutch and Italian railways, respectively).

In practice, duties often contain properties that are unpopular to train drivers. In the freight railway business, there are two main types of unpopular properties: duties starting in the early morning and duties that require hotel rests.

Because many freight trains are operated during night time, train drivers often start their duty in the late evening or early morning hours. At Freightrail, train drivers receive allowances for working at night times, and the total number of working hours at night times is limited by the labor contract. Working in night shifts that start in the late evening is widely accepted by train drivers. Duties that start in the early morning, on the contrary, are unpopular. Train drivers receive lower allowances for such duties than for duties that start in the evening. Additionally, there is scientific evidence that early morning shifts particularly interfere with the human circadian rhythm. Early morning shifts negatively affect the length of prior sleep (e.g., Knauth, 1993; Akerstedt, 2003). Some authors explain this effect with social demands in the late afternoon and early evening hours, which encourages morning shift workers to go to bed too late (Rosa et al, 1996). Other researchers attribute the shorter sleep length prior to an early morning shift to the circadian acrophase, i.e., the period at which the peak of a human's achievement potential is reached. For most people, this period is reached in the early evening hours, such that falling asleep in this period is especially difficult (Akerstedt, 1998); some authors refer to this period as the "forbidden zone" for sleep (Folkard and Barton, 1993). Additionally, early morning shifts conflict with the circadian nadir, i.e., the period at which the low point of a human's achievement potential occurs. Typically, this period is reached in the early morning hours. As a consequence, waking up in this period to start working is difficult (Akerstedt, 1998). Some authors deduce recommendations for the earliest starting time of a morning shift (e.g., Kecklund and Akerstedt, 1995, who recommend a starting time of 7 a.m. or later). However, there is no consensus on the exact starting times that should be avoided. At Freightrail, duties starting between midnight and 4 a.m. are considered as most contradicting to the circadian rhythm.

Duties with hotel rests are also unpopular. Because most trains operate at night times, many hotel rests are scheduled at day times. At Freightrail, about 7% of the duties of a typical crew schedule contain a hotel rest. Approximately 60% of those hotel rests include noon; the average duration of a hotel rest is 10.5 hours.

The basic set covering formulation (SC) used for optimizing crew schedules does not differentiate between popular and unpopular properties when generating duties for each crew depot. The resulting crew schedules can contain a large number of unpopular duties. Since the objective function does not include fairness components, some crew depots typically obtain a higher-than-average share of unpopular duties, while others have a smaller-than-average share.

In the past, Freightrail used penalty costs for unpopular duties to reduce the total amount of unpopularity in a crew schedule (see Jütte et al, 2011, for an example of reducing the number of hotel rests). Still, the smaller number of unpopular duties remained unevenly distributed among crew depots. At Freightrail, train drivers are organized in local work councils at the depot level which have to approve the duties that are assigned to their depot. The work councils typically show high aversion towards higher-than-average unpopularity assignments for their depots. (Smaller-than-average unpopularity assignments for their depot, on the contrary, apparently do not negatively influence the fairness perception at a work council.) When negotiating with the work councils, the company management must indicate that unpopular duties are similarly distributed between depots.

Unpopularity and fairness in crew scheduling has been addressed in the literature, but for problems that differ from the one that we consider. In most problems that are discussed in the literature, the total amount of unpopularity that has to be distributed among crew depots is known beforehand. When scheduling nurses in a hospital, for example, the number of night duties that have to be distributed among the nurses is part of the problem input. The fair share of unpopularity per group of crew members can be calculated a priori, and fairness can be easily integrated into the set covering formulation (SC) by adding fixed upper bounds on the amount of unpopularity attributed to each group (e.g., De Causmaecker and Vanden Berghe, 2011). For the sources of unpopularity we are analyzing in this paper, however, integrating fairness considerations is more difficult. Both the number of duties with hotel rests and the number of duties conflicting with the circadian rhythm are not known a priori, but are determined during optimization. We cannot establish fixed upper bounds on the amount of unpopularity attributed to each depot in our case. Instead, we will compare the amount of unpopularity attributed to a depot to the total amount of unpopularity of the current schedule at each iteration of our algorithm. Using soft constraints, we will penalize each positive deviation from this shifting reference value.

3 Literature Review

Fairness aspects have been analyzed for several decades. The literature distinguishes between output-based and intention-based fairness (e.g., Fehr and Schmidt, 1999). While the latter focuses on the purpose of an individual's action, output-based fairness solely judges a situation based on the result of an action. Adams (1965) provides an early analysis of output-based fairness and sheds light on the consequences of a distribution of goods that is "not meeting the norms of justice". The author underlines the findings of Homans (1961) that inequity in a social exchange is perceived whenever the ratio of output to input differs for the players involved. According to their research, inequity in favor of one player evokes feelings of guilt, while inequity in favor of the opponent results in anger. In line with these findings, Fehr and Schmidt (1999) define (output-based) fairness as "self-centered inequality aversion". They present a basic model to include fairness aspects in an individual's utility function: In a game with two players, the utility of one player is defined by this player's output, but linear deductions are made for an inequality in the output of the players (both in advantage and in disadvantage of the player in question). Typically, disadvantageous inequality is rated more severe than advantageous inequality, resulting in a comparatively larger aversion factor in the utility function. Bolton and Ockenfels (2000) refer to the equal division of payoffs among several players as the "social reference point" and embed the findings of Fehr and Schmidt (1999) in a comprehensive model to capture people's behavior in various experiments.

Because scheduling and rostering of employees is closely connected with the distribution of various types of work among individuals, discussions on the fairness of a schedule are inevitable in practice. Blöchliger (2004) presents a basic idea of integrating fairness aspects within staff scheduling. In a setting with n individuals, the author suggests to measure fairness as either the difference between the maximum and minimum outcome of all individuals, or as the standard deviation of all outcomes.

Most crew scheduling applications including fairness aspects can be found in the stationary context, where all crew members are present at the same physical location and can theoretically be assigned to all tasks for their qualification. Examples include scheduling nurses in a hospital (e.g., Millar and Kiragu, 1998; Bard and Purnomo, 2005; Maenhout and Vanhoucke, 2013) and ground personnel at an airport (e.g., Dowling et al, 1997; Mason et al, 1998; Stolletz, 2010). In the stationary context, crew scheduling and crew rostering problems are typically solved simultaneously. A common approach to include fairness in this process is to generate cyclic rosters, i.e., sequences of duties that are rotated among all crew members. At the end of a scheduling horizon, each crew member has then been assigned to each of the duties exactly once. More recent approaches rely on agent-based cooperative search to incorporate fairness constraints (see Martin et al, 2013).

Fairness models from stationary contexts can also be transferred to a mobility context, where crew members are not fixed to one location, but start their duty at one location, move around a network of locations during work and return to their home location at the end of the duty. In contrast to stationary contexts, problem instances in mobility contexts are typically much larger, such that crew scheduling and crew rostering must be treated sequentially. For crew rostering in mobility contexts, implementing cyclic rosters is still a common approach to include fairness aspects (e.g., Caprara et al, 1998). Within the literature of crew scheduling in mobility contexts, however, only few publications have yet brought up fairness considerations. Abbink et al (2005) show how to integrate fairness aspects in a crew scheduling implementation at Netherlands Railways. In their setting, sources of unpopularity include operating old rolling stock, operating trains in unpopular areas, and having a low variation within the duty. With their "sharing-sweet-and-sour" approach, the authors distribute popular and unpopular properties among crew depots by imposing fixed upper bounds on the amount of unpopularity per depot. Other applications indirectly reduce the unfairness of a schedule by explicitly limiting the amount of unpopularity that is allowed per duty (e.g., Schaefer et al, 2005, for an airline crew scheduling example).

In this paper, we include fairness aspects in a crew scheduling problem in a mobility context. Similar to the general approach of Fehr and Schmidt (1999), we assume that the utility of a group of crew members is reduced by the amount of inequity that is perceived compared to other groups, but we will focus on disadvantageous inequity. Rather than considering an interaction of individuals, we are dealing with fairness aspects between several groups of individuals and assume that the preferences of the individuals within one group are equal. We only analyze fairness aspects within crew scheduling. Unlike Abbink et al (2005), we refrain from imposing fixed upper bounds on the amount of unpopularity per depot. Instead, we compare the amount of unpopularity per depot to the changing total amount of unpopularity that is inherent in the current crew schedule during the course of the optimization. As we will see, higher fairness can be achieved not only by changing the assignment of a given amount of unpopularity among depots, but also by changing the total amount of unpopularity of the crew schedule.

4 Solution Approach

Prior to presenting our approach to integrate fairness considerations into crew scheduling, we provide an outline of the basic crew scheduling solution algorithm that is currently implemented at Freightrail. Details on the algorithm can be found in Jütte et al (2011).

Most state-of-the-art solution algorithms that are used to solve crew scheduling problems rely on column generation: The basic idea of this procedure is to stepwise add to the optimization problem only some variables that are known to improve the current best solution. The overall solution process then alternates between a master problem (to solve the optimization problem restricted to the number of variables that have been generated) and a sub problem (to generate new variables that can improve the solution to the current master problem); we refer the reader to Barnhart et al (1998) and Desrosiers and Lübbecke (2005) for a comprehensive introduction to column generation. For our specific railway crew scheduling problem, we have implemented the column generation algorithm **RCS** that is used for solving large real-world test instances with tens of thousands of trips. To deal with large problem sizes, we resort to solving the LP relaxation of the master problem and use a variable fixing technique to stepwise induce integrality in the course of the algorithm: If we do not find new variables in the sub problem, we choose some of the variables with non-integer values in the current solution and round up their value to 1. Re-solving the sub problem after each variable fixing phase, one can typically regain much of the solution quality that is lost with this heuristic procedure; see Gamache et al (1999) and Borndörfer et al (2001) for details on variable fixing.

In the following, we will focus on a single source of unpopularity. However, our model can easily be adapted to multiple sources of unpopularity.

For a given source of unpopularity, we define the *duty unpopularity* u_d as the amount of unpopularity that is inherent to duty d. As an example, considering hotel rests as the source of unpopularity, we set $u_d = 1$ if duty d contains a hotel rest and $u_d = 0$ else. (By union rules, the number of hotel rests per duty is limited to a maximum of one per duty.) For each crew depot $j \in J$, we calculate the *depot unpopularity* U_j as

$$U_j = \frac{\sum_d b_{jd} u_d x_d}{k_j}, \tag{6}$$

where we use the depot capacities k_j, i.e., the number of train drivers available at each depot, as weighting factors to account for the different sizes of the depots. At Freightrail, depot capacities are close to demand.

We denote the *schedule unpopularity* by U and compute it by

$$U = \frac{\sum_d u_d x_d}{\sum_j k_j}. \tag{7}$$

The schedule unpopularity corresponds to an average unit unpopularity across all depots when taking into account the total capacity across all depots.

Fehr and Schmidt (1999) define fairness as "self-centered inequity aversion". Considering a setting with two individuals A and B and outcomes O_A and O_B, respectively, the authors define the utility function of individual A as

$$Utility(A) = O_A - \alpha(O_B - O_A)^+ - \beta(O_A - O_B)^+, \tag{8}$$

where $x^+ = max(x, 0)$ for a real number x. The utility of individual A decreases linearly in $O_B > O_A$ and $O_B < O_A$. α and β are aversion factors accounting for inequity in disadvantage and advantage of individual A, respectively. Fehr and Schmidt (1999) assume advantageous inequity to have a smaller effect on the utility of an individual in general, i.e., $\beta < \alpha$. Still, advantageous inequity is assumed not to increase the utility of an individual, i.e., $\beta \geq 0$.

We build on the general concept of Fehr and Schmidt (1999) and define the *schedule unfairness* F as

$$F = \sum_j k_j (U_j - U)^+. \tag{9}$$

The terms $U_j - U$ measure the difference between the unpopularity of depot j and the average unpopularity of the schedule. This models the situation at Freightrail, where management compares the unfairness of duties assigned to a depot with the overall schedule unpopularity. Our discussions with Freightrail indicated that drivers are primarily concerned about an above average fraction of unpopular duties assigned to their depot; they seem much less concerned about a below average fraction. Therefore, we only include deviations of the unpopularity from the average unpopularity in the fairness measure if the depot unpopularity exceeds the average. In the fairness measure, we weight the fairness values of the individual depots by their capacities, because a large depot with high unfairness affects more drivers than a small depot.

The following *fairness formulation* is an extension of (SC) that takes the unpopularity and fairness concerns into account:

$$\text{minimize} \quad \sum_{d \in D} c_d x_d + \sum_{j \in J} \hat{c}_j \hat{y}_j + \sum_{j \in J} k_j \tilde{c} \tilde{y}_j \tag{10}$$

$$\text{subject to} \quad \sum_{d \in D} a_{td} x_d \geq 1 \quad \forall\, t \in T \tag{11}$$

$$\sum_{d \in D} b_{jd} x_d - \hat{y}_j \leq k_j \quad \forall\, j \in J \tag{12}$$

$$\frac{1}{k_j} \sum_{d} b_{jd} u_d x_d - \frac{1}{\sum_j k_j} \sum_{d} u_d x_d - \tilde{y}_j \leq 0 \quad \forall\, j \in J \tag{13}$$

$$x_d \in \{0, 1\} \; \forall\, d \in D \tag{14}$$

$$\hat{y}_j \geq 0 \quad \forall\, j \in J \tag{15}$$

$$\tilde{y}_j \geq 0 \quad \forall\, j \in J \tag{16}$$

Compared to the standard set covering formulation, we have added fairness Constraints (13): If, for depot j, the depot unpopularity U_j exceeds the schedule unpopularity U, the deviation of U_j from U is captured in a penalty variable $\tilde{y}_j$. The new penalty variables are included in the extended objective Function (10) with a penalty cost factor $k_j \tilde{c}$. Multiplying the standard penalty cost factor $\tilde{c}$ with the respective depot capacity k_j, we ensure that we penalize deviations in unpopularity proportionally to the sizes of the depots.

The fairness formulation above is still linear in the decision variables x_d, $\hat{y}_j$, and $\tilde{y}_j$, which is a prerequisite for applying this formulation when solving large-scale problem instances. Using the fairness formulation as the master problem within our RCS algorithm, we obtain the *railway fairness crew scheduling algorithm* ($\mathbf{RCS - F}$). In the following section, we analyze the performance of the $RCS - F$ algorithm regarding solution quality of the resulting crew schedule, unpopularity level, and fairness level obtained.

5 Computational Results

We tested our algorithm using data from Freightrail. Our test set consists of $3,911$ trips from $1,420$ trains operating between 106 stations. 33 of these stations are depots, i.e., locations where a duty can start or end. Each train driver is assigned to a distinct depot. In total, $2,033$ train drivers are available across all depots. Depot capacities vary between 1 and 335 train drivers. Figure 1 shows a histogram of the depot capacities.

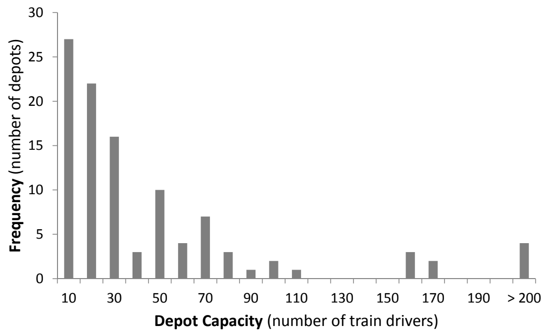


Fig. 1 Histogram of Depot Capacities

We will analyze the distribution of unpopular duties among the 33 depots and the effects of including fairness considerations on the properties of the solution schedule. First, we will choose conflicts with the circadian rhythm as the source of unpopularity. Then, we will move on to hotel rests.

To compare the outcome of our test runs, we will use three different key performance indicators: Net schedule costs z_{net}, schedule unpopularity U, and schedule unfairness F. The net schedule costs are the costs of the solution schedule excluding the fairness penalty costs and capacity penalty costs, i.e., $z_{net} = \sum_d c_d x_d$. Schedule unpopularity and schedule unfairness are defined in Equations 7 and 9, respectively.

All test runs were executed on a computer with an Intel Xeon processor, 8 cores, 2.13 GHz, 24 GB RAM, and a 64 bit operating system. The LP relaxation of the master problem was solved using the Barrier algorithm of CPLEX 12.1. All code was executed in parallel using the Open MP programming technique.

5.1 Circadian Rhythm Conflicts

Following the current approach of Freightrail, we consider duties starting between midnight and 4 a.m. as unpopular. We set $u_d = 1$ if duty d starts between midnight and 4 a.m., and set $u_d = 0$ otherwise. Then, U_j measures the proportion of unpopular duties at depot j, and U measures the proportion of unpopular duties within the schedule. F represents the number of unfairly assigned unpopular duties in the schedule, i.e., the number of unpopular duties that have been assigned to the depots beyond their respective fair share.

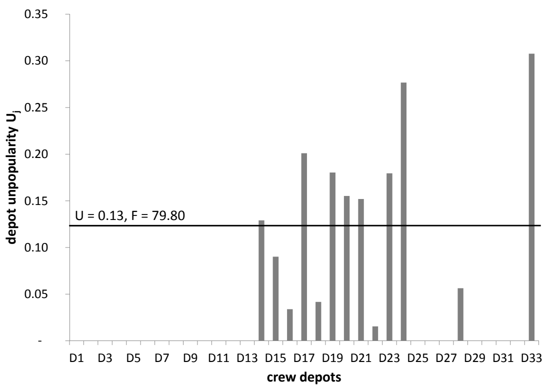


Fig. 2 Depot Unpopularities, Circadian Rhythm, Base Case

To obtain a benchmark solution, we applied the RCS algorithm, i.e., the solution algorithm without fairness constraints, to our test data. In the following, we will refer to this solution as *base case*. The RCS algorithm is a heuristic, but generates close-to-optimal solutions. In our numerical experiments, the costs of the crew schedule were less than 0.4% above a known lower bound on the optimal crew schedule. The resulting crew schedule consists of $1,658$ duties. Of those duties, 270 start between midnight and 4 a.m. The schedule unpopularity of the solution is $U = 0.13$, i.e., 13% of the duties start between midnight and 4 a.m. The unpopular duties are unevenly distributed among crew depots, with U_j ranging between 0 and 0.31, and the solution exhibits a schedule unfairness of $F = 79.80$. Figure 2 shows the depot unpopularities for the depots in the base case. Note that 20 depots have not obtained any unpopular duties, but depots $D33$ and $D24$ have 31% and 28% unpopular duties, respectively. It is not surprising that such a situation is considered unfair by the train drivers of the depots with high fractions of unpopular duties.

Next, we applied the $RCS - F$ algorithm to our test data. To analyze the effect of the fairness constraints on our three key performance indicators, we stepwise increased the fairness penalty costs $\tilde{c}$ from 0 to 3000. Fairness penalty costs of 1, for example, imply a 1 Euro increase in schedule costs for each unpopular duty that is unfairly assigned in the schedule. Figure 3 shows the results in terms of net schedule cost and level of fairness achieved. The x-axis of the graph shows the net schedule costs of a solution schedule in relation to those of the base case. A value of 101%, for instance, indicates that the net costs of the solution schedule are 1 percent above those of the base case schedule. The y-axis shows the corresponding unfairness values F. The diamonds represent the numerical solutions under different penalty costs. The line shows the corresponding exponential regression. The upper left point denotes the base case solution and has a net cost of $z_{net} = 100\%$ and a schedule unfairness of $F = 79.80$.

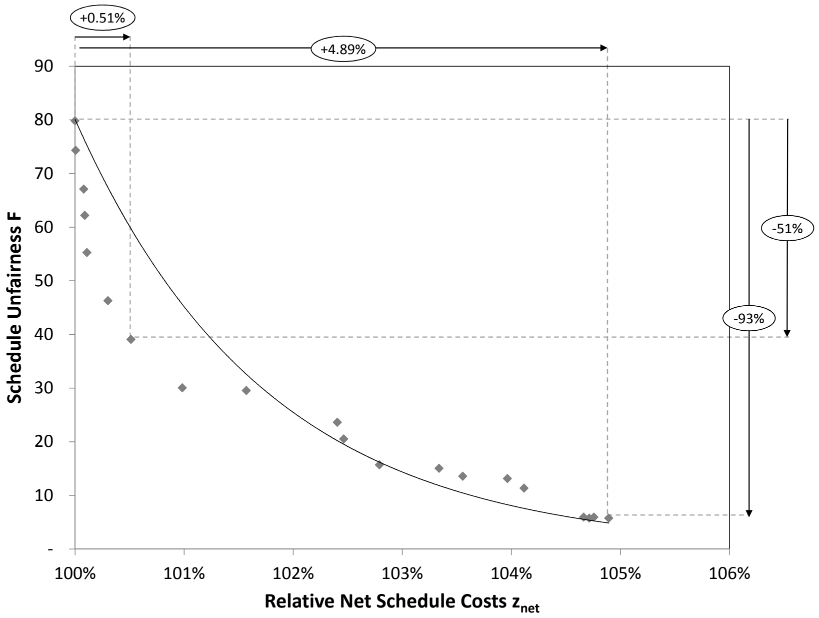


Fig. 3 Reduction of Schedule Unfairness, Circadian Rhythm

As we can see from the graph, the high level of unfairness in the base case schedule can be reduced from $F \approx 80$ to $F \approx 40$ at a moderate cost increase of 0.51%. Reducing the schedule unfairness further is relatively expensive. For example, a reduction of the schedule unfairness to a value of $F \approx 6$ increases net schedule costs by 4.89%, which corresponds to several millions of Euros per year. For some ranges of the fairness penalty costs, we observe particularly small improvements in fairness. Changing fairness penalty costs from 500 to 1000, for instance, reduces schedule unfairness only from 15 to 13, but increases costs by more than 1%.

Having examined the effects of decreased unfairness on the schedule costs, we will next shed light on the relation between schedule unpopularity and schedule unfairness for our test runs. In general, there are three means of decreasing the schedule unfairness F:

1. Redistributing the current schedule unpopularity U among the crew depots,
2. reducing the number of unpopular duties at depots with high depot unpopularity (which implies reducing U), and
3. increasing the number of unpopular duties at depots with small depot unpopularity (which implies increasing U).

Figure 4 shows the schedule unpopularity U (y-axis) as a function of the schedule unfairness F (x-axis) for our test runs. The upper right point corresponds to the base case solution with $F = 79.80$ and $U = 0.13$. As we can

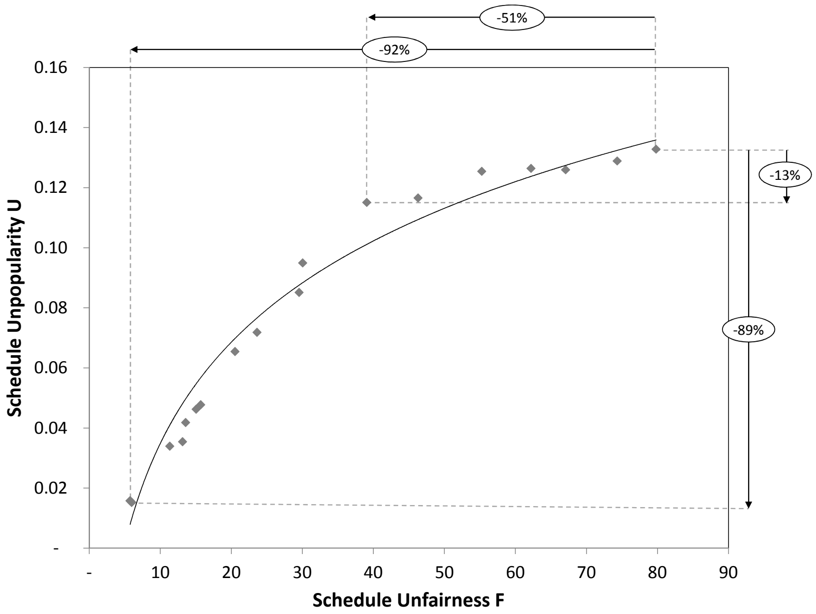
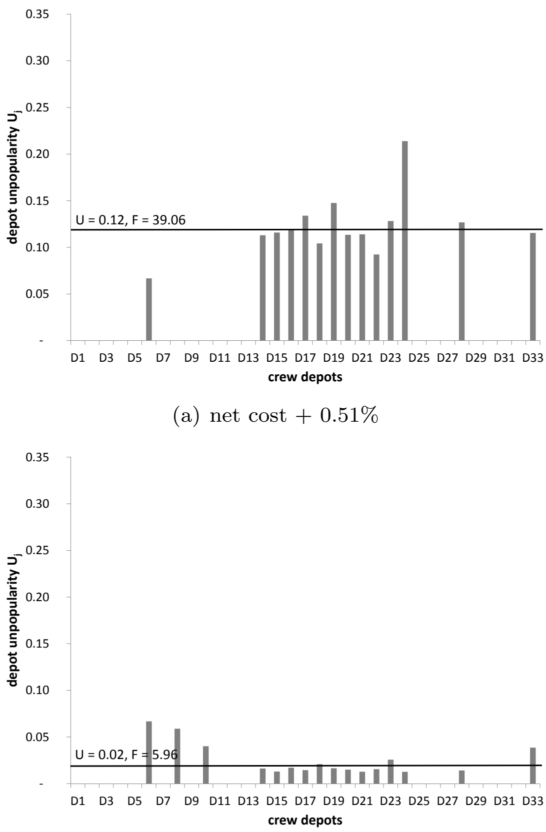


Fig. 4 Relation of Schedule Unpopularity and Schedule Unfairness, Circadian Rhythm

see from the figure, the schedule unpopularity U decreases with the schedule unfairness F. Reducing the schedule unfairness F by about half, however, only results in a small decrease of the schedule unpopularity U, while further reductions induce a large decrease of U. We conclude that the algorithm initially mainly redistributes the current unpopularity among the depots to reduce schedule unfairness. A further reduction of schedule unfairness is then achieved by reducing the number of unpopular duties in the schedule.

Figures 5(a) and 5(b) show the depot unpopularities for a reduction of the schedule unfairness F by 51% and 93% (at cost increases of 0.51% and 4.89%, respectively). While the schedule unpopularity U decreased only marginally from 0.13 to 0.12 in the first case (i.e., from 270 to 234 unpopular duties), it decreased substantially to 0.02 (i.e., 31 unpopular duties) in the second case.

Figure 6 shows the number of duties per starting time for the solution schedule of the base case (solid line) and for the solution schedule of the case with maximum achieved fairness (dashed line). As the figure shows, the small number of duties with starting times between midnight and 4 a.m. has been achieved by increasing the number of duties starting before midnight and after 4 a.m.



(a) net cost + 0.51%

(b) net cost + 4.89%

Fig. 5 Depot Unpopularities, Circadian Rhythm, Fairness Improvements

5.2 Hotel Duties

Hotel duties are the second main source of unpopularity among train drivers at Freightrail. In the European freight railway business, a hotel rest within a duty can last between 9 and 24 hours. Each duty can contain at most one hotel rest. Interviews with train drivers have shown that the length of the hotel rest is not a major driver of the unpopularity of a duty – if a train driver has to stay at a hotel within a duty, the train driver typically does not attach much relevance to the duration of the stay. We chose to use $u_d = 1$ if d contains a hotel rest and set $u_d = 0$ else.

The base case solution contains 112 hotel duties, which corresponds to a schedule unpopularity of $U = 0.06$. The hotel duties are distributed unevenly among the depots; the number of hotel duties per depot range between 0 and 34. The schedule unfairness of the base case solution is $F = 58.96$, i.e.,

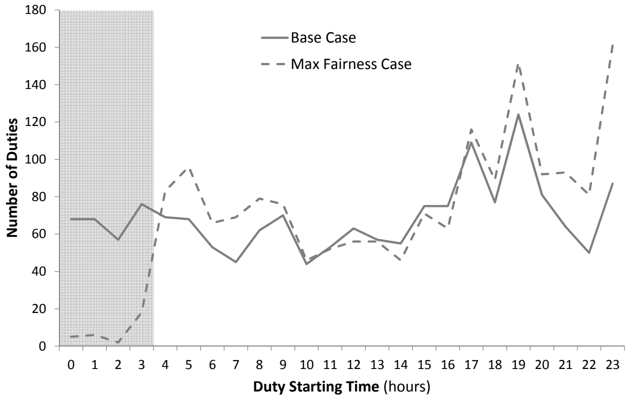


Fig. 6 Number of Duties per Starting Time, Circadian Rhythm

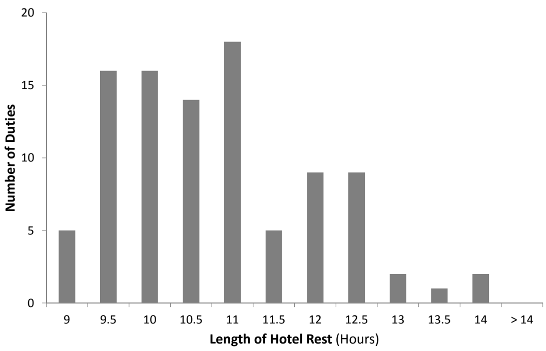


Fig. 7 Length of Hotel Rests, Base Case

approximately 59 of the 112 unpopular duties are unfairly assigned in the schedule.

When increasing the fairness constraint penalty factor, we observed a similar pattern as in the circadian rhythm case (see Figure 8): Initially, we can significantly reduce schedule unfairness with marginal increases in cost. With a net cost increase of 0.45%, for example, schedule unfairness can be reduced from 59 to 25 (−58%). Additional schedule unfairness reductions are substantially more expensive. Reducing schedule unfairness further from 25 to 4 (i.e., from 42% to 7% of the base case), for instance, requires a net cost increase from 0.45% to 2.55%.

The reduction of schedule unfairness is partly achieved by an overall reduction of schedule unpopularity (see Figure 9). Compared to the circadian

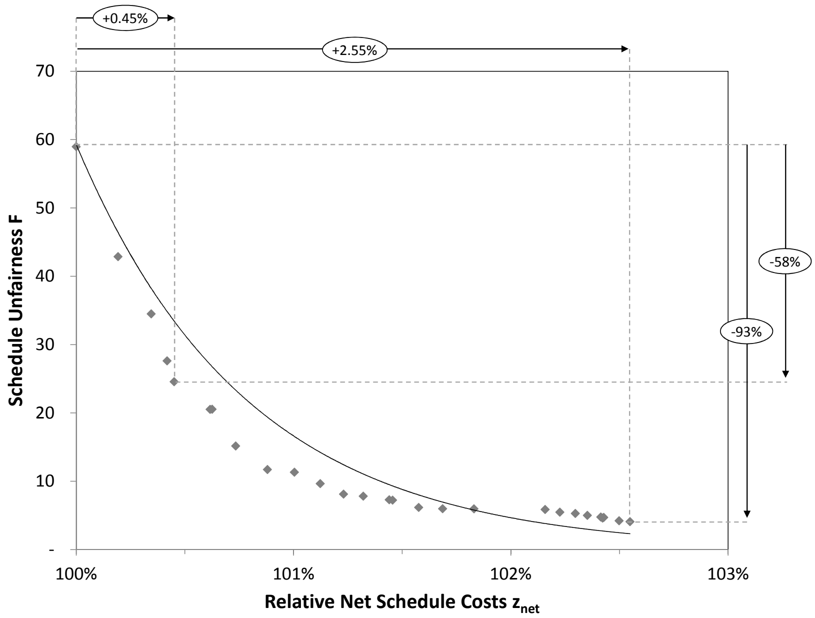


Fig. 8 Reduction of Schedule Unfairness, Hotel Duties

rhythm case, however, we do not observe a monotonous decrease of schedule unpopularity with decreasing schedule unfairness. Instead, there are situations where decreasing schedule unfairness is achieved by increasing schedule unpopularity. For example, to decrease schedule unfairness from $F = 28$ to $F = 25$, the algorithm increases the number of hotel duties from 55 to 67, which corresponds to an increase in schedule unpopularity of 22%. We attribute this behavior to the fact that some hotel duties are highly beneficial. Hence, when moderately increasing the fairness penalty costs, the algorithm rather adds additional hotel duties at depots with currently small share of unpopularity to obtain higher fairness. The highly beneficial hotel duties are only removed from the solution schedule if fairness penalty costs are further reduced.

5.3 Balancing Objectives

As we have seen in our test runs, depending on the value of the fairness penalty costs, the $RCS-F$ algorithm emphasizes either net schedule costs or schedule unfairness (and hence, in most cases, indirectly emphasizes schedule unpopularity). Judging from our computational results, we recommend the following steps to choose a fairness penalty cost value that balances net schedule costs, schedule unpopularity, and schedule unfairness:

First, management should decide on the maximum cost increase that is tolerated for a decrease of schedule unpopularity and schedule unfairness. Be-

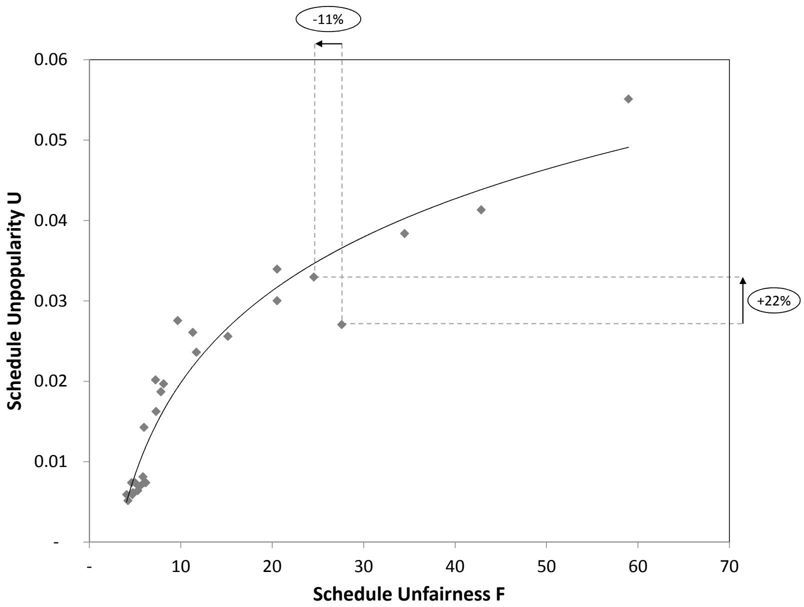


Fig. 9 Relation of Schedule Unpopularity and Schedule Unfairness, Hotel Duties

cause of the high absolute costs associated with a crew schedule, this limit is probably set rather low (see also Jütte et al, 2011). For decreasing schedule unpopularity and schedule unfairness regarding hotel duties, for example, Freightrail might impose a maximum tolerated cost increase of 0.5%.

Then, test runs with the $RCS - F$ algorithm with varying fairness penalty cost values can be used to find the maximum possible decrease in schedule unfairness within the tolerated cost limits. For small cost increases, we observed a high decrease in unfairness for our test runs. For a maximum tolerated cost increase of 0.5%, for example, the best solution of our test runs reduced unfairness regarding hotel duties from an initial value of 59 to a value of 25 (at a cost increase of 0.45%).

Finally, nearby solutions should be examined. If another schedule with similar schedule unfairness, but significantly smaller schedule unpopularity can be determined, this new schedule should be preferred over the current one. In our example, a slightly smaller fairness penalty cost value results in a schedule with schedule unfairness of 28 instead of 25, but only 55 instead of 67 unpopular duties, at a similar cost increase of 0.42%.

6 Conclusion

The traditional objective of railway crew scheduling is to generate train driver duties that cover all trips at minimal cost. However, the acceptance of a crew schedule by train drivers is essential to ensure successful operations. Train driver satisfaction with a schedule is partly determined by the unpopularity associated with the schedule and by the distribution of this unpopularity, and railway companies face a triad of objectives (see Figure 10): Minimizing net schedule costs, minimizing schedule unpopularity, and minimizing schedule unfairness. Minimizing the objectives simultaneously is not feasible; optimizing only one of the objectives typically results in an inferior solution with respect to at least one other objective.

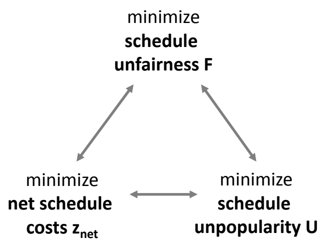


Fig. 10 Conflict of Objectives

We have analyzed the crew scheduling problem of a major European freight rail company. The company had focused on minimizing net schedule costs and schedule unpopularity; the issue of minimizing schedule unfairness had not been considered.

In this paper, we have described a mathematical model to capture fairness within the crew scheduling problem. We have modeled unfairness as the deviation of the actual share of unpopularity assigned to one group of train drivers from the overall average. Fairness constraints have been included by adding penalty costs in the objective function. We have explicitly modeled the objectives of minimizing net schedule costs and minimizing schedule unfairness. Schedule unpopularity, in contrast, has not been modeled explicitly, but was influenced indirectly by decreasing schedule unfairness.

We have conducted computational tests with a problem instance arising at the company that motivated our research. For two main sources of unpopularity, we found that the unfairness of a solution schedule could be reduced substantially at marginal increases in net schedule costs. Reductions of the schedule unfairness exceeding 50%, however, could only be achieved with large increases in net schedule costs. In most cases, increased fairness was achieved by either redistributing unpopular duties or reducing the total unpopularity of the schedule. Overall, we noticed a positive correlation between schedule unpopularity and schedule unfairness.

In our future research, we are planning to further analyze the relation between unpopularity and unfairness in the crew scheduling context. For a mere consideration of fairness aspects, increasing the schedule unpopularity to achieve higher fairness is a feasible approach and is not inferior to other approaches such as redistributing a fixed amount of schedule unpopularity among groups of train drivers. However, a smaller total number of hotel duties or circadian rhythm conflicts within a crew schedule contributes to the general public interest of all train drivers. To account for this thought, we intend to investigate how further bounding constraints on the schedule unpopularity affect the performance of our algorithm.

References

Abbink E, Fischetti M, Kroon L, Timmer G, Vromans MJCM (2005) Reinventing Crew Scheduling at Netherlands Railways. Interfaces 35(5):393–401

Adams JS (1965) Inequity in Social Exchange. Advances in experimental social psychology 2:267–299

Akerstedt T (1998) Is there an optimal sleep-wake pattern in shift work? Scandinavian Journal of Work, Environment & Health 24(suppl 3):18–29

Akerstedt T (2003) Shift work and disturbed sleep/wakefulness. Occupational Medicine 53(2):89–94, DOI 10.1093/occmed/kqg046

Bard JF, Purnomo HW (2005) Preference scheduling for nurses using column generation. European Journal Of Operational Research 164:510–534

Barnhart C, Johnson EL, Nemhauser GL, Savelsbergh MWP, Vance PH (1998) Branch-and-Price: Column Generation for Solving Huge Integer Programs. Operations Research 46(3):316–329

Blöchliger I (2004) Modeling staff scheduling problems. A tutorial. European Journal Of Operational Research 158:533–542

Bolton GE, Ockenfels A (2000) ERC: A Theory of Equity, Reciprocity, and Competition. The American Economic Review 90(1):166–193

Borndörfer R, Grötschel M, Löbel A (2001) Scheduling Duties by Adaptive Column Generation. Tech. Rep. 01-02, Konrad-Zuse-Zentrum für Informationstechnik, Berlin, Germany

Caprara A, Toth P, Vigo D, Fischetti M (1998) Modeling and Solving the Crew Rostering Problem. Operations Research 46(6):820–830

Caprara A, Kroon L, Monaci M, Peeters M, Toth P (2007) Passenger Railway Optimization. In: Barnhart C, Laporte G (eds) Handbooks in Operations Research and Management Science, Vol. 14 Transportation, Elsevier B.V., Amsterdam, chap 3, pp 129–187

Cordeau JF, Toth P, Vigo D (1998) A Survey of Optimization Models for Train Routing and Scheduling. Transportation Science 32(4):380–404

De Causmaecker P, Vanden Berghe G (2011) A categorisation of nurse rostering problems. Journal of Scheduling 14(1):3–16

Desrosiers J, Lübbecke ME (2005) A Primer in Column Generation. In: Desaulniers G, Desrosiers J, Solomon MM (eds) Column Generation, Springer, New York, chap 1, pp 1–32

Dowling D, Krishnamoorthy M, Mackenzie H, Sier D (1997) Staff rostering at a large international airport. Annals of Operations Research 72:125–147

Fehr E, Schmidt KM (1999) A Theory of Fairness, Competition, and Cooperation. The Quarterly Journal of Economics pp 817–868

Folkard S, Barton J (1993) Does the 'forbidden zone' for sleep onset influence morning shift sleep duration? Ergonomics 36(1-3):85–91

Folkard S, Tucker P (2003) Shift work, safety and productivity. Occupational Medicine 53(2):95–101

Gamache M, Soumis F, Marquis G, Desrosiers J (1999) A Column Generation Approach for Large-Scale Aircrew Rostering Problems. Operations Research 47(2):247–263

Harrington JM (2001) Health effects of shift work and extended hours of work. Occupational and Environmental Medicine 58(1):68–72

Homans GC (1961) Social behavior: its elementary forms. Harcourt, Brace

Jütte S, Albers M, Thonemann UW, Haase K (2011) Optimizing Railway Crew Scheduling at DB Schenker. Interfaces 41(2):109–122

Kecklund G, Akerstedt T (1995) Effects of timing of shifts on sleepiness and sleep duration. Journal of Sleep Research 4(suppl 2):47–50

Knauth P (1993) The design of shift systems. Ergonomics 36(1-3):15–28

Kroon L, Fischetti M (2001) Crew Scheduling for Netherlands Railways "Destination: Customer". In: VoßS, Daduna JR (eds) Computer-Aided Scheduling of Public Transport, Lecture Notes in Economics and Mathematical Systems 505, Springer, Berlin, Heidelberg, Germany, pp 181–201

Maenhout B, Vanhoucke M (2013) An integrated nurse staffing and scheduling analysis for longer-term nursing staff allocation problems. Omega 41(2):485–499

Martin S, Ouelhadj D, Smet P, Vanden Berghe G, Özcan E (2013) Cooperative search for fair nurse rosters. Expert Systems with Applications 40(16):6674–6683

Mason AJ, Ryan DM, Panton DM (1998) Integrated Simulation, Heuristic and Optimisation Approaches to Staff Scheduling. Operations Research 46(2):161–175

Millar HH, Kiragu M (1998) Cyclic and non-cyclic scheduling of 12 h shift nurses by network programming. European Journal Of Operational Research 104:582–592

Rosa RR, Härmä M, Pulli K, Mulder M, Näsman O (1996) Rescheduling a three shift system at a steel rolling mill: effects of a one hour delay of shift starting times on sleep and alertness in younger and older workers. Occupational and Environmental Medicine 53:677–685

Schaefer AJ, Johnson EL, Kleywegt AJ, Nemhauser GL (2005) Airline Crew Scheduling Under Uncertainty. Transportation Science 39(3):340–348

Smet P, Martin S, Ouelhadj D, Özcan E, Vanden Berghe G (2013) Fairness in nurse rostering. Tech. rep., University of Portsmouth, Portsmouth, United Kingdom

Stolletz R (2010) Operational workforce planning for check-in counters at airports. Transportation Research Part E: Logistics and Transportation Review 46(3):414–425