Dynamic Railway Crew Planning with Fairness over Time

B.T.C. van Rossum[1,⋆], T. Dollevoet[1], D. Huisman[1,2]

[1]Econometric Institute and Erasmus Center for Optimization in Public Transport
Erasmus University Rotterdam, The Netherlands

[2]Process quality and Innovation, Netherlands Railways
Utrecht, The Netherlands

⋆Corresponding author

vanrossum@ese.eur.nl, dollevoet@ese.eur.nl, huisman@ese.eur.nl

Econometric Institute Report Series, EI 2022-10

Abstract

Railway operators typically employ large numbers of drivers and guards, and are interested in providing them with fair and attractive working conditions. At Netherlands Railways (NS), the largest passenger railway operator in the Netherlands, this challenge is addressed through the use of Sharing-Sweet-and-Sour rules, which specify a fair allocation of sweet (attractive) and sour (unattractive) work over the different crew bases. While these rules are currently implemented at the crew base level and in the tactical planning phase, NS is considering formulating these rules at the individual level, in the operational planning phase, and over a given planning period. This gives rise to a new problem, which we call the dynamic railway crew planning problem with fairness over time. We propose a rolling horizon approach with a penalty-based feedback mechanism and a column generation heuristic to solve this problem. On several real-life instances from NS, including up to 572 unique crew members, this method is able to satisfy the individual rules for on average 95.2% of the employees.

Keywords: Railway optimisation, Integrated crew planning, Fairness over time, Column generation, Rolling horizon

1 Introduction

Railway operators typically employ large numbers of drivers and guards, who ensure that daily operations run smoothly and passengers experience high service levels. Apart from the general goal of improving employee well-being, there are several reasons why these operators are concerned with the attractiveness of the working conditions of their employees. First, it is important to remain an attractive employer and be able to attract new employees in today's highly competitive labour market. Second, better working conditions can reduce the amount of sick leave and increase overall performance. Third, a favourable working environment reduces the risk of conflicts and strikes, a threat that is particularly imminent in the railway sector where many workers are represented by strong labour unions. Recently, massive strikes in the United Kingdom and the Netherlands demonstrated the power of the unions [Topham, 2022, Albers and Nandramn, 2022].

At Netherlands Railways (NS), the largest passenger railway operator of the Netherlands, the challenge of shaping attractive working conditions is addressed through the use of Sharing-Sweet-and-Sour rules [Abbink et al., 2005]. These rules, initially developed to resolve large nationwide strikes in 2001, specify that sweet (attractive) and sour (unattractive) work should be allocated fairly over the 28 different crew bases of NS. For example, one such rule states that the average percentage of aggression work, i.e., work on trains with high risk of passenger aggression, must not exceed 50% for each crew base.

In the current crew planning process at NS, the Sharing-Sweet-and-Sour rules are incorporated in the tactical planning phase. Here, a generic crew schedule for the annual plan is constructed based on the yearly timetable and rolling stock schedule. This schedule contains *duties*, i.e., days of work, that are assumed to be performed every week of the year. The Sharing-Sweet-and-Sour rules require that the sweet and sour work in the generic schedule is distributed fairly over the different crew bases. Next, within each crew base duties are assigned to various roster groups. Each roster group consists of a set of crew members that rotate through a cyclic roster. Under the assumption of generic duties that are repeated every week, members of the same roster group thus perform the same work. Finally, in the operational phase the daily timetable and rolling stock schedule may deviate from the yearly ones to account for, e.g., planned maintenance works and events. As a result, some of the generic duties might be modified as well.

Unfortunately, the current implementation of Sharing-Sweet-and-Sour rules at NS suffers from two main drawbacks. First of all, since these rules impose restrictions on the attractiveness of work at the crew base level, they do not offer any guarantees on the attractiveness of any particular roster group. As such, large differences in the quality of rosters of individuals at the same crew base can persist. In practice this is especially stringent among roster groups of different seniority levels. Second, the rule set is imposed on the yearly schedule in the tactical planning phase, whereas the daily operated schedules might significantly deviate from this yearly schedule and differ between subsequent weeks. As a result, it is unclear to what extent the realised crew schedules actually satisfy the Sharing-Sweet-and-Sour rules.

To tackle these two issues, NS is currently designing a new crew planning process. Here, in the tactical phase a *capacity planning* is constructed, i.e., a general roster specifying the days and time windows at which each crew member is expected to work. In the operational planning phase this capacity planning is detailed to a crew schedule by assigning duties to specific crew members. Here, the inputs, i.e., the tasks to be performed on each day, are dynamically revealed, whereas it is required that duties are communicated to crew members well before the day of operation. In the new setting, Sharing-Sweet-and-Sour rules would be enforced at the individual level. This is now possible since duties are assigned directly to individuals. In addition, they would be enforced over a planning period of several months, as it is too restricting to enforce them on each particular day or week. As such, the new rules result in fairness over time between the various crew members. Moreover, it is less likely that operated duties deviate much from the planned ones, as the duties are constructed closer to the day of operation. All in all, the Sharing-Sweet-and-Sour rules in the new process would yield stronger and more reliable guarantees to crew members. An auxiliary benefit of this approach is the additional flexibility for the railway operator, as the exact work content can be determined closer to the day of operation.

The resulting problem, which we call the dynamic railway crew planning problem with fairness over time, features a challenging combination of characteristics that have rarely been studied jointly. In particular, it is a dynamic resource allocation problem without complete information, in which fairness over time with respect to multiple, conflicting attributes is to be attained across individuals, and in which the set of feasible allocations varies over time and between individuals. Static allocation problems with fairness considerations have received considerable attention in the literature, successful applications include crew rostering [Breugem et al., 2022a] and vehicle routing [Matl et al., 2018, 2019]. Lodi et al. [2021] study fair dynamic resource allocation under the assumption that the set of feasible allocations is fully known, time-invarying, and identical for all individuals. Bampis et al. [2018] relax this assumption and are therefore most closely related to our work. In their work, however, feasible allocations can be identified in polynomial time, whereas in our case it requires solving a $\mathcal{NP}$-hard crew scheduling problem. Existing literature thus does not provide any practical methods towards solving a large-scale practical problem with this challenging combination of characteristics. It is our goal to address this gap in the literature. In particular, we aim to provide an effective solution approach for attaining fairness over time in a large-scale railway application. Our methods can be generalised to other settings where work must be fairly allocated to individuals in a dynamic fashion.

The contributions of this paper are threefold. First, we introduce the dynamic railway crew planning problem with fairness over time. In other words, we formulate the problem that arises in the operational planning phase of the proposed crew planning process of NS and which combines several challenging characteristics. Second, to solve this problem we propose a rolling horizon approach with a penalty-based feedback mechanism and a column generation heuristic. The feedback mechanism is used to steer the work allocation towards one that respects the Sharing-Sweet-and-Sour rules. We propose an integrated method, which simultaneously determines the work content and assigns it to employees, as well as a sequential approach that tackles these two problems sequentially and forms a benchmark for our integrated approach. Third, we apply both approaches to large practical instances from NS with up to 572 crew members. The integrated approach is able to satisy the individual rules for on average 95.2% of the crew members, as compared to 86.7% in the sequential approach. Strikingly, this increase in satisfaction levels comes at negligible additional computational cost.

The remainder of this paper is structured as follows. We formally introduce the dynamic railway crew planning problem with fairness over time in Section 2. We provide an overview of relevant literature in Section 3, and in Section 4 we introduce our rolling horizon approach and feedback mechanism. We compare the integrated and sequential approaches on several real-life instances from Netherlands Railways in Section 5, and we conclude in Section 6.

2 Problem Description

We present an overview of the dynamic railway crew planning problem with fairness over time, as well as some practical considerations, in Section 2.1. We discuss the relevant duty and rostering rules in detail in Sections 2.2 and 2.3, respectively. Finally, in Section 2.4 we present the individual Sharing-Sweet-and-Sour rules.

2.1 Overview

We assume a planning period of given length, and consider crew planning in the operational planning phase. In this phase, all strategic and tactical crew planning decisions have already been taken. In particular, we assume that a capacity planning is given in the form of a template-based roster for each crew member. Here, a *template* specifies the day and time window in which an employee is scheduled to work. To allow some flexibility, this time window is larger than the length of a typical duty. Moreover, we assume that a set of Sharing-Sweet-and-Sour rules is given, specifying bounds on the attractiveness of work along several dimensions. We require that, for each crew member, the total work performed in the planning period satisfies all Sharing-Sweet-and-Sour rules, resulting in fairness over time across crew members.

The total work to be performed consists of a set of *tasks* that result from the timetable and rolling stock schedule. Each task has a start and end station, and start and end time. It is not known in advance which tasks are to be covered on any particular day of the planning period. Instead, the timetable and rolling stock schedule for specific days are constructed on a rolling horizon basis. This allows the railway operator some flexibility in accounting for, e.g., events and maintenance works. As a result, the tasks to be covered are revealed in a dynamic fashion. The goal of the dynamic railway crew planning problem with fairness over time is then to convert the template-based rosters to actual rosters by assigning to each template a matching duty, such that all tasks in the planning period are covered and the work of each crew member satisfies all Sharing-Sweet-and-Sour rules. Note that this is essentially a feasibility problem, since all crew costs are already determined by the capacity planning.

We conclude this section with some remarks on the practical setting of the dynamic crew planning problem. First, we note that the timetable and rolling stock schedule for a specific day are usually constructed four to six weeks in advance. The crew planning problem for this day would then need to be solved at least two weeks before the day of operations, as to leave enough time for planners to perform some post-processing and to communicate the final schedule to crew members. Second, while we assume that a fixed template-based roster is given for each crew member, in reality changes to the capacity planning may occur even in the operational planning phase. For example, crew members may request extra days off during the planning period. We do not consider such modifications in this work, but the first remark shows that such changes can be easily accomodated in our setting as long as they are known before the crew schedule is communicated. Thirdly, disruptions on the day of operation may cause the performed schedule to deviate from the planned schedule, potentially compromising our ability to satisfy the Sharing-Sweet-and-Sour rules. The effect of such disruptions on the ability to satisfy individual Sharing-Sweet-and-Sour rules is outside the scope of this paper.

2.2 Duty Rules

In the course of the planning period we detail the capacity planning by assigning a duty to each template. A *duty* is a sequence of tasks, constituting a day of work, that satisfies the following duty rules. First, subsequent tasks should end and start at the same station, respectively, and respect a minimum transition time. This transition time depends on whether two tasks are on the same rolling stock unit or not. Second, each duty should start and end at the same crew base. Third, duties should respect a maximum duty length. The maximum duty length depends on the start time of the duty, e.g., duties starting at the middle of the day have a longer maximum length than duties starting late at night. Finally, each duty should contain a meal break of at least half an hour at a station with a dedicated canteen. The time before and after the meal break should be at most 5.5 hours, ensuring that the meal break is positioned approximately halfway a duty.

2.3 Rostering Rules

We need to respect several roster rules in the dynamic crew planning problem. First, when assigning a duty to a crew member, this duty should match the template of the corresponding day in the roster of this crew member. In particular, the start and end time of the duty should be contained in the time window specified by the template. Second, there should be a minimum rest time of twelve hours between subsequent duties. Finally, we impose rotation constraints on duty patterns to increase the attractiveness of rosters. When the start times of two subsequent early templates are non-decreasing, we require that the start times of the duties assigned to these templates are non-decreasing too. Here, early templates are defined as those starting before 10:00. We say that a duty pattern satisfying this rule is forward rotating. Similarly, we require that the end times of duties in late templates are non-increasing when the end times of these templates are non-increasing. Here, late templates are defined as those ending after 20:00, and a duty pattern satisfying this constraint is called backward rotating. The two rotation constraints ensure that the rosters respect the employees' circadian rythms.

In practice, labour legislations and collective agreements specify additional rules that crew rosters should satisfy. In this paper, we assume that these need not be explicitly included in the dynamic crew planning problem for one of two reasons. First, there is a category of roster rules that can be incorporated in the construction of the template-based roster already and can therefore be discarded in the operational setting. Such rules include, among other things, the minimum number of days off and the minimum number of weekends off. Second, some roster rules govern special cases that do not occur regularly and are best incorporated implicitly. For example, one such rule specifies that each crew member performs at most 12 very long duties (at least 9 hours in length) per year. Instead of explicitly modelling this rule, we can keep track of the number of such duties that each crew member has performed until now, and we forbid the assignment of more such duties to members whose maximum has been reached.

2.4 Sharing-Sweet-and-Sour Rules

We consider Sharing-Sweet-and-Sour (SS&S) rules at the individual level, as compared to the current rules on crew base level. This implies that the fractions of sweet and sour work of each crew member are lower- and upper-bounded, respectively, thus guaranteeing a minimum level of attractiveness for each individual employee. The individual rules do not explicitly preclude some variation between crew members. However, choosing appropriate threshold values will automatically reduce such variation, and we do not consider variation problematic when the overall guarantee is tight.

Each individual SS&S rule concerns a specific *attribute*, i.e., a characteristic of work defining its (un)attractiveness. We assume that each piece of work can be given a *score* with respect to each attribute. Here, a piece of work refers to a set of tasks that can occur either in the form of a single task, duty, or roster. An individual SS&S rule then consists of a combination of an attribute and a lower or upper bound on the score of the roster performed in the planning period with respect to this attribute. We compute the score of a roster by averaging over all work in the roster. In practice, one might also consider a rolling window set-up where the score is computed over the duties performed in the last months.

We consider SS&S rules with regards to the following four attributes:

- *Duty length*: The duty length is defined as the difference in start and end times of a duty. Duties of shorter length are preferred.
- *Fraction of Type-A work*: Type-A work consists of desirable work on, for example, Intercity trains, which only stop at a limited number of stations, or modern rolling stock units.
- *Fraction of aggression work*: Work with a high risk of passenger aggression is undesirable. It occurs mostly in the metropolitan areas of the Netherlands.
- *Fraction of work on double-decker trains*: Work on double-decker trains is considered undesirable, as it requires guards to climb a large number of stairs.

3 Literature Review

Recently, several authors have started to study the theoretical properties of fairness over time in dynamic resource allocation problems. Lodi et al. [2021] consider the problem of achieving fair allocations over time, deriving some structural results regarding fairness-enhancing allocation policies. They apply their framework to the ambulance allocation problem. Salem et al. [2022] study the case where utility functions, governing the utility derived from a given allocation, vary over time. Bampis et al. [2018] consider feasible allocations that are not known in advance and might vary between individuals and time periods. They propose several offline and online approximation algorithms. While their setting most resembles ours, their methods do not translate to our case, where determining a feasible allocation requires solving a crew scheduling problem. To address this gap in the literature, we propose a rolling horizon approach with feedback mechanism that is able attain fairness over time in a large-scale crew planning application.

As such, our work belongs to the large body of literature on crew planning in public transport and airline operations. Caprara et al. [1999] and Abbink et al. [2018] offer excellent introductions to the field of railway crew planning. The crew planning problem is commonly decomposed into crew scheduling and crew rostering. We refer to Heil et al. [2020] for a recent survey on railway crew scheduling and to Kohl and Karisch [2004] for an introduction to airline crew rostering.

The dynamic crew planning problem we consider is situated after the tactical planning phase and before the day of operation. Several authors have studied a similar setting, mainly with the goal of minimising disturbances due to devations from the tactical planning. Stojković et al. [1998] consider the problem of generating modified airline crew pairings that minimise disturbances from a planned schedule. Huisman [2007] considers rescheduling railway crew due to maintenance works and solves this problem using a column generation heuristic. Breugem et al. [2022b] extend his work by integrating the rescheduling for multiple days simultaneously and providing a formulation for the integrated crew replanning problem. They solve the problem using valid inequalities and a column generation heuristic. Saddoune et al. [2009] solve an operational airline crew pairing problem using a rolling horizon approach and column generation, showing that schedules obtained in early planning phases have little added value when the flight schedule is not regular. Rählmann et al. [2021] perform robust tactical planning by choosing *duty frames* that are able to cover duties in different demand scenarios, i.e., that lead to robust solutions in the operational planning phase. In our work we consider crew scheduling in the operational phase given only a template-based capacity planning.

In this work we consider both duty and rostering rules, hereby partly integrating the crew scheduling and crew rostering problems. This integration has received relatively little attention in the literature. Ernst et al. [2001] propose an integrated model to compute cyclic and acyclic rosters. Their solution approach relies on a complete enumeration of all possible duties, and hence does not scale well beyond the sparse Australian railway network under consideration. Mesquita et al. [2013] develop a heuristic Benders decomposition method to solve the integrated problem of vehicle scheduling, crew scheduling, and crew rostering. Borndörfer et al. [2017] consider the integrated duty scheduling and rostering problem. Here, duties and rosters are linked through *duty templates*, coarse duty representations that are similar to our templates, and the problem is solved using Benders decomposition. Saddoune et al. [2012] solve the integrated crew pairing and crew assignment problem using column generation and dynamic constraint aggregation. Finally, Breugem et al. [2022b] propose a mathematical formulation of the integrated crew replanning problem and solve this using a column generation-based fixing heuristic.

The goal of the SS&S rules at NS is to obtain fair and attractive crew schedules, a topic that is gaining popularity in recent years (see Wolbeck [2019] for a review of personnel scheduling with fairness aspects). Most research in this area focuses on including fairness and attractiveness in the crew rostering problem. Hartog et al. [2009] describe the role of SS&S rules in the crew rostering process of NS. Borndörfer et al. [2015] and Maenhout and Vanhoucke [2010] propose heuristic solution methods for the crew rostering problem with personal preferences and fairness considerations, whereas Er-Rbib et al. [2021] tackle the problem of constructing cyclic rosters with group-based preferences through matheuristic approaches. Nishi et al. [2014] suggest a two-level decomposition for crew rostering, attaining fairness by minimising the maximum workload per roster group. Breugem et al. [2022a] propose a branch-price-and-cut algorithm to analyse the trade-off between fairness and attractiveness in crew rostering and evaluate this method on real-life instances from NS.

A smaller stream of literature considers fairness and attractiveness explicitly in the crew scheduling problem. Abbink et al. [2005] describe how the SS&S rules at NS are incorporated in the crew scheduling problem. Jütte et al. [2017] consider crew scheduling for a railway freight carrier where unpopular duties must be distributed fairly over different crew bases. Similar to the SS&S rules, constraints limit the unpopularity and unfairness per crew base. A related stream of literature considers the problem of balancing workload across routes in vehicle routing problems, see Matl et al. [2018] for an overview thereof. Matl et al. [2019] distinguish between variable-sum and constant-sum resources, i.e., resources whose total sum does or does not depend on the chosen resource allocation. We contribute to the applied fairness literature as we consider a dynamic setting with data that is revealed over time, and we are one of the few to consider fairness at the level of the individual crew member. Moreover, we consider both variable-sum (duty length) and constant-sum (all other attributes) resources.

4 Solution Approach

We propose a rolling horizon approach to cope with the dynamic nature of the problem. Similar approaches have been successfully applied to crew planning problems, though mainly with the aim of decomposing the problem of planning for a long period into multiple problems spanning a shorter period (see, e.g., Saddoune et al. [2009] and Quesnel et al. [2020] for applications to airline crew pairing). For each day of the planning period, we solve a crew planning problem that covers this day and potentially the next day. We fix the duties in the solution for the first day of the problem in our crew plan and proceed to the next day, until a crew plan for the entire planning period has been constructed.

Within each planning problem we place penalties on the assignment of work to individuals to steer towards satisfaction of individual SS&S rules at the end of the planning period. At each day we compute, for each crew member and each SS&S attribute, the scores of the work assigned to this crew member until now. When the current score is (nearly) outside the allowed interval, we place penalties on the assignment of more sour work, with regards to this attribute, to this particular crew member. By minimising the sum of penalties in each planning problem, we aim to establish a feedback mechanism that increases the quality of work of this individual and steers the SS&S score towards the allowed interval.

In each planning problem we need to (i) generate duties covering all tasks and (ii) assign these duties to crew members as to minimise penalties. In our proposed *integrated* approach we integrate these two steps: we generate duties that minimise penalties and simultaneously assign them to individual crew members. Since no obvious benchmark method for our problem exists, we propose the *sequential* approach, in which we do not integrate the two steps. Instead, we first generate duties that satisfy the SS&S rules at the crew base level only, thereby mimicking the current crew planning process of NS. Within each crew base, we then assign these duties to crew members in a way that minimises the sum of penalties.

The remainder of this section is structured as follows. We provide a mathematical formulation of the crew planning problem in Section 4.1 and describe the column generation heuristic used to solve the crew planning problems, as well as several acceleration strategies, in Section 4.2. We define the penalty function used in the feedback mechanism in Section 4.3. Finally, we present the integrated and sequential approaches in Sections 4.4 and 4.5, respectively.

4.1 Mathematical Formulation of the Crew Planning Problem

We now provide a mathematical formulation of the crew planning problem that is repeatedly solved in the rolling horizon approach. Recall that we consider problems spanning either one or two days. Here, we limit ourselves to the formulation of the single-day problems, as this is the set-up used to obtain most of our results. We show how it can be extended to multi-day problems by introducing coupling roster constraints in Appendix A.

Let t be the next planning day encountered in the rolling horizon approach, and denote by K_t and R_t the set of tasks to be covered and the set of crew members working on this day, respectively. For each crew member $r \in R_t$, let Δ_t^r be the set of duties that can be performed by this crew member on this day. This set is implicitly defined by the duty rules, the template in the roster, and the work of the previous day (through the rest time and rotation constraints). The objective coefficient of duty $\delta \in \Delta_t^r$, defined as the sum of penalties on its tasks, is given by c_{rt}^{δ}, and the binary parameter a_{tk}^{δ} indicates whether task $k \in K_t$ is covered by duty δ.

For each $r \in R_t$ the binary decision variable x_{rt}^{δ} indicates whether duty δ is assigned to crew member r on day t. The crew planning problem for day t can then be formulated as the following binary linear program:

$$\min \quad \sum_{r \in R_t} \sum_{\delta \in \Delta_t^r} c_{rt}^{\delta} x_{rt}^{\delta} \tag{1}$$

$$\text{s.t.} \quad \sum_{r \in R_t} \sum_{\delta \in \Delta_t^r} a_{tk}^{\delta} x_{rt}^{\delta} \geq 1 \qquad \forall k \in K_t \tag{2}$$

$$\sum_{\delta \in \Delta_t^r} x_{rt}^{\delta} = 1 \qquad \forall r \in R_t \tag{3}$$

$$x_{rt}^{\delta} \in \mathbb{B} \qquad \forall r \in R_t, \delta \in \Delta_t^r. \tag{4}$$

The objective (1) is to minimise the penalties of selected duties, thereby balancing the penalties of the different crew members. In this way we aim to obtain a duty assignment where sour work is shifted away from crew members currently not satisfying the rules to those that are currently well below the threshold values. Constraints (2) ensure that each task is covered at least once, allowing for potential overcoverage through deadheading, while constraints (3) guarantee that each crew member, scheduled to work on day t, is assigned exactly one duty. The domain of the decision variables is defined by constraints (4).

4.2 Column Generation Heuristic

We propose to solve all crew planning problems with a column generation heuristic (see Desaulniers et al. [2006] and Lübbecke [2010] for general introductions to column generation, and Huisman [2007], Potthoff et al. [2010], and Breugem et al. [2022b] for successful applications to railway crew planning problems). In particular, we relax constraints (4) to $x_{rt}^{\delta} \geq 0$ and initialise model (1)-(4) with slack variables for constraints (2) and (3) only. We refer to the resulting model as the restricted master problem (RMP). In each iteration of the algorithm, we solve the RMP with incumbent columns to optimality and solve a series of pricing problems to identify columns, i.e., duty variables, with negative reduced cost that could potentially improve the objective value. These columns are added to the RMP and the algorithm repeats until no such columns are found.

Let $\lambda_{tk} \in \mathbb{R}_{\geq 0}$ and $\pi_{rt} \in \mathbb{R}$ be the dual variables corresponding to constraints (2) and (3), respectively. The reduced cost of duty x_{rt}^{δ} is then given by

$$c_{rt}^{\delta} - \sum_{k \in K_t} \lambda_{tk} a_{tk}^{\delta} - \pi_{rt}. \tag{5}$$

The problem of finding the duty with most negative reduced cost can be modelled as a series of resource constrained shortest path problems (RCSPPs), one for each template. For each such problem a directed acyclic graph is constructed containing all tasks that lie within the time window specified by the template. The duals can be decomposed on the arcs in this graph, and, as we will explain in Section 4.4, so can the penalties used in the integrated approach.

We solve each RCSPP using a heuristic labelling algorithm with completion bounds [Breugem et al., 2022b]. Since we use a heuristic pricing algorithm, the RMP is not necessarily solved to optimality when the column generation algorithm terminates.

We apply several acceleration techniques to speed up the column generation algorithm (see Desaulniers et al. [2002] for an overview of commonly used techniques). First, to avoid the notorious tailing-off effect, we terminate the algorithm when the percentual decrease in objective value in two subsequent iterations is below a threshold parameter ϵ. Second, to limit the number of columns entering the RMP, we do not add all columns returned from the pricing problems, but only those that are sufficiently task-disjoint (see Breugem et al. [2022b] for a definition). Finally, we perform column management to limit the number of columns in the RMP. More specifically, every iteration we remove all columns that have had a reduced cost above a certain threshold for more than a given number of iterations.

The algorithm outlined above yields a fractional solution to the RMP, whereas in our rolling horizon approach we aim to fix duties for the current day in our crew plan. As such, we need to convert the fractional solution to a binary one. Necessitated by the large scale of our problem, we use a diving heuristic which proceeds as follows (see Joncour et al. [2010] for an overview of column generation based primal heuristics). Once the column generation algorithm terminates, all duties whose solution value is above a threshold parameter τ are fixed. The threshold value must be strictly larger than $\frac{1}{2}$ to avoid double assignments to the same crew member. If no such duties exist, we fix the duty with the highest solution value. The remaining problem, which is now strictly smaller in size, is re-solved using column generation and the fixing rule is applied again until the solution is completely binary. We remark that the strong heuristic nature of this approach may cause several tasks to remain uncovered.

4.3 Feedback Mechanism

Our feedback mechanism relies on a penalty function to penalise the allocation of more sour work to crew members whose work history does not satisfy the individual SS&S rules. Minimisation of penalties in the objective function of the model outlined in Section 4.1 should then lead to a more favourable work assignment for said crew members, thereby bringing their SS&S scores (closer) to the allowed interval. We compute penalties for all SS&S attributes separately. The penalty for a piece of work is defined as the sum of penalties over all attributes.

The penalty for a given combination of work, attribute, and crew member depends on two factors: the score of the work with respect to the attribute, and the score of the work history of the crew member with respect to the attribute. The penalty is increasing in the score of the work, as a higher score corresponds to less favourable work and is therefore more likely to drive the crew member away from satisfying the individual SS&S rules. The penalty is also increasing in the score of the work history, since we prioritise the allocation of attractive work to employees who have already performed relatively much unfavourable work.

We now formally define the penalty function. Without loss of generality, assume that the score of each attribute $a \in A$ takes values in the range $[0, r_a]$, and that the SS&S rule specifies that the score should be at most b_a, where $b_a < r_a$. The penalty $f_a(u, s)$ of assigning a piece of work with attribute score u to a crew member with work history score s is then defined as the product $f_a(u, s) = g_a(u) \cdot h_a(s)$. In our work we use $g_a(u) = \frac{u}{r_a}$. For $h_a(s)$ we use the piece-wise linear function illustrated in Figure 1. This piece-wise structure prioritises the assignment of sweet work to crew members whose current score is above the threshold, but also attempts to avoid the assignment of sour work to crew members that currently satisfy the SS&S rule yet risk crossing the threshold.

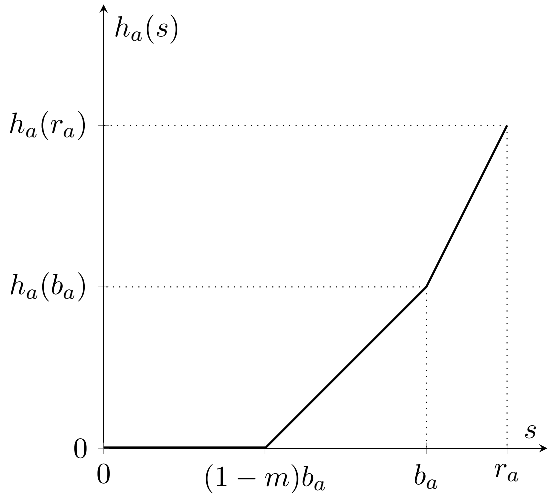


Figure 1: Example of penalty function.

The function h_a is mathematically defined as follows:

$$h_a(s) = \begin{cases} 0 & \text{if } 0 \leq s \leq (1-m)b_a, \\ \alpha(1-\gamma) \cdot \frac{s-(1-m)b_a}{r_a} & \text{if } (1-m)b_a < s \leq b_a, \\ h_a(b_a) + \alpha \cdot \frac{s-b_a}{r_a} & \text{if } b_a < s \leq r_a. \end{cases} \tag{6}$$

Here, $\alpha > 0$ is a scaling parameter, $m \in (0,1)$ is a parameter indicating how close the history score must be to the threshold for the penalty function to be activated, and $\gamma \in (0,1)$ regulates the scaling factor between penalties for crew members that currently satisfy the rule versus those that do not. The function evaluates to zero when $s \leq (1-m)b_a$, increases linearly in s when $s > (1-m)b_a$ and $s < r_a$, and increases linearly at a faster rate when $s > b_a$. Both g_a and h_a are normalised by dividing through r_a, to ensure that different attributes contribute equally to the objective function. While computational experiments showed that this function performs well, our approach allows for any choice of functional form for f_a.

4.4 Integrated Approach

In the integrated approach we integrate the generation of duties and the minimisation of penalties in each planning problem. We do so by placing penalties, computed using the penalty function described in Section 4.3, on the assignment of tasks to specific individuals in each crew planning problem. Each penalty is weighted by the duration of the task divided by the average length of a duty, to avoid small tasks contributing disproportionally to the objective function. The advantage of this approach is that the penalties can be easily decomposed on the arcs of the pricing graph without increasing the complexity of the pricing problems. A downside is that in general, due to the nonlinearities in the

SS&S attributes, the penalty of a duty does not equal the sum of penalties of its tasks. As such, the penalties used in the integrated approach constitute an approximation of the true penalties. The assignment of duties to crew members in the solution to the crew planning problem is used to build the crew plan in the rolling horizon approach.

We now illustrate the use of penalties in the integrated approach by means of a small example with two SS&S attributes, a single crew member, and a single duty. Assume that the only SS&S attributes are the fractions of aggression (agg) and double-decker (dd) work with thresholds $b_{\texttt{agg}} = 0.30$ and $b_{\texttt{dd}} = 0.50$, respectively. Furthermore, assume that the current scores of our crew member on these attributes are $s_{\texttt{agg}} = 0.29$ and $s_{\texttt{dd}} = 0.55$, respectively. In other words, the level of aggression work performed so far is slightly below the threshold, whereas the fraction of double-decker work is currently above the maximum allowed value. Finally, the duty we consider assigning to this crew member is given in Table 1: a round trip from Rotterdam (Rtd) to Utrecht (Ut), Nijmegen (Nm), Amsterdam (Asd), and back to Rotterdam.

Table 1: Example duty containing four tasks. Durations are denoted in minutes.

Task			Aggression work		Double-decker work	
Start	End	Duration	Duration	Fraction	Duration	Fraction
Rtd	Ut	37	0	0.00	37	1.00
Ut	Nm	54	0	0.00	54	1.00
Nm	Asd	82	23	0.28	82	1.00
Asd	Rtd	75	18	0.24	0	0.00
Total		248	41	0.17	173	0.70

We compute the penalties that the integrated approach will place on the assignment of this duty to our crew member using the following parameters for the penalty function: $\alpha = 1{,}000$, $m = 0.1$, and $\gamma = 0.5$. Recall that the penalty of a duty is the sum, over all attributes, of the penalties of all tasks. The penalty of a single task can be obtained using the penalty function $f = g \cdot h$ as described in Section 4.3, and is then weighted by the relative duration of the task. To weigh the relative contribution of each task, we assume an average duty length of four hours. For the aggression attribute, the function h evaluates to

$$h_{\texttt{agg}}(0.29) = 1{,}000 \cdot (1 - 0.5) \cdot \frac{0.29 - (1 - 0.1) \cdot 0.30}{1} = 10.$$

For each single task, the component g equals the fraction of aggression work as listed in Table 1. Summing over all tasks in the duty, we find that the total penalty for aggression equals

$$10 \cdot \left(\frac{37}{240} \cdot 0.00 + \frac{54}{240} \cdot 0.00 + \frac{82}{240} \cdot 0.28 + \frac{75}{240} \cdot 0.24\right) = 1.71.$$

Since the current score on double-decker work exceeds the threshold, the value $h_{\mathtt{dd}}(0.55)$ is computed as follows:

$$h_{\mathtt{dd}}(0.55) = 1{,}000 \cdot (1 - 0.5) \cdot \frac{0.50 - (1 - 0.1) \cdot 0.50}{1} + 1{,}000 \cdot \frac{0.55 - 0.50}{1} = 75.$$

The total penalty for double-decker work then equals

$$75 \cdot \left(\frac{37}{240} \cdot 1.00 + \frac{54}{240} \cdot 1.00 + \frac{82}{240} \cdot 1.00 + \frac{75}{240} \cdot 0.00\right) = 54.06.$$

The total penalty of this duty in the integrated method thus evaluates to 55.77. The relatively high contribution of double-decker work is caused by the high fraction of double-decker work in this duty in combination with the current score of the crew member exceeding the threshold.

4.5 Sequential Approach

In this section we propose the sequential method, which we will use to evaluate the relative quality of the integrated method. The sequential method consists of two steps. First, duties are generated by solving model (1)-(4) without the inclusion of any penalties in the objective function. Instead, we include SS&S rules at the crew base level to increase the likelihood that these duties allow for satisfaction of SS&S rules at the individual level. In the second step, the first stage solution can be improved by swapping duties between crew members. Here, similar to the integrated method, we place a penalty on the allocation of duties to crew members and aim to minimise the sum of penalties. A practical benefit of this approach is that it would not require any modifications to the way in which NS currently generates duties.

We now describe the sequential method in more detail. It starts by solving model (1)-(4), augmented with SS&S at the crew base level, using the column generation heuristic of Section 4.2. Let B be the set of crew bases, R_t^b be the set of crew members working from crew base $b \in B$ at day t. Moreover, let s_a^δ be the score of duty δ with respect to attribute a. Finally, let t_a^δ equal one in case a is the duty length attribute, and the total duration of tasks in δ otherwise. The SS&S rules then read

$$\frac{\sum_{r \in R_t^b} \sum_{\delta \in \Delta_t^r} s_a^\delta x_{rt}^\delta}{\sum_{r \in R_t^b} \sum_{\delta \in \Delta_t^r} t_a^\delta x_{rt}^\delta} \leq b_a \qquad \forall a \in A, b \in B. \tag{7}$$

We linearise constraints (7) as follows. First consider the case where attribute a is the duty length attribute. Let ℓ^δ be the length of duty δ. We then add the following constraints to model (1)-(4):

$$\sum_{r\in R_t^b}\sum_{\delta\in\Delta_t^r}(\ell^\delta-(1-\eta)b_a)x_{rt}^\delta\le p_{ab} \qquad \forall b\in B. \tag{8}$$

Here, $p_{ab}\ge 0$ is a slack variable which enters the objective function as a penalty, and ensures that a feasible solution always exists. The parameter $\eta\in(0,1)$ is used to slightly decrease the value of the SS&S threshold, since it is nearly impossible to satisfy SS&S rules at the individual level with duties that only just satisfy the rules at the crew base level. Computational experiments showed that $\eta=0.025$, i.e., a reduction of the threshold by 2.5%, worked well.

Now consider the case where attribute a is not the duty length, but the Type-A, aggression, or double-decker attribute. In this case we need to formulate the SS&S constraint at the task level. Let u_a^k be the total time task k contributes to attribute a, e.g., the total time of Type-A work, and w^k be the total work time in task k. We add the following constraints:

$$\sum_{r\in R_t^b}\sum_{\delta\in\Delta_t^r}x_{rt}^\delta\sum_{k\in K_t}a_{tk}^\delta(u_a^k-(1-\eta)b_aw^k)\le p_{ab} \qquad \forall b\in B. \tag{9}$$

Both types of constraints are easily amendable to the pricing problems: the duals of (8) can be decomposed over source and sink arcs, whereas the duals of (9) can be placed on the incoming arcs of each task.

Once a solution is obtained, we remove all columns whose solution value is zero. For each remaining duty variable we then iterate over all crew members, and we add a duplicate of this variable whenever a crew member is allowed to perform this duty. A crew member is allowed to perform the duty when it starts at his or her crew base, and when it is compatible with the template and work history of this crew member. Finally, we place a penalty on each duty variable, measuring how desirable it is to assign the duty to a particular crew member. In contrast to the integrated approach, these penalties are not approximations but equal the true penalties. We then re-solve this model with the requirement that all duty variables on the first day of the planning problem are binary, and fix the duties of the first day in our crew plan.

Note that the first step of the sequential approach mimicks the current crew planning process at NS, where SS&S rules are also included at the crew base level. The second step of the sequential approach effectively constitutes a form of post-optimisation: we allow for the exchange of duties between crew members as to minimise the sum of penalties, hoping to steer the allocation of work towards satisfaction of individual SS&S rules. We have also experimented with applying this post-optimisation on top of the integrated method, but this did not yield any significant gains.

We now return to the example from Section 4.4 to see what penalty would be used in the sequential method. The penalty is now computed over the score of the full duty, and not over the single tasks. The values $h_{\texttt{agg}}(0.29) = 10$ and $h_{\texttt{dd}}(0.55) = 75$ depend only on the work history and are therefore identical to those in the integrated approach. The attribute scores of the duty equal $g_{\texttt{agg}}(0.17) = 0.17$ and $g_{\texttt{dd}}(0.70) = 0.70$, respectively. The penalty of the duty therefore equals

$$10 \cdot 0.17 + 75 \cdot 0.70 = 54.20.$$

Note that this penalty is slightly below the one computed in the integrated approach.

5 Computational Experiments

In this section we show the effectiveness of the integrated approach by evaluating it on three real-life instances from Netherlands Railways. We describe the instances and the parameter settings in Sections 5.1 and 5.2, respectively. In Section 5.3 we show that, in contrast to the sequential method, the integrated method is able to achieve high satisfaction levels of individual SS&S rules. We further illustrate the strength of the integrated approach by analysing the results of one instance in detail in Section 5.4, and show that extending the optimisation horizon from one to two days has no apparent benefits in Section 5.5. In Section 5.6 we evaluate computation times, showing that the integrated approach is competitive with the sequential approach and fast enough for use in an operational setting. Finally, we justify our parameter settings with a sensitivity analysis in Section 5.7.

5.1 Instances from Netherlands Railways

We consider the problem of planning guards on three real-life instances from Netherlands Railways, which we call `Center`, `South-West`, and `East`. Each instance corresponds to a distinct region of the Netherlands and contains several crew bases that are geographically close to each other, as illustrated in Figure 2. Here, the tracks operated by Netherlands Railways are indicated by black lines, and the crew bases included in the instances are illustrated by white labels. Instance `Center` contains the crew bases Amersfoort (Amf), Amsterdam (Asd), and Utrecht (Ut). Instance `South-West` contains the crew bases Dordrecht (Ddr), The Hague (Gvc), Roosendaal (Rsd), Rotterdam (Rtd), and Vlissingen (Vs). Finally, instance `East` contains the crew bases Arnhem (Ah), Enschede (Es), Hengelo (Hgl), Nijmegen (Nm), and Zwolle (Zl).

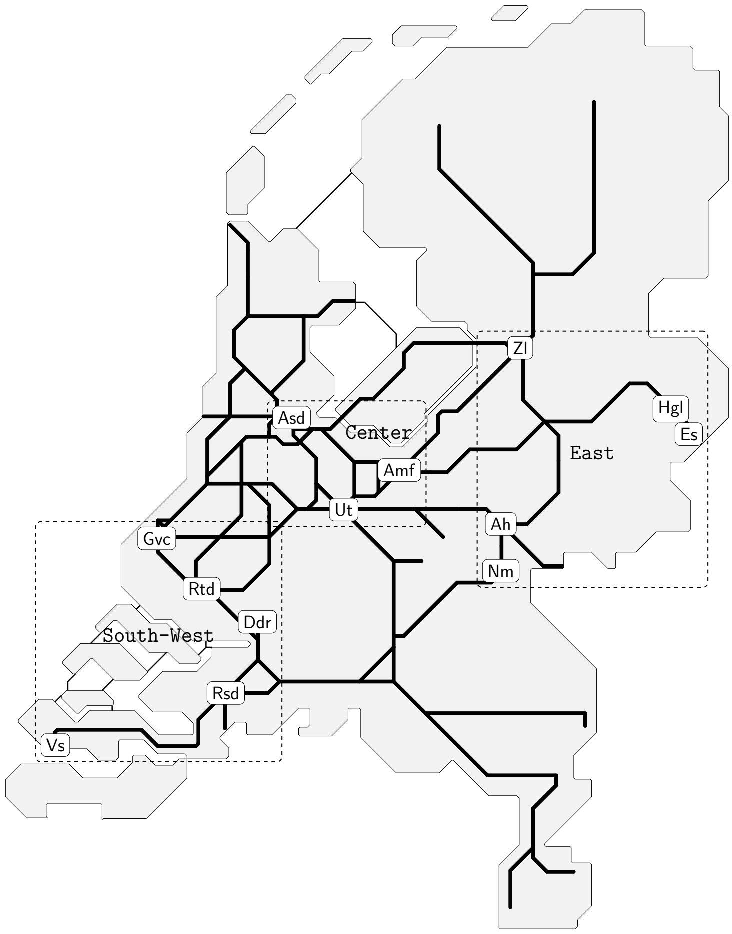


Figure 2: Location of the three instances on the Dutch railway network.

Our planning period spans the seven weeks from October 4 to November 21 of 2021. This period is representative of regular operations as the timetable was largely unaffected by COVID-19 measures. Since the proposed planning process is not in use at NS, we artificially construct the necessary inputs to our problem. We do so by converting the cyclic rosters to template-based individual rosters. The tasks to be performed on each day of the planning period are obtained from the realised crew plan of NS. For each instance and day, we group all duties that were actually performed by members of the crew bases of this instance on this day. The set of tasks to be performed then equals the set of tasks spanned by these duties.

Table 2 displays some characteristic of the three instances. For each crew base the number of crew members, number of templates in the rosters, and number of tasks to be performed in the 49-day planning period are shown. Moreover, for each crew base the average score of these tasks on SS&S attributes Type-A work, aggression work, and work on double-decker trains are given. These scores are given in grey as they are based on the realised crew plan, and are therefore not strictly spoken input to our problem. However, these scores do reflect the strong geographical variation in the distribution of sweet and sour work. Finally, totals and averages per instance are given in bold.

Of the three instances `South-West` is the largest, with a total of 572 crew members and 112,729 tasks to be performed over the seven-week planning period. In other words, the average daily crew scheduling problem contains 2,301 tasks, yielding a computationally challenging instance. We find that the ratios of templates to crew members and of tasks to templates are approximately constant over the different crew bases: the average crew member works 19 days during the planning period and the average duty contains about 10 tasks. Moreover, we find that sweet and sour work are not distributed evenly over the country. In particular, we find that crew bases located in `East`, an area that is not very densely populated, perform relatively more Type-A work. Work with high passenger aggression, on the other hand, is performed more frequently by crew members working in `Center` or `South-West` as their crew bases are located in the more metropolitan areas of the Netherlands. A strong discrepancy occurs even within instances, however: for example, the fraction of aggression work in Amsterdam (0.31) is almost twice that of Utrecht (0.17).

Table 2: Characteristics of the three instances.

Instance	Crew base	Crew	Templates	Tasks	Type-A	Aggression	Double-decker
`Center`	Amf	104	2,282	22,260	0.51	0.20	0.15
	Asd	230	4,235	39,696	0.34	0.31	0.30
	Ut	227	4,557	38,595	0.32	0.17	0.20
	Total	**561**	**11,074**	**100,551**	**0.37**	**0.23**	**0.23**
`South-West`	Ddr	61	1,267	15,522	0.37	0.17	0.41
	Gvc	203	3,605	41,861	0.41	0.22	0.16
	Rsd	80	1,778	19,547	0.48	0.11	0.47
	Rtd	186	2,639	27,584	0.38	0.22	0.34
	Vs	42	721	8,215	0.40	0.11	0.80
	Total	**572**	**10,010**	**112,729**	**0.41**	**0.19**	**0.34**
`East`	Ah	146	2,415	23,907	0.51	0.14	0.38
	Es	60	1,260	15,626	0.74	0.08	0.25
	Hgl	28	469	4,085	0.53	0.04	0.16
	Nm	90	1,974	20,658	0.48	0.13	0.45
	Zl	124	2,788	27,535	0.47	0.15	0.10
	Total	**448**	**8,906**	**91,811**	**0.52**	**0.13**	**0.27**

Finally, Table 3 shows the threshold values of the individual SS&S rules. For each instance the maximum average duty length, minimum fraction of Type-A work, maximum fraction of aggression work, and maximum fraction of work on double-decker trains are listed. We have chosen to let these values differ across instances, since Table 2 revealed sharp differences in the allocation of sweet and sour work over the different regions of the Netherlands and we aim to make each instance equally challenging. For example, we use a relatively high threshold value of 0.30 for aggression in `Center`, whereas the bound for double-decker work is highest at 0.50 in `South-West`. We use a higher average duty length in `South-West` and `East`, since here the ratio of tasks to templates is slightly higher than in `Center`. Generally, we have chosen bounds that are close to the average attribute scores of the work of each instance, but not too close as to become unattainable. In practice, a railway operator might prefer the use of one set of bounds that applies to each employee. We also note that, in a real-life setting, bounds would have to be chosen based on historical data.

Table 3: Sharing-Sweet-and-Sour rules of the three instances.

Instance	Duty length (h, $\leq$)	Type-A ($\geq$)	Aggression ($\leq$)	Double-decker ($\leq$)
`Center`	8:00	0.35	0.30	0.30
`South-West`	8:15	0.35	0.25	0.50
`East`	8:15	0.45	0.20	0.45

5.2 Settings

In order to apply the integrated and sequential methods described in Section 4, we need to choose values for three classes of parameters: (i) parameters that govern the column generation acceleration techniques, (ii) parameters that regulate the aggressiveness of the column generation algorithm and fixing heuristic, and (iii) parameters that define the penalty function. Note that, at least in theory, parameters in the first category affect the computation time but not the solution quality, whereas parameters in the second and third categories can affect both. In the first category, we use the following values. We return at most 50 columns per pricing problem, only select columns that are at least 50% task-disjoint from each other, and in each iteration we remove columns from the RMP whose reduced cost has been at least 250 in the last 5 iterations. In the second category, we choose to terminate the column generation process when the percentual increase in objective value is below $\epsilon = 0.01\%$, and we fix all duties whose solution value is at least $\tau = 0.55$. We justify the choice for these parameters with a sensitivity analysis in Section 5.7. Finally, for the penalty function we use a scaling constant $\alpha = 1{,}000$, a margin threshold $m = 0.1$, and a margin scaling factor $\gamma = 0.5$.

5.3 Satisfaction of Individual Sharing-Sweet-and-Sour Rules

We now analyse the practical performance of the integrated approach, i.e., to what extent it is able to construct rosters that satisfy the individual SS&S rules. Table 4 shows, for each combination of instance and method, the percentage of crew members satisfying each of the four SS&S rules. We will refer to this value as the satisfaction level. We find that the integrated approach yields high satisfaction levels on nearly all instances and attributes, attaining scores close to 100% on `Center` and `East`. The scores on instance `South-West` are slightly lower, with for example a satisfaction level of 88.1% on the double-decker attribute. This can be attributed to the fact that crew base Vlissingen

(Vs) is reachable through double-decker trains only (see Table 2). Moreover, we observe that the satisfaction levels of the sequential method are well below those of the integrated approach. This is especially the case for the duty length attribute, the only attribute whose total score is not independent of the generated duties and which thus benefits the most from an integrated approach. The differences are also large for attributes whose threshold are close to the instance average, such as Type-A in `Center` or aggression in `South-West`. Finally, we note that, on all instances and for both approaches, our heuristic leaves less than 1% of the tasks uncovered. Additional experiments indicated that this is largely caused by a discrepancy between the templates derived from the cyclical rosters and the true capacity demand during the planning period.

Table 4: Satisfaction levels (%) of the sequential (S) and integrated (I) approaches using a one-day optimisation horizon.

Instance	Method	Duty length	Type-A	Aggression	Double-decker
`Center`	S	64.6	85.1	86.6	88.0
	I	99.1	98.0	97.6	96.9
`South-West`	S	74.6	90.0	91.9	88.2
	I	83.2	91.8	95.8	88.1
`East`	S	87.7	92.8	97.3	93.5
	I	96.9	97.8	98.9	98.4

Additional computations show that violations of SS&S rules in the integrated approach are not clustered among crew members, i.e., there is no large group of crew members violating multiple rules. We find that, averaged over all three instances, 82.8% of the crew members satisfies each of the four SS&S rules. For 14.3% and 2.7% of the crew members we observe one or two rules being violated, respectively. It is for only 0.2% of the crew that three rules are simultaneously violated, and for all these crew members it holds that they performed less than ten duties over the planning period. In other words, there is little room for the feedback mechanism to provide them with a fair allocation of work. It never occurs that a crew member violates all four rules.

5.4 Detailed Results of `Center` Instance

We further illustrate the power of the integrated approach, as compared to the sequential one, by analysing in detail the characteristics of the solutions obtained on the `Center` instance. We choose this instance since (i) it contains the two largest crew bases in the Netherlands, Amsterdam and Utrecht, and (ii) the results in Table 4 showed a strong performance gap between the sequential and integrated approaches on this instance. In the following, we first show how the integrated approach is able to shift the distribution of the attribute scores of crew members towards the desired side of the threshold. We then show that the integrated approach converges to high satisfaction levels already halfway in the planning period. Finally, we explore the effect of the penalty mechanism on the work of individual crew members in detail and show that it is able to rapidly correct violations of SS&S rules.

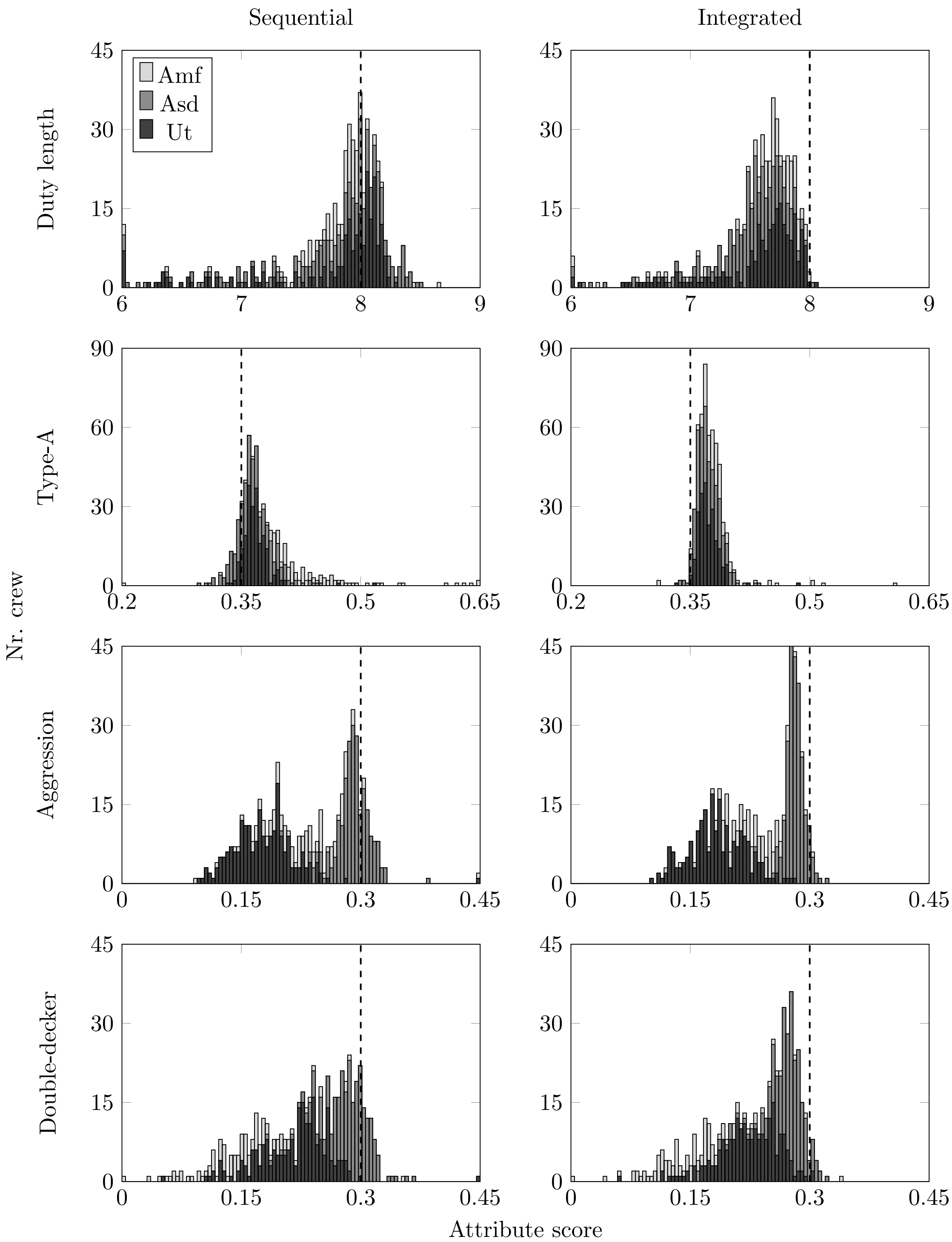


Figure 3: Distribution of Sharing-Sweet-and-Sour attribute scores over crew members of crew bases Amersfoort (Amf), Amsterdam (Asd), and Utrecht (Ut) of the sequential and integrated approaches on instance `Center`.

Figure 3 shows how the SS&S scores of the rosters of the `Center` instance, obtained using either the sequential or the integrated approach, are distributed over the crew members. Here, the attribute score and the number of crew members are displayed on the horizontal and vertical axes, respectively. The values of the three different crew bases are plotted in different colours. The threshold of each rule is indicated by a dashed line. We find that the distributions obtained using the integrated approach are located on the desired side of the boundaries and display a sharp peak right next to those boundaries. This is a direct result of the piece-wise structure of our penalty function. In contrast, the rosters obtained using the sequential approach are distributed on both sides of the threshold, illustrating the lower satisfaction levels of this method. Moreover, the distributions of the integrated approach display a lower degree of inequality between crew bases. This is nicely illustrated by the Type-A scores: while Amersfoort enjoys high Type-A scores compared to Amsterdam in the sequential approach (recall Table 2), the distribution of their scores are highly comparable in the integrated approach. Finally, as can be observed from the narrower range of the distributions, the inequality between individuals of the same crew base also tends to decrease when using the integrated approach.

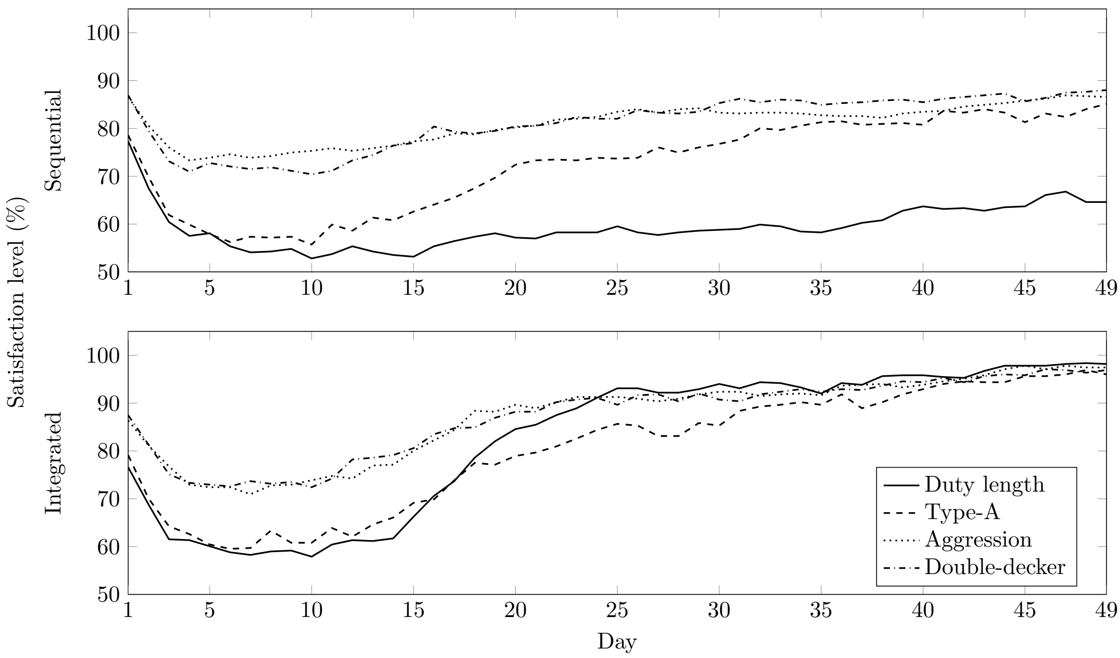


Figure 4: Temporal evolution of satisfaction levels of the sequential and integrated approaches on instance `Center`.

We now study how fast the satisfaction levels of the sequential and integrated approaches converge to acceptable levels. Figure 4 shows the temporal evolution of satisfaction levels of the rosters obtained using both methods. The days in the planning period are listed on the horizontal axis, whereas the satisfaction levels of each of the four SS&S rules are shown on the vertical axis. In both approaches, the satisfaction levels show a drop at the start of the planning period. Since we do not initialise the rosters with any work history, a single sour duty can cause a roster to violate rules for multiple days. Halfway through the planning period, however, all satisfaction levels of the integrated approach have reached values above 90%, after which they slowly rise to near 100% and flatten out near the end of the planning period. This last property is desirable in practice as it shows that stable satisfaction levels can be attained. The satisfaction level of Type-A work experiences a sharp drop at the start of the planning period and slightly lags behind the other attributes. This is most likely caused by the fact that the large crew bases Amsterdam and Utrecht naturally do not perform much Type-A work (see Table 2), and more time is needed to correct for this initial imbalance. The sequential approach fails to deliver compared to the integrated approach, since its satisfaction levels increase at a slower rate and plateau at a lower level.

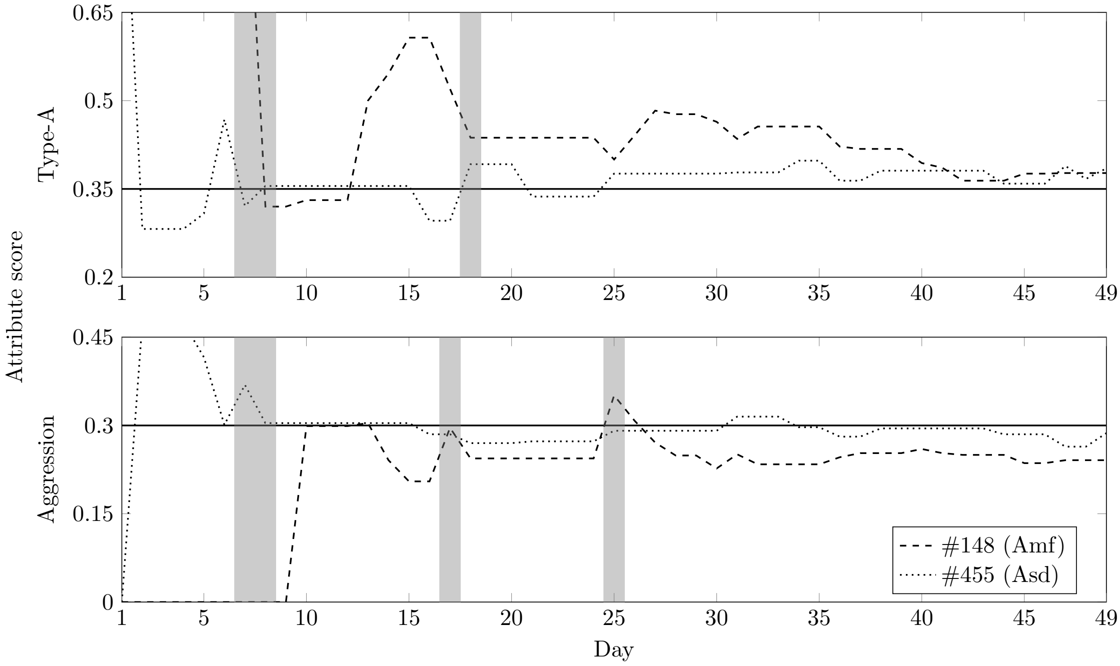


Figure 5: Temporal evolution of individual Sharing-Sweet-and-Sour scores of the integrated approach on instance `Center`.

Finally, we analyse the attribute scores of several individual rosters to gain a better understanding of the effect of penalties in the integrated approach. Figure 5 shows, for the SS&S attributes Type-A and aggression, the evolution of the scores of the rosters of crew members #148 and #455 who work at crew bases Amersfoort and Amsterdam, respectively. The day in the planning period is listed on the horizontal axis and the attribute score on the vertical axis. The solid line indicates the threshold of the respective attribute. Finally, we highlighted periods of interest, to be discussed in the next paragraph, in grey.

We find that the scores heavily fluctuate at the start of the planning period and slowly dampen towards the end of the planning period, a logical result of the fact that attribute scores are averaged over a larger number of duties. When the attribute score of a crew member exceeds the threshold, we find that, usually already in the next duty, this score is driven towards the desired side of the boundary. Examples include the Type-a score of #455 on days 7 and 18, and the aggression score of #148 on day 25, respectively. This shows that the penalties used in the integrated approach are an effective way of steering towards satisfaction of SS&S rules. Moreover, we observe the same pattern of rapid convergence as before: from day 27 onwards, both crew members satisfy both SS&S rules, except for a single spike in the aggression score of #455 at day 31. We also find that our method is able to steer on multiple attributes simultaneously, see for example the development in Type-A and aggression scores of crew member #455 from day 7 to day 8. In addition, the margin in our penalty function successfully prevents attribute scores from crossing the thresholds, as illustrated by the aggression score of guard #148 on day 17. Finally, we note that differences between individuals do persist throughout the planning period. In this instance this is especially the case for aggression work, reflecting the initial imbalance across crew bases. Given our choice of penalty function, there is no mechanism in our algorithm that actively corrects such differences.

5.5 Effect of Two-Day Optimisation Horizon

We now analyse the potential benefits of increasing the optimisation horizon, i.e., the number of days included in each crew planning problem, from one to two days. In doing so we only consider the integrated approach, as it has shown to outperform the sequential approach in all dimensions. Table 5 shows, for each combination of instance and horizon, the percentage of crew members satisfying each of the four SS&S rules. The results show that there is no apparent benefit in enlarging the optimisation horizon from one to two days, as the satisfaction levels are not significantly affected. This might indicate that, due to the scale of the instances, there is sufficient capacity and flexibility in the rosters, and that explicitly incorporating look-ahead information about roster rules therefore does not yield any improvements. Given that computation times are significantly higher when using a two-day horizon, as we will show in Section 5.6, we conclude that it is preferable to use a one-day optimisation horizon.

Table 5: Satisfaction levels (%) of the integrated approach using an optimisation horizon of one versus two days.

Instance	Horizon	Duty length	Type-A	Aggression	Double-decker
`Center`	1	99.1	98.0	97.6	96.9
	2	98.1	96.7	98.5	95.5
`South-West`	1	83.2	91.8	95.8	88.1
	2	84.4	92.6	94.2	89.6
`East`	1	96.9	97.8	98.9	98.4
	2	97.3	97.5	98.0	98.2

5.6 Computational Performance

In this section we analyse the computational performance of our methods to determine whether they are suitable for use in an operational setting. To this extent, note that all experiments were conducted using CPLEX 20.1.0 and four cores of an Intel Xeon Gold 6130 processor. Table 6 shows the computational performance of the sequential approach with a one-day optimisation horizon, and the integrated approach with a one- or two-day horizon, to each of the three instances. For each combination of instance, method, and horizon, the average computation time per day of the planning period is shown. This is composed of time spent solving pricing problems, solving the RMP, adding and removing columns to and from the RMP, and swapping duties between crew members (i.e., the second step of the sequential method). In addition, the average number of column generation iterations and number of iterations in which columns are fixed are shown.

We find that the computation times of the integrated approach with a one-day horizon, which proved to be the preferable configuration, are more than reasonable for use in a daily, operational setting. Even on the largest instance `South-West`, the average total computation time per day does not exceed eight minutes. By far the largest share of computation time is required to solve the RMP, followed at a distance by the time needed to solve pricing problems. Integrating the duty generation and assignment does not come at a large computational cost, since the sequential method is only marginally faster than the integrated one. Interestingly, in each iteration it spends significantly less time in the RMP and more time in the pricing problem. As expected we find that performing the post-optimisation step in which duties are swapped requires little computation time. Finally, we observe that computation times of the integrated approach increase drastically when the optimisation horizon is enlarged. This effect is mainly driven by an increase in the RMP time per iteration, as the total number of iterations stays relatively constant.

Table 6: Computational performance of the sequential (S) and integrated (I) approaches. The values are averaged over all days of the planning period.

			Time (s)					Iterations	
Instance	Method	Horizon	Total	Pricing	RMP	Columns	Swap	Total	Fixing
`Center`	S	1	217.2	68.1	143.6	5.3	0.2	103.7	14.1
	I	1	228.0	39.2	182.8	6.0	-	111.3	12.0
	I	2	1,333.6	177.8	1,135.0	20.8	-	109.8	11.5
`South-West`	S	1	447.7	153.0	284.3	10.3	0.1	178.6	15.0
	I	1	452.2	94.6	347.7	9.9	-	165.7	13.8
	I	2	1,714.8	282.8	1,404.0	28.0	-	169.2	12.5
`East`	S	1	123.0	33.7	85.6	3.6	0.1	93.6	13.5
	I	1	145.5	28.7	112.6	4.2	-	94.5	11.1
	I	2	569.7	75.7	483.0	11.0	-	85.4	9.8

5.7 Effect of Column Generation Parameters

In order to cope with the magnitude of our crew planning problems, we have used a rather aggressive configuration of our column generation heuristic. In particular, we used a fixing threshold of $\tau = 0.55$, and have terminated column generation iterations when the improvement in objective value between consecutive iterations was below $\epsilon = 0.01\%$. In this section we experiment with varying these parameters to analyse the trade-off between computation time and solution quality. Figure 6 shows, for various combinations of τ and ϵ, the average satisfaction levels and average total computation time per day of the planning period when applying the integrated approach with a one-day optimisation horizon on the `Center` instance. Here, parameters τ and ϵ refer to the fixing threshold and threshold on improvement in objective value, respectively.

We find that prematurely terminating the column generation iterations, i.e., using any $\epsilon > 0$, drastically reduces the computation times of the algorithm. In particular, computation times are more than halved when switching from $\epsilon = 0.00$ to $\epsilon = 0.10$. While increasing ϵ from 0.00 to 0.01 does not seem to have a strong effect on the satisfaction levels, we find that further increasing ϵ does reduce the solution quality. The effect of the fixing threshold τ appears to be less pronounced, and while a higher threshold generally increases the computation time, it need not necessarily lead to a higher solution quality. We conclude that our current parameter setting $(\tau, \epsilon) = (0.55, 0.01)$ is on the Pareto front and adequately compromises solution quality and computation time.

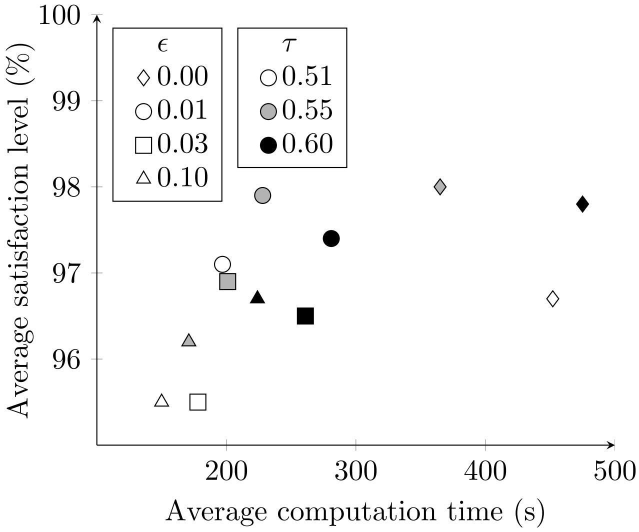


Figure 6: Effect of column generation parameter on satisfaction levels and computation times of the integrated approach on instance `Center`. Satisfaction levels are average over all four attributes, and computation times are averaged over all days of the planning period.

6 Conclusion

In this paper we consider fairness over time in a large-scale public transport application. In particular, we introduce the dynamic railway crew planning problem with fairness over time. Here, the goal is to assign duties to crew members over the course of a planning period, such that all tasks are covered and the total work of each crew member satisfies all Sharing-Sweet-and-Sour rules. We propose sequential and integrated rolling horizon approaches to solve the problem, and evaluate both methods on three large real-life instances from NS spanning a seven-week planning period. We find that the integrated approach is able to achieve high satisfaction levels and significantly outperforms the sequential approach at negligible additional computational cost. The integrated approach is able to accurately and rapidly steer towards a fair allocation of attractive and unattractive work, and it converges to high satisfaction levels already halfway in the planning period.

Our approach has high potential for railway operators, as it shows that tight guarantees on the attractiveness of individual work schedules can be made. The set-up we consider can be modified in several interesting directions. First, our approach allows for differential preferences across crew bases or even crew members. Our method can be used to evaluate the impact of using one set of nationwide bounds, instead of the regional bounds, which reflect the geographical distribution of sweet and sour tasks, used in this work. In addition, one could consider a setting where senior employees receive tighter guarantees, or where employees agree to relax Sharing-Sweet-and-Sour rules in return for a monetary compensation. Second, one could evaluate the Sharing-Sweet-and-Sour rules on a rolling window basis instead of over a fixed planning period. The stable performance of the integrated approach towards the end of planning period is certainly promising in this respect. Finally, as our approach is situated in the operational planning phase, it allows for dynamic classifications of attractive and unattractive work.

The work in this paper raises some relevant practical questions. While we assume that individual template-based rosters are given, it is interesting to investigate how such rosters can be constructed from scratch in a clever manner. Moreover, we have not yet explored the efficiency gains arising from the fact that the operator can determine the exact work content closer to the day of operation. Next, while in this work our only goal is to ensure that all individuals satisfy a certain rule set, is interesting to what extent a penalty-based approach is able to reduce the remaining variation between crew members. Finally, it is important to analyse the effect of disruptions, i.e., discrepancies between the planned schedule and the operated schedule, on the ability to satisfy Sharing-Sweet-and-Sour rules. We expect that our penalty mechanism largely mitigates this effect.

While our work shows the high potential of the penalty mechanism, more sophisticated solution methods are required to apply this approach in practice, e.g., on the full Dutch railway network. One promising option is to use a modified version of the powerful Lagrangian-based solver currently used by NS [Abbink et al., 2011]. Whereas it is hard to explicitly include roster rules in this solver, our results showed this to be of little added value. Alternatively, one could apply a form of dynamic constraint aggregation [Saddoune et al., 2012] or geographical decomposition techniques [Jütte and Thonemann, 2012].

Acknowledgements

We would like to thank Gábor Maróti, Joël van 't Wout, and Maaike Vollebergh of Netherlands Railways for their fruitful suggestions and support throughout the research project leading to this work.

References

E. Abbink, M. Fischetti, L. Kroon, G. Timmer, and M. Vromans. Reinventing crew scheduling at Netherlands Railways. *Interfaces*, 35(5):393–401, 2005.

E. Abbink, L. Albino, T. Dollevoet, D. Huisman, J. Roussado, and R.L. Saldanha. Solving large scale crew scheduling problems in practice. *Public Transport*, 3(2):149–164, 2011.

E. Abbink, D. Huisman, and L. Kroon. Railway crew management. In *Handbook of optimization in the railway industry*, pages 243–264. Springer, 2018.

M. Albers and A. Nandramn. Vrijdag in heel Nederland geen NS-treinen wegens staking, bonden onvermurwbaar. *De Volkskrant*, September 2022. URL https://www.volkskrant.nl/nieuws-achtergrond/vrijdag-in-heel-nederland-geen-ns-treinen-wegens-staking-bonden-onvermurwbaar~bbb3a673/.

E. Bampis, B. Escoffier, and S. Mladenovic. Fair resource allocation over time. In *AAMAS 2018-17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 766–773. International Foundation for Autonomous Agents and Multiagent Systems, 2018.

R. Borndörfer, M. Reuther, T. Schlechte, C. Schulz, E. Swarat, and S. Weider. Duty rostering in public transport - Facing preferences, fairness, and fatigue. Technical report, ZIB, 2015.

R. Borndörfer, C. Schulz, S. Seidl, and S. Weider. Integration of duty scheduling and rostering to increase driver satisfaction. *Public Transport*, 9(1):177–191, 2017.

T. Breugem, T. Dollevoet, and D. Huisman. Is equality always desirable? Analyzing the trade-off between fairness and attractiveness in crew rostering. *Management Science*, 68(4):2619–2641, 2022a.

T. Breugem, B.T.C. van Rossum, T. Dollevoet, and D. Huisman. A column generation approach for the integrated crew re-planning problem. *Omega*, 107:102555, 2022b.

A. Caprara, M. Fischetti, P.L. Guida, P. Toth, and D. Vigo. Solution of large-scale railway crew planning problems: The Italian experience. In *Computer-Aided Transit Scheduling*, pages 1–18. Springer, 1999.

G. Desaulniers, J. Desrosiers, and M. Solomon. Accelerating strategies in column generation methods for vehicle routing and crew scheduling problems. In *Essays and surveys in metaheuristics*, pages 309–324. Springer, 2002.

G. Desaulniers, J. Desrosiers, and M. Solomon. *Column generation*, volume 5. Springer Science & Business Media, 2006.

S. Er-Rbib, G. Desaulniers, I. Elhallaoui, and P. Munroe. Preference-based and cyclic bus driver rostering problem with fixed days off. *Public Transport*, 13(2):251–286, 2021.

A.T. Ernst, H. Jiang, M. Krishnamoorthy, H. Nott, and D. Sier. An integrated optimization model for train crew management. *Annals of Operations Research*, 108(1):211–224, 2001.

A. Hartog, D. Huisman, E. Abbink, and L. Kroon. Decision support for crew rostering at NS. *Public Transport*, 1(2):121–133, 2009.

J. Heil, K. Hoffmann, and U. Buscher. Railway crew scheduling: Models, methods and applications. *European Journal of Operational Research*, 283(2):405–425, 2020.

D. Huisman. A column generation approach for the rail crew re-scheduling problem. *European Journal of Operational Research*, 180(1):163–173, 2007.

C. Joncour, S. Michel, S. Ruslan, D. Sverdlov, and F. Vanderbeck. Column generation based primal heuristics. *Electronic Notes in Discrete Mathematics*, 36:695–702, 2010.

S. Jütte and U.W. Thonemann. Divide-and-price: A decomposition algorithm for solving large railway crew scheduling problems. *European Journal of Operational Research*, 219 (2):214–223, 2012.

S. Jütte, D. Müller, and U.W. Thonemann. Optimizing railway crew schedules with fairness preferences. *Journal of Scheduling*, 20(1):43–55, 2017.

N. Kohl and S.E. Karisch. Airline crew rostering: Problem types, modeling, and optimization. *Annals of Operations Research*, 127(1):223–257, 2004.

A. Lodi, P. Olivier, G. Pesant, and S. Sankaranarayanan. Fairness over time in dynamic resource allocation with an application in healthcare, 2021. URL https://arxiv.org/abs/2101.03716.

M.E. Lübbecke. Column generation. *Wiley Encyclopedia of Operations Research and Management Science. Wiley, New York*, pages 1–14, 2010.

B. Maenhout and M. Vanhoucke. A hybrid scatter search heuristic for personalized crew rostering in the airline industry. *European Journal of Operational Research*, 206(1): 155–167, 2010.

P. Matl, R.F. Hartl, and T. Vidal. Workload equity in vehicle routing problems: A survey and analysis. *Transportation Science*, 52(2):239–260, 2018.

P. Matl, R.F. Hartl, and T. Vidal. Workload equity in vehicle routing: The impact of alternative workload resources. *Computers & Operations Research*, 110:116–129, 2019.

M. Mesquita, M. Moz, A. Paias, and M. Pato. A decomposition approach for the integrated vehicle-crew-roster problem with days-off pattern. *European Journal of Operational Research*, 229(2):318–331, 2013.

T. Nishi, T. Sugiyama, and M. Inuiguchi. Two-level decomposition algorithm for crew rostering problems with fair working condition. *European Journal of Operational Research*, 237(2):465–473, 2014.

D. Potthoff, D. Huisman, and G. Desaulniers. Column generation with dynamic duty selection for railway crew rescheduling. *Transportation Science*, 44(4):493–505, 2010.

F. Quesnel, G. Desaulniers, and F. Soumis. Improving air crew rostering by considering crew preferences in the crew pairing problem. *Transportation Science*, 54(1):97–114, 2020.

C. Rählmann, F. Wagener, and U.W. Thonemann. Robust tactical crew scheduling under uncertain demand. *Transportation Science*, 55(6):1392–1410, 2021.

M. Saddoune, G. Desaulniers, and F. Soumis. A rolling horizon solution approach for the airline crew pairing problem. In *2009 International Conference on Computers & Industrial Engineering*, pages 344–347. IEEE, 2009.

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

T. S. Salem, G. Iosifidis, and G. Neglia. Enabling long-term fairness in dynamic resource allocation, 2022. URL https://arxiv.org/abs/2208.05898.

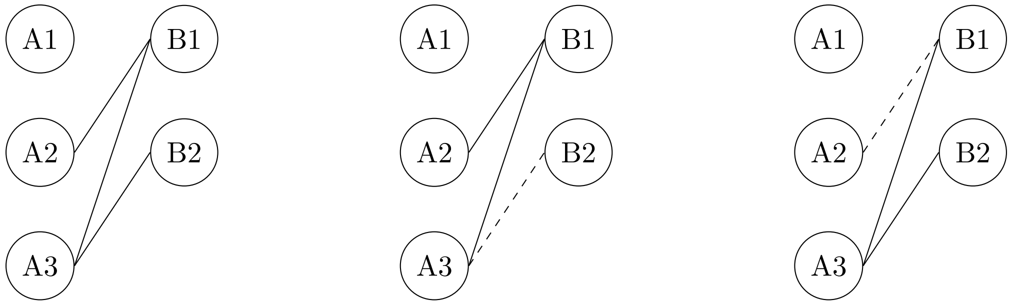
M. Stojković, F. Soumis, and J. Desrosiers. The operational airline crew scheduling problem. *Transportation Science*, 32(3):232–245, 1998.

G. Topham. Great Britain faces biggest rail strike in 30 years. *The Guardian*, June 2022. URL https://www.theguardian.com/business/2022/jun/20/great-britain-faces-biggest-rail-strike-in-30-years-starting-on-tuesday.

L.A. Wolbeck. Fairness aspects in personnel scheduling. Diskussionsbeiträge, 2019. URL http://hdl.handle.net/10419/210988.

A Coupling Constraints for the Multi-Day Crew Planning Problem

We now show how model (1)-(4) can be extended to span multiple days. In such a setting the roster rules, i.e., the minimum rest time and rotation constraints, act as coupling constraints between pairs of subsequent days. Following Breugem et al. [2022b] we model these rules using clique inequalities. Such inequalities are obtaining by enumerating maximal bicliques in the bipartite violation graph connecting duties whose joint selection would yield a constraint violation. Here, a maximal biclique is defined as a clique in a bipartite graph that is not fully contained in any other clique. Let P be the set of pairs of subsequent days in the crew planning problem, and denote by R_p the set of crew members working on both days of pair $p \in P$. Finally, denote by Q_p^r the set of clique constraints linking days in p for crew member r, and let binary parameter $o_{rpt}^{\delta q}$ indicate whether duty $\delta \in \Delta_t^r$, for $t \in p$, appears in constraint $q \in Q_p^r$.



(a) Violation graph. **(b)** First maximal biclique. **(c)** Second maximal biclique.

Figure 7: Modelling the forward rotation constraint using maximal bicliques.

Figure 7 presents an example of how maximal bicliques are obtained from the violation graph corresponding to the forward rotation constraints. We assume that the duty of a given crew member starts with either task A1, A2, or A3 on the first day, and with task B1 or B2 on the subsequent day. Moreover, we assume that all tasks start before 10:00. The bipartite violation graph corresponding to the forward rotation constraint of this crew member, as shown in Figure 7a, connects start tasks of the first day with all earlier start tasks of the second day. This graph contains two maximal bicliques, shown in Figures 7b and 7c, respectively. Each such biclique yields a clique inequality stating that at most one duty with a start task in the clique is assigned to this crew member.

Let T be set of days in the multi-day problem. The formulation of the multi-day crew planning problem now reads:

$$\min \quad \sum_{t \in T} \sum_{r \in R_t} \sum_{\delta \in \Delta_t^r} c_{rt}^{\delta} x_{rt}^{\delta} \tag{10}$$

$$\text{s.t.} \quad \sum_{r \in R_t} \sum_{\delta \in \Delta_t^r} a_{tk}^{\delta} x_{rt}^{\delta} \geq 1 \qquad \forall t \in T, k \in K_t \tag{11}$$

$$\sum_{\delta \in \Delta_t^r} x_{rt}^{\delta} = 1 \qquad \forall t \in T, r \in R_t \tag{12}$$

$$\sum_{t \in p} \sum_{\delta \in \Delta_t^r} o_{rpt}^{\delta q} x_{rt}^{\delta} \leq 1 \qquad \forall p \in P, r \in R_p, q \in Q_p^r \tag{13}$$

$$x_{rt}^{\delta} \in \mathbb{B} \qquad \forall t \in T, r \in R_t, \delta \in \Delta_t^r. \tag{14}$$

Here, the clique inequalities are given by constraints (13). Note that these constraints are easily amendable to the pricing problem, as their duals can be placed on source or sink arcs of tasks in the bicliques. In the multi-day problem we use the same diving rule as before, i.e., we fix duties of the first-day of the planning problem until the first-day solution is completely binary.