RAILWAY CREW PAIRING OPTIMIZATION

Izabela DZIADUCH, Paweł HANCZAR

Wrocław University of Economics, Department of Logistics, Wrocław, Poland, EU,
izabela.dziaduch@ue.wroc.pl, pawel.hanczar@ue.wroc.pl

Abstract

Crew scheduling is one of the major phases in crew management in large transportation networks such as railway, bus and airline systems. The scheduling of crews is usually considered as two problems: the crew pairing problem and the crew assignment (rostering) problem. These problems are solved sequentially. In this paper we focus on the crew pairing problem in the railway systems. An optimization model for solving train drivers pairing problem arising from a practical application is presented. The model allows to define a number of duties with minimal total duration, covering all segments that the company has to carry out during a predefined time period. In order to evaluate the results from the developed model, they are compared with the results obtained by the planner. The results show that the proposed model is capable of efficiently generating the optimal work schedules for the train drivers in a reasonable computation time.

Keywords: Railway scheduling, crew pairing, passenger transportation, exact model, optimization

1. INTRODUCTION

Before explaining the crew pairing problem a few definitions should be given. In railway scheduling, a segment is a section of the trip from one station to another with specified departure and arrival times. A duty period (or duty) - mostly a working day of a crew - consists of a sequence of segments with short rest periods or sits separating them. A sit connection during a duty period mainly consists of the waiting time of the crew for changing planes on to the next segment. Each duty starts with a sign-in and ends with a sign-out by the crew. A pairing (also called tour of duty, rotation or trip) is a sequence of duties and each pairing begins and ends at the same crew base. Crew base is a city where crews are stationed. In a pairing there is an overnight rest between each duty. To reposition a crew from one base to another base, a pairing might include crews as passengers and this kind of task is called deadhead (or positioning). Generally deadheads are used to transport a crew where they are needed to cover a segment or to return to their home base. **Figure 1** illustrates an example of a two-day crew pairing, showing duty periods, sits within duty periods, overnight rests, and sign-in and sign-out times.

In the crew pairing problem the objective is finding a set of pairings over the planning horizon. While selecting pairings all segments must be covered exactly once and cost of the all selected pairings should be kept in the minimum [1]. Cost structure depends on the individual railway companies. Some railway companies pay a fixed salary for a working hour. Others calculate the salary based on credit hours which are the maximum of the travel time and some guaranteed minimum hours. The cost of a pairing also includes a cost paid to the crew for staying away from their base (e.g., accommodations, transportation), deadheading (when crew are transported on a trip as passengers), and other penalty costs. Also selected pairings should be legal according to the specified work rules and rail traffic regulations.

The crew pairing problem is one of the most widely studied problems in the field of Operations Research. Mathematically, crew pairing problem is usually modeled via the set partitioning problem or the set covering problem with additional constraints [3-7]. The problem is difficult because of the number of constraints and variables; they are large-scale mixed integer problem [8]. The three most common solution methodologies are Lagrangian relaxation [9-12], Benders decomposition [13-15] and branch-and-price [16-19]. Since the 1990s, the most popular approach has been the set covering problem with column generation embedded in branch-and-bound [20-22].

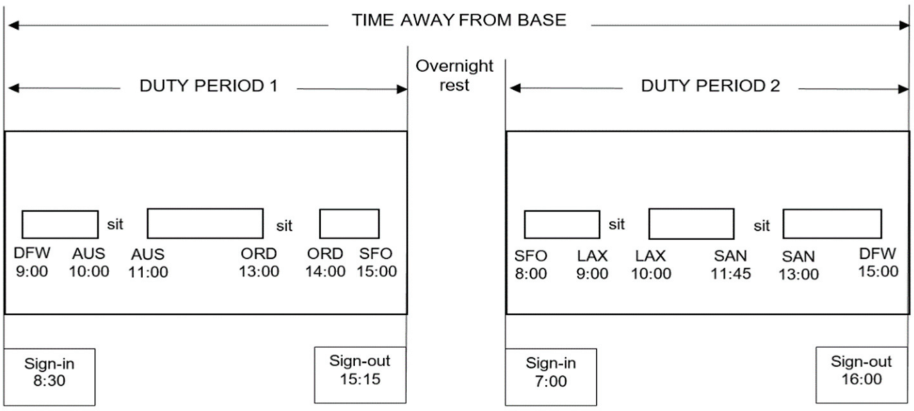

Figure 1 An example of crew pairing with DFW station as crew base [2]

The outline for the rest of the paper is as follows. In the next section, a detailed description of the problem is provided. The decision model, which will allow to construct a work schedule (a set of feasible duties) for train drivers, is defined in Secion 3. In Section 4, the results of our computational experiments are presented. Finally, the last section concludes the paper and gives directions for further research.

2. PROBLEM DESCRIPTION

In this study, we consider the problem of constructing a cyclical work schedule of minimum duration for train drivers spread over several crew bases in a rail network. Cyclical means that each day has the same work schedule. A feasible duty is a duty that satisfies all constraints imposed by labor contracts and union regulations, among others. The following constraints were used to create work schedule for drivers:

- each driver belongs to one of the crew bases. There are 5 crew bases, mainly located at the main nodes of the railway network in Lower Silesia. **Figure 2** presents the rail network.
- Each driver operates from his crew base. That is, each of his duties starts and ends in his crew base.
- Each duty consists of a series of segments to be carried out by a single driver on a single day.
- The duties should be such that each segment is covered by (at least) one driver.
- The minimum length of a duty is 8 hours.
- The maximum length of a duty is 12 hours.
- The duration of duty for train driver includes besides the time of reading the duty plan, the time of driving a train, the time of cleaning and refueling of railbus, also the gaps between the trips, the time necessary to complete the duty (if the duty duration is less than 8 hours), the travel time from the crew base to the start station of the first segment of the duty, the travel time from the end station of the last segment of the duty to the crew base and travel time from the end station of his/her previous trip to the start station of his/her next trip. A driver has to travel as passenger in a train driven by another driver or take a taxi. Assuming that the travel time by car is the same as the average travel time of train.

Efficiency is one of the main criteria for scheduling process of drivers. Efficiency means that the percentage of productive time in the duties is high. Productive time of work includes the times of 4 above mentioned tasks i.e. driving a train from one station to another, reading the duty plan, cleaning and refueling of train, where as the times of all other tasks must be considered as non-productive time.

The input data comes from the daily plan of trains circulation obtained from regional operator of passenger trains. Company runs mostly inside Lower Silesian Province of Poland. Each circuit is served by diesel rail buses. The plan consists of segments that the driver has to perform. Each segment is characterized by a departure station, a departure time, an arrival time, and an arrival station. It should be pointed out that the work schedule for the crew is created manually by employee of the company (qualified person who is responsible for scheduling of crews' work). Manual scheduling is time consuming and error prone. Currently the company doesn't have any decision support system to crew scheduling.

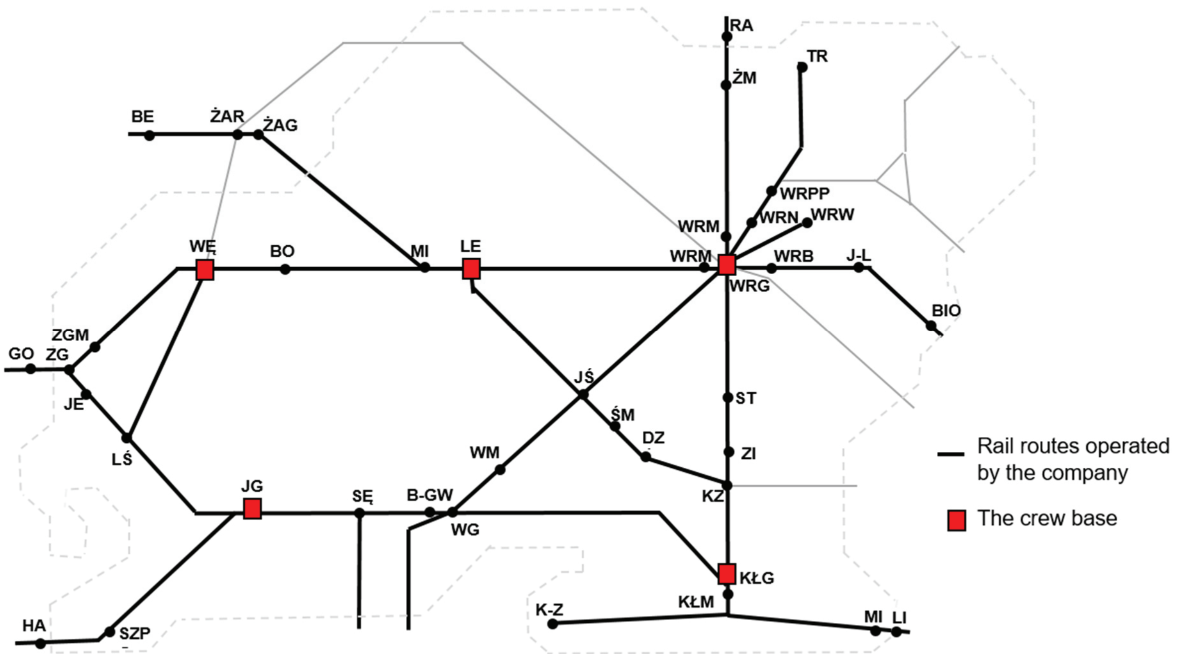

Figure 2 Schematic representation of the geographical railway network used by the regional operator of passenger trains in Lower Silesia in Poland

3. MATHEMATICAL MODEL

In the formulation written with formula (1) to (10), five groups of decision variables were used. The first group contains the decision variables *x*, *y* and *z* which are responsible for ensuring consistency of the routes obtained in the solution. The variable $x_{r1,r2,s}$ is a binary variable and is equal to the value of 1 if and only if segment r_2 is served directly after segment r_1 by duty *s*. Variables $y_{r,s}$ and $z_{r,s}$ are equal to 1 if the segment *r* is served as the first in a duty *s* (variable *y*) or is served as the last in a duty *s* (variable *z*). The next group of variables consists of variables $q_{r,s}$ and w_s, which are important for the times of the duties. The variable $q_{r,s}$ is equal to the duration of the duty in minutes from the beginning of the duty until the segment *r* is realized. On the other hand variable w_s is equal to the largest value of $q_{s,r}$, so it represents the duration of the duty *s*.

The objective function expressed by the formula (1) causes that the solution will have the shortest length of duties - it means sum of durations of all duties expressed in the decision variable w_s:

$$\sum_{s \in S} w_s \rightarrow min \qquad (1)$$

Constraints expressed by formula (2) and formula (3) are used to ensure that each segment will be served, i.e. it is served after another segment or is served first -formula (2), and is followed by another segment or is handled last - formula (3):

$$\sum_{s \in S} y_{r,s} + \sum_{s \in S, r1 \in R} x_{r1,r,s} \geq 1 \text{ for every } r \in R \qquad (2)$$

$$\sum_{s \in S} z_{r,s} + \sum_{s \in S, r1 \in R} x_{r,r1,s} \geq 1 \text{ for every } r \in R \qquad (3)$$

The constraint expressed by formula (4) initializes the value of the $q_{r,s}$ variable. If the duty starts from the segment r (that is variable $y_{r,s}$ is equal to 1) then the value of variable $q_{r,s}$ after handling the segment r must be equal to the length of the duration of the segment r:

$$q_{r,s} \geq y_{r,s} \cdot \left(c_r^k - c_r^p\right) \text{ for every } r \in R,\ s \in S \tag{4}$$

where:

c_r^k - end time of segment r,

c_r^p - end time of segment r.

The constraint expressed by formula (5) ensures the continuity of one duty in the time. If the value of the variable $x_{r1,r2,s}$ is 1 then this constraint is active and causes that q_{r2},s is equal to the value of $q_{r1,s}$ increased by the duration of segment r_1 and the interval between segment r_1 and r_2. In addition, this restriction takes into account the situation when the segment begins one day before it ends:

$$q_{r,s} + y_{r,s} \cdot \left(c_{r'}^k - c_{r'}^p\right) + \begin{cases} c_{r'}^p - c_{r'}^k & if \quad c_{r'}^p \geq c_{r'}^k \\ 24 - \left(c_{r'}^k - c_{r'}^p\right) & if \quad c_{r'}^p \geq c_{r'}^k \end{cases} - 24 \cdot (1 - x_{r,r',s}) \leq q_{r',s} \text{ for every } r \in R,\ r' \in R \tag{5}$$

The constraint expressed by formula (6) is used to combine the value of variable w_s with decision variables $q_{r,s}$. The value of variable w_s is equal to the largest of all values of $q_{r,s}$ for a given duty:

$$q_{r,s} + \sum_{r' \in R} y_{r',s} e_{g_s,r'}^p + \sum_{r' \in R} z_{r',s} e_{g_s,r'}^k \leq w_s \text{ for every } s \in S,\ r \in R \tag{6}$$

where:

g_s - depot of duty s,

$e_{g_s r}^p$ - travel time between g_s and start station of segment r,

$e_{g_s r}^k$ - travel time between g_s and end station of segment r.

Two further constraints expressed by formula (7) and formula (8) ensure that the duration of the duty will not be less than 8 hours and longer than 12 decreased by 20 minutes, when the employee has to read the duty plan:

$$q_s \leq 11.67 \cdot \sum_{r \in R} y_{r,s} \text{ for every } s \in S \tag{7}$$

$$q_s \geq 7.67 \cdot \sum_{r \in R} y_{r,s} \text{ for every } s \in S \tag{8}$$

The last group of constraints is related to the consistency of the segments. Constraints expressed by formula (9) and formula (10) ensure that the duty used to serve segments must have a beginning ($y_{r,s}$=1) and end ($z_{r,s}$ equal to 1):

$$\sum_{r \in R, r' \in R} x_{r',r,s} \leq M \sum_{r' \in R} y_{r',s} \text{ for every } s \in S \tag{9}$$

$$\sum_{r \in R, r' \in R} x_{r',r,s} \leq M \sum_{r' \in R} z_{r',s} \text{ for every } s \in S \tag{10}$$

where:

M - very large number.

Then constraints expressed by formula (11) and formula (12) ensure that the duty can not start and end more than one segment:

$$\sum_{r \in R} y_{r,s} \leq 1 \text{ for every } s \in S \quad (11)$$

$$\sum_{r \in R} z_{r,s} \leq 1 \text{ for every } s \in S \quad (12)$$

Similarly, restrictions expressed by formula (13) and formula (14) ensure that the duty will not serve ($x_{r1,r2,s}$=1) more than one segment:

$$\sum_{r' \in R} x_{r',r,s} \leq 1 \text{ for every } s \in S,\ r \in R \quad (13)$$

$$\sum_{r' \in R} x_{r,r',s} \leq 1 \text{ for every } s \in S,\ r' \in R \quad (14)$$

The last limitation expressed by formula (15) causes that the sum of segments started by duties is equal to the sum of completed segments:

$$\sum_{r' \in R} x_{r',r,s} + y_{r,s} = \sum_{r' \in R} x_{r,r',s} + z_{r,s} \text{ for every } s \in S,\ r \in R \quad (15)$$

4. COMPUTATIONAL RESULTS

Computational experiments were carried out using instances based on the daily trains circulation plan of the regional operator of passenger trains in Lower Silesia in Poland to compare the results obtained by the proposed model proposed with the results received by planner (qualified person who is responsible for scheduling of crews' work in the company). The plan of train circulation included 7 circuits, which included 82 segments in total. Each circiut consisted of a different number of segments (from 6 to 15).

In total, 6 instances were checked, each of which consisted of 3 circuits. In the first instance, cycles 1, 2 and 3, containing 27 segments, were considered, in the second one, the cycles 4, 5 and 6, containing 40 segments, were analyzed, in the third one there were 43 segments making up the 5, 6 and 7 cycles, in the fourth one the circuits 1, 2 and 5, containing 31 segments, were considered, in the fifth one there were 29 segments making up the 1, 2 and 6 cycles, while, in the sixth one, analysis of circuits 1, 2 and 7, containing 31 segments were analyzed.

The model proposed in Section 3 was solved by using the AIMMS (Advanced Integrated Multidimensional Modeling Software) optimization package (version 12.7.1, CPLEX). All test runs are made on personal computer with an Intel® Core™ i3-8130U Processor. The obtained results are presented in **Table 1**.

The results presented in **Table 1** were compared with the results of computational experiments performed by the planner in order to determine the efficiency of drivers' working time. **Table 2** shows the results of computational experiments performed by the planner.

The study shows that the duration of duties, according to the optimizer, is shorter than the times in the plans of the planner. The average duration of duties according to the optimizer is 63.94 hr, while, according to planner plan, it equals average 67.78 hr. The results indicate, on average, an approx. 5.7 % improvement in the use of drivers' working time.

It is worth noting, that the duration of effective work during the duty is shorter on average by 0.17 hr in schedules performed using optimizer, than by the planner. This time depends mainly on the number of duties.

The analysis of the data, presented in the **Table 1** and **Table 2**, demonstrates, that the planner developed 5 drivers' working time schedules for 7 duties, and 1 schedule for 8 duties, while the optimizer, that was used to solve the six test instances, indicated, that it is possible to develop 4 drivers' working time schedules for 7 duties, and 2 schedules for 6 duties.

Table 1 Computational results of 5 instances solved by AIMMS.

Number of test instances	1	2	3	4	5	6
Number of duties	7	7	7	7	6	6
Duration of duties (hr)	63.91	64.82	65.03	67.45	61.41	61.01
Deadheads (hr)	9.61	6.91	8.13	8.07	9.74	9.63
Sits within duty periods (hr)	6.82	10.65	12.75	8.45	7.52	10.63
Waiting time for tasks (hr)	3.65	6.23	5.08	8.50	2.07	1.80
Time of effective work	43.83	41.03	39.07	42.43	42.08	38.95
Solving time (min)	4.5	5.2	3.8	3.2	2.2	0,5

Table 2 Computational results of 5 instances obtained by the planner.

Number of test instances	1	2	3	4	5	6
Number of duties	7	7	7	8	7	7
Duration of duties (hr)	66.10	65.27	67.36	74.65	66.59	66.72
Deadheads (hr)	9.63	7.48	8.99	13.89	10.29	9.86
Sits within duty periods (hr)	6.97	11.83	14.37	7.62	8.45	12.15
Waiting time for tasks (hr)	5.67	4.93	4.93	10.37	5.43	5.43
Time of effective work	43.83	41.03	39.07	42.77	42.42	39.28
Solving time (min)	20.2	22.1	21	18.5	18.0	19.1

The results from response optimizer indicated that, the length of time of trip between the crew base and the start or end station of the segment, was shorter on average by 1.34 hr, than the time in the planner schedules.

In addition, that sits within duties is shorter on average by 0.76 hr in schedules prepared using optimizer, than by the planner.

Moreover, according to data obtained from the optimizer, the mean time of waiting for tasks is 4.56 hr and there is shorter by 1.57 hr compared to the time obtained by the planner.

The three parameters mentioned above i.e. deadheads, sits within duty periods and waiting time for tasks are important from the point of view of the transport company, because it is not connected with productive work. The results from response optimizer indicated that, the average percentage of the non-productive work time in the total driver work time is 35.46 % per duty, while the planner has developed schedules, in which the average percentage of the non-productive work time in the total driver work time is 38.83 % per duty.

The data also demonstrates, that the optimizer needs less time to designate the solution than the planner does. The optimizer presented solutions on average after 3.23 minutes, while the planner needed on average 19.82 minutes to perform driver working time schedule.

5. CONCLUSION

In this paper, a railway crew scheduling problem was presented. The constraints, which were needed to construct a work schedule of minimum duration for train drivers spread over several crew bases in a rail network operated by the regional company of passenger trains in Lower Silesia in Poland, were described. Next, the model to find a solution of the problem was formulated. Finally, the results from the model were compared with the results obtained by the planner. For each instance solved by using the model, shorter durations of train drivers' duties were achieved. The results indicated, on average, an approx. 5.7 % improvement in the use of drivers' working time, which would reduce personnel costs in the company. In addition, the optimizer needed less time to plan the solution than the planner - on average, 16.59 minutes.

Unfortunately, at the current stage of the work, it should be noted that limiting the number of segments to 50, quite significantly reduces the possibilities of using the proposed model. It should also be noted, that this model has been used as a exact model, while its implementation in practice means the necessity of frequent updating of parameters and at least several solving during the day. Taking into account the above limitations, further research will aim to develop a heuristic model, that will allow to determine a work schedule (a set of feasible duties) for more than 50 segments.

REFERENCES

[1] ÖZDEMIR, U. *Methodology for crew-pairing problem in airline crew scheduling* [online]. Master Theses, Bogazici Univesity, 2009. [viewed 2017-11-10]. Available from https://www.cmpe.boun.edu.tr/~gungort/theses/Methodology%20for%20Crew-Pairing%20Problem%20in%20Airline%20Crew%20Scheduling.pdf

[2] XUGANG, Y. *Airlines' Crew Pairing Optimization: A Brief Review* [online]. Technical Report, Department of Applied Mathematics and Statistics, Johns Hopkins University, 2007. [viewed 2017-11-12]. Available from http://www.cis.jhu.edu/~xye/papers_and_ppts/ppts/airline_crewpairing_copy1.pdf

[3] FRELING, R., LENTINK, R. M. and WAGELMANS, A. P. M. A Decision Support System for Crew Planning in Passenger Transportation Using a Flexible Branch-and-Price Algorithm. *Annals of Operations Research*. 2004. Vol. 127, iss. 1-4, pp. 203-222.

[4] WREN, A., FORES, S., KWAN, A., KWAN, R., PARKER, M. and PROLL, L. A flexible system for scheduling drivers. *Journal of Scheduling*. 2003. Vol. 6, iss. 5, pp. 437-455.

[5] SUYABATMAZ, A. C. and SAHIN, G. A. Column-and-Row Generation Algorithm for a Crew Planning Problem in Railways. In: KLATTE, D., LUTHI, H.-J. and SCHMEDDERS, K. *Operation Research Proceedings 2011*. Berlin, Heidelberg: Springer, 2012, pp. 335-340.

[6] DESAULNIERS, G., DESROSIERS, J., DUMAS, Y., MARC, S., RIOUX, B., SALOMON, M. M. and SOUMIS, F. Crew pairing at Air France. *European Journal of Operational Research*. 1997. Vol. 97, iss. 2, pp. 245-259.

[7] PEZZELA, F. and FAGGIOLI, E. Solving large set covering problems for crew scheduling. *Top*. 1997. Vol. 5, iss. 1, pp. 41-59.

[8] KASIRZADEH, A., SADDOUNE, M. and SOUMIS, F. Airline crew scheduling: models, algorithms, and data sets. *EURO Journal on Transportation and Logistics*. 2017. Vol. 6, iss. 2, pp. 111-137.

[9] GEOFFRION, A.M. Lagrangian relaxation for integer programming. *Mathematical Programming Study*. 1974. No. 2, pp. 82-114.

[10] FISHER, M. L. The Lagrangian relaxation method for solving integer programming problems. *Management Science*. 2004. Vol. 50, no. 12, pp. 1861-1871.

[11] HEARN, D. W., AND LAWPHONGPANICH, S. Lagrangian dual ascent by generalised linear programming. *Operations Research Letters*. 1989. Vol. 8, pp. 189-196.

[12] KIPP, M. R. *Large Scale Linear and Integer Optimization: A Unified Approach*. 3rd ed. Boston/Dordrecht/London: Kulwer Academic Publishers, 2004.

[13] BENDERS, J. F. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*. 1962. No. 4, pp. 238-252.

[14] MINOUX, M. *Mathematical programming: theory and algorithms*. Chichester: John Wiley & Sons, 1986.

[15] CORDEAU, J.-F., STOJKOVIC, G., SOUMIS, F. and DEROSIERS, J. Benders Decomposition for Simultaneous Aircraft Routing and Crew Scheduling. *Transportation Science*. 2001. Vol. 35, no. 4, pp. 375-388.

[16] SOYKAN, B. and EROL, S. A Branch-and-Price Algorithm for the Robust Airline Crew Pairing Problem. *The Journal of Defense Sciences*. 2014. Vol. 13, iss. 1, pp. 37-74.

[17] DESAULNIERS, G., DESROSIERS, J., IOACHIM, I., SOLOMON, M. M., SOUMIS, F. and VILLENEUVE, D. A unified framework for deterministic time constrained vehicle routing and crew scheduling problems. In: CRAINIC, T.G. and LAPORTE, G., eds. *Fleet Management and Logistics*. New York: Springer, 1998, pp. 57-93.

[18] BARNHART, C., JOHNSON, E. L., NEMHAUSER, G. L., SAVELSBERGH, M. W. P. and VANCE, P. H. Branch-and-price: Column generation for solving huge integer programs. *Operations Research*. 1998. Vol. 46, no. 3, pp. 316-329.

[19] VANCE, P., ATAMTURK, A., BARNHART, C., GELMAN, E., JOHNSON, E., KRISHNA, A., MAHIDHARA, D., NEMHAUSER, G. and REBELLO, R. A heuristic branch-and-price approach for the airline crew pairing problem. *Technical Report LEC-97-06*. Georgia Institute of Technology, 1997.

[20] DESROSIERS, J. and LUBBECKE, M.E. A primer in column generation. In: DESAULNIERS, G., DESROSIERS, J. and SOLOMON M. M., eds. *Column generation*. US: Springer, 2005.

[21] DESROSIERS, J., DUMAS, Y., SOLOMON, M. M. and SOUMIS, F. Time constrained routing and scheduling. In: BALL, M. O., MAGNANTI, T. L., MONMA, C. L. and NEMHAUSER, G. L., eds. *Network Routing, Handbooks in Operations Research and Management Science*. North-Holland, Amsterdam: Elsevier, 1995, chapter 2, pp. 35-139.

[22] ANBIL, R., FORREST, J. J. and PULLEYBLANK, W. R. Column Generation and the Airline Crew Pairing Problem. *Documenta Mathematica*. 1998. Vol. 3, pp. 677-686.