

# A Solution Approach for Railway Crew Scheduling with Attendance Rates for Multiple Networks



Julia Heil

## 1 Introduction

Since the 1990s, with the advances of computational power, computer-aided crew scheduling has gained momentum in the railway industry. By applying state-of-the-art Operations Research techniques, operators generate cost-efficient schedules, which fulfill desired levels of robustness and employee satisfaction at the same time. In recent years, as the size of railway crew scheduling problems increases in order to meet operators' practical needs, researchers developed solution methods for large-scale optimization in this field [4].

In the German regional railway passenger transport, state transport authorities tender railway networks publicly. Network parameters, such as geography, lines, timetable, quality of rolling stock and others, are specified in the transportation contract. Based on this contract and further restrictions, e.g., labor tariffs, railway operators plan their operations and make an offer (“planning for tender”). After offer acceptance, the nominated operator operates the network for the contractual period with recurring operational planning on (half-)yearly basis. Due to this process, operators plan and operate each single network separately. Only in few cases, based on their knowledge and experience, planners create duties across multiple networks manually. Assuming potential for further cost reduction, our research project investigates the effect on personnel cost of scheduling crews for multiple networks collectively.

Typically a crew in regional railway passenger transport consists of two types of workers: a train driver who operates the train and a conductor who, for instance, controls tickets or secures departures. In Germany, the conductors' presence is not

---

J. Heil (✉)  
TU Dresden, Dresden, Germany  
e-mail: [jheil@tu-dresden.de](mailto:jheil@tu-dresden.de); [julia.heil@mailbox.tu-dresden.de](mailto:julia.heil@mailbox.tu-dresden.de)

required for 100% of the train trips: State transport authorities define attendance rates per network, which vary depending on product type, time window and others. The attendance rate  $a$  is satisfied when the  $a$ -share of all kilometers assigned to attendance rate  $a$  are covered by at least one conductor and can be defined as  $a \leq \text{kilometers}_{a,attended} / \text{kilometers}_{a,total}$ .

Our work is based on the multi-period railway crew scheduling problem with attendance rates for conductors developed by Hoffmann et al. [3]. A hybrid column generation approach, which solves the pricing problem via a genetic algorithm, is applied. In this work, we present the extension to the multiple network problem and develop a two-phase optimization method to solve the corresponding very large-scale railway crew scheduling problem in reasonable time. We discuss computational experiments with a real case of 12 networks and determine the effect of multi-network crew scheduling on personnel cost.

## 2 Problem Definition

Crew scheduling is part of the operational planning process at a railway operator. Its objective is to generate a set of feasible duties, a *schedule*, which covers all trips at minimum cost. A *trip* symbolizes the atomic unit of a train run between two *relief points*. At a relief point, crew members can change vehicle, at some breaks are allowed. Hence, a *duty* is a combination of trips, which starts and ends at a *crew depot*. We define a single railway *network* as the entity of train lines with its corresponding parameters determined by the transportation contract. The *union of multiple networks* can be interpreted as a synthetic single network, which usually is very large and/or complex. Evidently, unifying single networks is only reasonable if they are interlinked, i.e. two networks share at least one relief point.

Both the entire schedule and an individual duty are limited by a number of various restrictions. This includes operational conditions, legal and work regulations, e.g., maximum duty time or break time rules, and contractual terms of the transportation contract, such as attendance rates. For simplification, we assume that the same restrictions apply to all networks with the exception of attendance rates, which must be satisfied for each network individually.

### 2.1 Mathematical Formulation

We formulate the multi-network crew scheduling problem with attendance rates as set covering model. Let  $M$  denote the set of trips  $i$  and  $N$  the set of duties  $j$ . For network-specific requirements, we define  $R$  as the set of networks  $r$  and  $G_r$  as the set of required attendance rates  $g$  of network  $r$ . Furthermore, let  $d_i$  be the travel distance of trip  $i$  and  $c_j$  the cost of the duty  $j$ . The binary assignment matrix  $A$

defines the trips  $i$  covered by duty  $j$ . The model formulation is:

$$\min \sum_{j \in N} c_j x_j \quad (1)$$

$$s.t. \sum_{i \in M_r} d_{ig} y_i \geq g_r \sum_{i \in M_r} d_{ig} \quad \forall g \in G_r, \forall r \in R \quad (2)$$

$$\sum_{j \in N} a_{ij} x_j \geq y_i \quad \forall i \in M \quad (3)$$

$$y_i \geq a_{ij} x_j \quad \forall i \in M, \forall j \in N \quad (4)$$

$$x_j, y_i \in \{0, 1\} \quad \forall i \in M, \forall j \in N. \quad (5)$$

The objective function (1) minimizes the total cost of all duties. Constraint (2) ensures that the minimal required total distance of attended trips is satisfied for each attendance rate  $g$  of network  $r$ . Constraint (3) links the trips of a duty to the attendance rate constraint: if a duty covers trip  $i$ , the corresponding distance can add to the attendance rate constraint (2). Finally, constraint (4) ensures that if a trip is covered by at least one of the duties in the optimal solution, the distance of the trip add to the attendance rate fulfillment. In other words, the inequality of (3) allows deadheading, i.e. crew traveling on a train without being assigned to a work task, but only if the trip is already covered by another duty (cf. constraint (4)).

### 3 Two Solution Approaches

Railway crew scheduling problems are known to be very large and complex [5]. This is mainly because of the immense number of trips and an explosion of their combinatorial possibilities. For instance, in our data set, the average trip length is around 18 min. Hence, a cost-efficient duty of approximately 8 h combines on average 10 to 20 trips, in extreme cases up to 40. As a result, different approaches to manage the problem size and to reduce the computational time were presented in recent years. These approaches can be categorized as network-size reduction by trip combination (e.g., [5]) or as network decomposition into smaller sub-problems (e.g., [1, 4]). Additionally, acceleration techniques for solution methods, especially for the popular column generation heuristic, have been investigated extensively [2].

#### 3.1 A Hybrid Column Generation Approach with Genetic Algorithm

Our work is based on a solution method developed by Hoffmann et al. [3]. It consists of a hybrid column generation approach with genetic algorithm (*CGGA*).

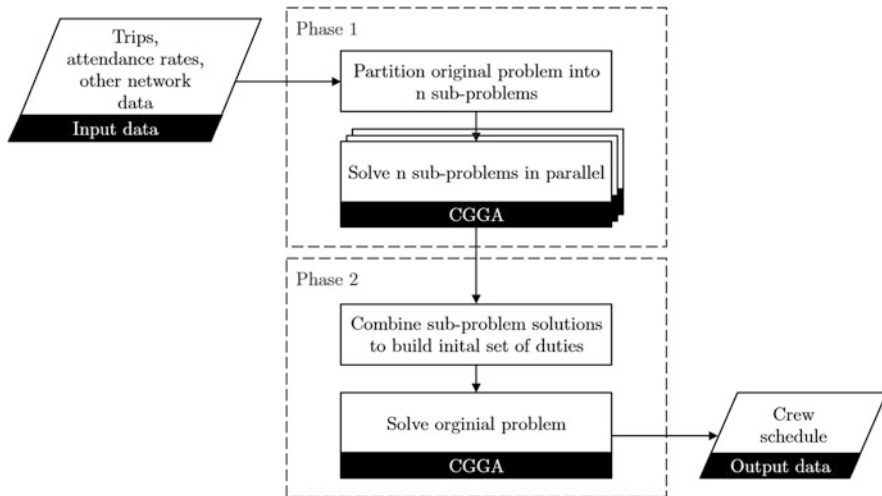
**Table 1** A hybrid column generation approach with genetic algorithm

Step 0	<i>Initialization and initial set of duties.</i> Read input data and generate initial set of columns by combining trips to efficient blocks
Step 1	<i>Computation of dual multipliers.</i> Solve linear relaxed master problem with the current set of columns using a commercial solver and retrieve dual multipliers for trips
Step 2	<i>Generation of columns.</i> Generate columns with negative reduced cost using a genetic algorithm. If new columns with negative reduced cost are generated, go to Step 3 otherwise go to Step 4
Step 3	<i>Deletion of columns.</i> If a column was not part of the optimal solution in Step 1 for a defined number of iterations, delete it and return to Step 1
Step 4	<i>Optimal integer solution.</i> Generate integer solution of master problem using a commercial solver with defined stopping criterion

Acceleration techniques such as efficient trip combinations in the initial set of duties or column deletion are applied. For the sake of brevity, we outline the algorithm in Table 1 and refer the reader to Hoffmann et al. [3] for more details.

### 3.2 A Two-Phase Optimization Method

In order to accelerate computational time while maintaining high quality solutions, we propose a novel two-phase optimization method (*2PH*) based on a partitioning-and-re-combining strategy (see Fig. 1).



**Fig. 1** Framework of the two-phase optimization method (*2PH*)

In the first phase (Phase 1), we decompose the original problem into smaller sub-problems, which are solved in parallel using *CGGA*. Subsequently, in Phase 2, we solve the original problem using *CGGA* building on an initial set of duties, which combines the duties of each sub-problem solution and additional efficient trip combinations (cf. Step 0 of *CGGA*). Starting with the feasible and already good quality solution of Phase 1, *CGGA* further improves the total schedule by creating duties across the sub-problems. The main advantage of this approach is that the generation of separate sub-problems enables parallel computation. The second phase ensures that the original problem is solved close to optimality and thereby all attendance rates requirements per network are fulfilled.

Sub-problems can be generated on the basis of different dimensions, such as network structure, geography and distance, time or historical schedule knowledge, e.g., the likelihood of two trips being combined in a duty [1]. Here, we take advantage of the underlying network structure and define a single network as a sub-problem.

## 4 Computational Experiments

We evaluate both approaches by means of a real data set of 12 interconnected networks. The 1-day problem consists of 6.491 trips and 136 relief points, thereof 28 are potential crew depots. The attendance rate distribution over all networks shows as follows: 79% of kilometers are assigned to an attendance rate of 100%, 3% to 50%, 13% to 20% and 6% to 0%. We manually add taxi trips (with attendance rate 0%) to ensure solution feasibility. All tests are executed on a Intel(R) Xenon(R) CPU E5-2630 with 2.6GHz clock speed (384GB RAM). The linear and integer programming problems are solved using the commercial solver Gurobi, version 7.5. We limit the computational time for the relaxed problem and genetic algorithm to 48h. The integer programming model is terminated when the optimal solution is found (i.e. 0% gap between the best solution found and the current lower bound), at a gap  $\leq 1\%$  after 48h or at the latest after 48 and 120h in  $2PH_{Phase1}$  and in *CGGA* and  $2PH_{Phase2}$ , respectively.

Table 2 shows the average (*avg.*) as well as the minimal (*min.*) and maximal (*max.*) best objective value found (*Obj.*) of 10 test runs per solution approach. We also include the gap of the best integer solution found to the current lower bound (*Gap*), the total computational time (*Time*) and the share of mixed duties (*Mixed duties*), i.e. duties that consists of trips from two or more networks. In addition to the multi-network case, we solve each of the 12 single networks separately using *CGGA*. The objective values of those optimization runs add up to the best solution found of the single network scheduling case, which serves as reference point to evaluate the cost effect of multi-network scheduling ( $\Delta_{Obj}$ ).

As can be seen from Table 2,  $2PH$  achieves a better objective value than *CGGA* within less computational time ( $-53.72$ h on average). A cause for this is the difference in problem size, i.e. the number of decision variables, between

**Table 2** Computational results

	Solution approach		Obj.	Gap [%]	Time [h]	Mixed duties [%]	$\Delta_{Obj}$ [%]
Single networks	CGGA		7.775.820	–	–	–	–
Multiple networks	CGGA	Avg.	8.418.766	8.4	146.06	49.7	8.3
		Min.	8.313.055	8.1	167.98	46.5	6.9
		Max.	8.520.605	9.4	142.01	51.6	9.6
	2PH	Avg.	7.605.110	0.4	92.34	11.5	–2.2
		Min.	7.593.765	0.7	95.05	10.9	–2.3
		Max.	7.629.340	0.6	96.28	9.6	–1.9

*CGGA* and *2PH<sub>Phase2</sub>*, which both solve the original problem. *2PH<sub>Phase2</sub>* builds on a highly efficient and smaller initial set of duties. Therefore, a smaller number of decision variables can be maintained during the solution procedure. Hence, the number of iterations as well as the computational time per iteration is reduced.

Interestingly, the solution generated by *CGGA* contains a high share of mixed duties (49.7% on average). However, considering the best objective value found and the remaining gap, this doesn't lead to a better objective value than *2PH*. Instead, with only 11.5% mixed duties, *2PH* reduces the schedule cost by on average 2.2% in comparison to the single networks case.

## 5 Conclusion and Further Research

We presented a multi-network crew scheduling problem with attendance rates and compared two different approaches to solving the optimization problem. The two-phase optimization method based on a partitioning-and-re-combining strategy showed a better performance in both solution quality and computational time. We also showed that with multi-network crew scheduling personnel cost savings around 2% can be achieved.

Further studies aim to investigate the potential of different types of partitioning algorithms and metrics to further improve the performance of *2PH*. In this context, we address solving the crew scheduling problem of 12 networks for 1 week, the standard time period for crew scheduling of our partner company.

## References

1. Abbink, E.J.W., van't Wout, J., Huisman, D.: Solving large scale crew scheduling problems by using iterative partitioning. In: Liebchen, C., Ahuja, R.K., Mesa, J.A. (eds.) 7th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'07), pp. 96–106. Leibniz Center for Informatics, Wadern (2007). <https://dx.doi.org/10.4230/OASIS.ATMOS.2007.1168>

2. Desaulniers, G., Desrosiers, J., Solomon, M.M.: Accelerating strategies in column generation methods for vehicle routing and crew scheduling problems. In: Ribeiro, C.C., Hansen, P. (eds.) *Essays and Surveys in Metaheuristics*, pp. 309–324. Springer, Boston (2002). [https://dx.doi.org/10.1007/978-1-4615-1507-4\\_14](https://dx.doi.org/10.1007/978-1-4615-1507-4_14)
3. Hoffmann, K., Buscher, U., Neufeld, J.S., Tamke, F.: Solving practical railway crew scheduling problems with attendance rates. *Bus. Inf. Syst. Eng.* **59**, 147–159 (2017). <https://dx.doi.org/10.1007/s12599-017-0470-8>
4. Jütte, S., Thonemann, U.: Divide-and-price: a decomposition algorithm for solving large-scale railway crew scheduling problems. *Eur. J. Oper. Res.* **219**, 214–223 (2012). <https://dx.doi.org/10.1016/j.ejor.2011.12.038>
5. Kwan, R.S.K., Kwan, A.S.K.: Effective search space control for large and/or complex driver scheduling problems. *Ann. Oper. Res.* **155**, 417–435 (2007). <https://dx.doi.org/10.1007/s10479-007-0203-3>