Aalto University
School of Science
Degree programme in Engineering Physics and Mathematics

Optimization of Duties in Railway Traffic

Barchelor’s thesis
6.5.2014

Teemu Kinnunen

Aalto University

AALTO UNIVERSITY
SCHOOL OF SCIENCE
PO Box 11000, FI-00076 AALTO
http://www.aalto.fi

ABSTRACT OF THE BACHELOR’S THESIS

Author: Teemu Kinnunen

Title: Optimization of Duties in Railway Traffic

Degree programme: Degree Programme in Engineering Physics and Mathematics

Major subject: Systems Sciences | Major subject code: F3010

Supervisor: Professor Harri Ehtamo

Instructor: M.Sc. (Tech.) Mikko Alanko

Abstract:

This thesis discusses crew scheduling in railway transportation. A case of Finnish railway operator is used as an example. Literature review discusses crew scheduling in other countries. Finally, two crew scheduling problems are solved.

The objective of crew scheduling is to form duties from all the tasks, which need to be performed. There are multiple constrains related to the duties, which make the crew scheduling a complex problem. Also, the problem sizes are usually really large. Traditionally duties were constructed by hand, but the rising demands of the modern world have forced railway operators to acquire planning software.

In literature the crew scheduling problem is often modeled as a set covering problem with additional constraints. Objective is to minimize the total cost so that every task is covered at least once. Usually solution methods rely on column generation and different kinds of heuristics, because set covering problem is NP-hard. The problem is often solved in two phases, duty generation and duty selection.

Two crew scheduling problems, based on real railway data from the Finnish operator, are solved using the planning software of the operator. The effect of different optimization parameters on the final solution is tested. The objective is to find solutions that minimize the total working hours and the amount of duties, while the amount of tasks, which are used multiple times, is as low as possible. Some clearly bad values for the parameters are found, but there is not a single parameter combination that would produce best results in every case.

Date: 6.5.2014 | Language: English | Number of pages: 17 + 4

Keywords: Crew scheduling, railway transportation, planning software

Contents

1. Introduction

The problems related to railway transportation are large and complex. Railway transportation planning includes multiple subdivisions. Caprara et al. (2007) divide the planning in six operational areas: line planning, train timetabling, train platforming, rolling stock circulation, train unit shunting, and crew planning. In this thesis the focus is on crew planning, specifically for drivers, and from now on planning refers to the planning related to drivers.

Before starting the crew planning, the railway operator must assemble trains according to customer demand, make timetables for the trains and assign locomotives for each train (when needed). Thereafter every task needed to cover all the timetabled trains is generated. Crew planning process consists of two phases. The first phase is crew scheduling, in which individual *tasks* are coupled to form *duties*. A task is usually a driving trip between two stations, or a *positioning trip.* Positioning trip means travelling between stations by other means than driving a train, preferably travelling as a passenger in a train. If suitable trains are not available, alternative options are using a car, a bus or a taxi. Other task types include *walking* within a station in order to change train and *shunting*, for example coupling or uncoupling train units. A duty includes every task that a driver must complete during his/her day of work. Usually a duty originates and terminates at the driver's *home base*. After duties are generated, the second phase is to form *rosters* from them. Rosters of a certain personnel base include all the duties that must be performed by the personnel of that base and are formed by clustering duties.

There are multiple constraints, which must be satisfied when generating the duties, related to working time, breaks, driving time and starting and ending a duty. This makes planning of the duties difficult. The planning personnel faces even greater challenge, when there is a disruption in a part of railway network. Alternate routes/means for travelling must be applied fast for customers, and the duties must be altered. It was a long tradition in the railway industry, that experienced planning personnel constructed the duties by hand, which was very time consuming and left little or no room for analyzing and optimizing the duties. The quality of the duties relied heavily on the experience of planners. Nowadays there is a growing tendency that everything must be made more efficient. Also, at the same time, service for passengers and quality of work of train crews must have been improved. Support of computing power and optimization algorithms is needed to meet the ever rising efficiency, punctuality and crew satisfaction targets. In recent years many railway companies have adopted computer aided means to improve the results of crew scheduling. This has provided more efficient duties, improved the perceived quality of work for drivers and also improved punctuality of trains (Kroon et al. 2008).

This thesis provides an overview of crew scheduling. A Finnish railway operator is used as an example case to demonstrate the crew scheduling process in practice. Solutions used in other countries are described in literature review. The planning software of the Finnish operator is used to solve two crew scheduling problems, which use real world data. The remainder of this thesis is organized as follows: In chapter 2 is a description of crew scheduling process. The chapter includes also rules related to crew scheduling and presents a mathematical formulation for the problem. Chapter 3 contains a literature overview of crew scheduling. In chapter 4 two example cases of crew scheduling are solved and effect of altering different optimization parameters is tested. The objective is to find parameter values, which produce solutions with least total hours and duties. Chapter 5 summarizes the planning process and results from the test cases.

2. Crew scheduling

2.1. Crew scheduling in Finland

Railway crew scheduling is a larger and more complex problem than corresponding problems in airline or bus industry (Kroon et al. 2008). In addition to driving the train from one station to another, train drivers must perform the needed shunting work including coupling and uncoupling of carriages, moving carriages/engines to different platforms, driving an engine to be refueled, driving a train to be washed and so on. Duties usually include multiple tasks and multiple changes of rolling stock. Duties also often include positioning trips, i.e. travelling between stations in order to reach the starting point of next driving task. In the other transportation industries shunting work and transferring between vehicles is much less common, because the composition of vehicle will not change during its route and drivers rarely travel as a passenger for a part of a duty.

Crew scheduling process of a Finnish railway operator is used here as an example case. The fundamental objective of crew scheduling is to form duties from all the tasks, which need to be performed by drivers. First of the two phases of crew scheduling is to gather all of the base data which is needed for the planning software used by the operator. Base data is imported from multiple sources and includes for example train timetables, platform information and amount of available workforce. Main goals of the first phase are making the base data error free and gathering the data in time according to the schedule of planning process.

Second phase of crew scheduling is constructing duties. The most important issues that must be concerned while generating the duties include economical and efficiency targets, quality of work and various constraints related to duties and schedule of planning process. Duties are made as economical and acceptable as possible. Acceptability and economy are increased by avoiding resting periods outside home base and by giving more importance to the quality of overall solution than single duties. Acceptability is further increased by grouping tasks so that there are enough breaks, placing needed resting periods at night and avoiding positioning trips. Economy is further improved by increasing the efficiency of duties (effective working time divided by duty length) and minimizing salary supplements. Additional targets for crew scheduling include splitting the workload according to resources of different personnel bases, and when possible evening out working hours, night hours and weekend hours between bases. Duties which can be repeated on multiple days of the week and can be used for a long time are preferred.

2.2. Duty Constraints

To give an idea of the complexity of the crew scheduling problem, the following list presents some of the constraints related to constructing long distance duties for train drivers in Finland. Some rules and details are left out.

- Minimum length of a duty is 4 hours.
- Maximum length of a duty is 10 hours, when the duty does not have rest outside home base.
- Maximum length of a duty, which starts or ends outside home base, is 12 hours.
- Maximum length of a duty, which includes rest (at least 2 hours without work) outside home base is 15 hours, including resting time.

- When duty length is over 15 hours (must be under 24) it is separated into two duties. In this case the duties must be separated by resting period outside home base, which must last at least 6 hours.
- If continuous working time in a duty is over 7 hours, the duty must include a meal break of at least 30 minutes. When the break is outside home base, it must last at least 60 minutes.
- Driving time is calculated according to the timetabled departure and arrival times of trains. Stops that are under 12 minutes are included in driving time. Shunting work, traveling “towards work” and traveling between 22.00-6.00 o’clock are included in driving time.
- A break of at least 12 minutes is counted as relieving time.
- Relieving time must be placed between continuous driving time sections, and it must be at least 10 % of the calculated driving time.
- Minimum of relieving time in a duty is 20 % of duty length.
- In a duty, that does not include work between 02.00-05.00, maximum continuous driving time is 3.5 hours. This can be extended to 5 hours using relieving periods.
- For a duty, that includes work between 02.00-05.00, maximum length of duty is 9 hours. Maximum of continuous driving time is 3.5 hours and maximum driving time in total is 7 hours (using rests).

In regional traffic most of the constraints are the same, biggest difference is that minimum duty length is 7 hours.

2.3. Crew Scheduling Problem Formulation

Crew scheduling problem can be presented using following notation:

$i = [1,2, \dots, D]$ is the list of generated potential duties.

$j = [1, 2, \dots, T]$ is the list of tasks that need to be covered.

c_i are the costs of duties.

x_i are binary decision variables that indicate whether duty i is selected on the final solution $(x_i = 1)$ or not $(x_i = 0)$.

a_{ji} are binary indices that indicate whether task j is included in duty i $\left(a_{ji} = 1\right)$ or not $\left(a_{ji} = 0\right)$.

The crew scheduling problem can be formulated as follows (based on Fischetti et al. (2004)).

$$\min_{x} \sum_{i=1}^{D} c_i x_i \qquad (1)$$

subject to:

$$x_i \in \{0, 1\}, \qquad \forall\, i \in [1, 2, \dots, D] \qquad (2)$$

$$\sum_{i=1}^{D} a_{ji} x_i \ \geq 1, \qquad \forall\, j \in [1, 2, \dots, T] \qquad (3)$$

Equation (1) is the objective function with goal of minimizing total cost of the solution. Constraint (2) assigns binary values for the decision variables. Constraints (3) state, that each task must be included in at least one duty that is part of the final solution. Crew base level constraints can also be included in the formulation. Those can include for example minimum and maximum amounts for number of duties, total working hours and night hours for each crew base. Further discussion about crew base level constraints can be found in Fischetti et al. (2004). This formulation can be used to solve the crew scheduling problem, after a set of potential duties is generated. The solving is done by selecting an optimal subset of duties from the set of potential duties. Duty level constraints are not included in the formulation, because those are applied, when the set of potential duties is generated.

2.4. Rosters

Planners form rosters after the crew scheduling process. Rosters can be presented in matrix form, where each row is assigned to a single driver and each column represents a certain day of week or planning period. When a driver has done the work of certain row, next week/period he does the duties of the next row and so on. This circulation of drivers in a roster ensures that the work does not get monotonous and improves the perceived quality of work. There are multiple constraints related to making rosters from the duties included in the final solution, for example minimum time between duties, maximum amount of work and night work in a period and maximum amount of work before weekly resting period. Further discussion about forming rosters can be found in Hartog et al. (2009) and Caprara et al. (2007).

3. Literature review

In past two decades the railway transportation planning process has received considerable attention in the literature. A recent review of railway operations is provided by Narayanaswami and Rangaraj (2011). Caprara et al. (2007) give an overview of railway operations and provide optimization methods related to every operation. Huisman et al. (2005) describe the problems arising in different phases of planning and solution techniques related to each phase.

In literature the crew scheduling problem is often modeled as a set covering problem with additional constrains. Usually solution methods rely on column generation and different kinds of heuristics, because set covering problem is NP-hard (Kwan 2011). In addition problem sizes in crew scheduling are very large, which makes generating all possible solutions and finding an exact optimum impossible. Column generation uses a feasible solution as a starting point and generates new duties that can improve the solution. Constraints related to individual duties are applied in duty generation phase.

After each generation phase a subset of the generated duties is selected. The objective function is minimizing the cost of all duties. Decision variables are binary and indicate if a duty is selected in the solution or not. Each variable is multiplied with cost of the corresponding duty. Constraints ensure that every task is covered at least once. Additional constraints can for example balance the work load between crew bases and restrict the amount of duties containing night work. This type of formulation is used in multiple publications, see for example Jütte et al. (2011), Abbink et al. (2010), Fischetti et al. (2004) and Kwan (2011).

The Netherlands has been a focus country for the research related to railway transportation planning in the recent years. Before the year 2000 crew scheduling was done manually at NS (Nederlandse Spoorwegen, the largest railway operator of Netherlands), relying on experience and craftsmanship of the planners. Fischetti et al. (2004) provide a review of the implementation of computer aided means, the TURN1-system, to help the crew scheduling process. Also a new production model called “Sharing Sweet&Sour” was created. The model aims at sharing the attractive "sweet" work (driving long-distance trains, work with variation in routes and train series) and unattractive "sour" work (driving regional trains, driving old rolling stock, work in lines that have a lot of passenger aggression) equally between crew bases. TURN1 is based on a set covering problem with additional constraints and uses methods of operations research to solve it. Airline industry had used these types of models before, but the problems in found railway industry are a magnitude larger.

Kroon et al. (2008) describe how implementing operations research methods made a substantial improvement in the Dutch timetable. Crew scheduling problem at NS is also discussed. For NS, the three main objectives of Crew Scheduling are efficiency, acceptability and robustness. Efficiency means minimizing the total cost of duties. Acceptability includes the need of variation in duties and the specialty of NS, sharing sweet and sour work. Robustness means that the duties must be made so that propagation of delays is prevented, for example by adjusting transfer times of crews between trains. NS has over 6000 crew members and operates about 5500 trains daily. On average drivers of NS must complete roughly 1000 duties per day, which include 15 000 trips.

Crew schedules are generated using TURN1-system, which models the crew scheduling problem as a set covering problem. Using the set covering model includes two phases: generating a set of feasible duties and selecting a subset of these. Because of the complexity of the problem and its constraints, TURN1 uses column generation to generate potential duties on the fly whenever needed. The resulting extended set covering problem is solved using Lagrangean relaxation, subgradient optimization and multiple heuristics. The system makes sure, that in the solution each timetabled trip is covered by at least one duty and all crew base level constraints are satisfied, simultaneously minimizing the total cost of selected duties. In addition to duty-related constraints which are taken into account when duties are generated, the set of selected duties must satisfy multiple crew depot level constraints. Some of these include number of duties per crew base, average length of duties and fair division of the sweet and sour work. Implementing the TURN1 had multiple benefits for NS. Number of passengers increased, punctuality reached a record high, millions of Euros were saved and increase of railway transport was achieved without significant investment in infrastructure.

Abbink et al. (2007) present multiple methods to partition the large weekly crew scheduling problems of NS. Overcovering tasks is discussed; sometimes it might be beneficial to allow it. In overcovered task there is more than one driver, for one it is a driving task and for others positioning trip. During the first years of using TURN1, partitioning the weekly problem to sub-problems per weekday was the only partitioning method used. Three additional models for partitioning are presented. Geographical partitioning splits the country to equally sized regions, each of which contains a cluster of crew bases. Cluster sizes of 3 and 7 crew bases are used. Line based partitioning splits the problem in 4 parts according to 4 most important long distance lines. In partitioning based on column information the duties generated by TURN1 are analyzed as pairs and scores are given for each pair. Then the scores are graphed and the problem is partitioned using a genetic algorithm. Best results were achieved by using all partitioning methods one after another. The methods were implemented in 2007 and NS achieved a 2 % efficiency improvement.

Abbink et al. (2010) further discuss methods of solving large scale crew scheduling problems. A new algorithm called LUCIA is presented. It solves the weekly problems without splitting, using Lagrangian heuristics, column generation and fixing techniques. With the new algorithm efficiency improvement of 1 % was achieved in test cases compared to the previous methods, which splitted the problem (see Abbink et al. 2007).

Jütte et al. (2011) describe the crew scheduling problem of DB Schenker, a German freight railway operator. The difference between freight and passenger train crew scheduling is that some of the freight trains are added to schedule on short notice. In the case of DB Schenker, amount of these last-minute trains is about 20 percent. The freight trains are also often operated at night, so splitting the problem into daily instances cannot be done. At DB Schenker crew scheduling was performed manually until 2006. The country was segmented geographically and duties were formed relying heavily on the experience of planners. Thereafter a crew scheduling application was implemented, which helped to achieve savings in operations and planning process costs.

DB Schenker had 3 goals when designing the application for crew scheduling: feasibility, performance and applicability. Feasibility meant that the schedules had to incorporate all the related constraints. Performance meant that the application should outperform manual schedules in the solution speed and quality. Applicability meant that the software had to be easy to use and easy to integrate to current IT systems. The crew scheduling software uses shortest-path algorithm when generating duties. Dynamic column generation and fixing techniques are applied to achieve optimal solution. In each fixing step duties are scored and the ones with the highest scores are fixed. After that the problem is reduced by the tasks included in fixed duties and solved again.

Jütte and Thonemann (2012) present a column generation based decomposition algorithm, which is called divide-and-price. The new algorithm allows overlapping of the regions unlike usual decomposition algorithms, which reduces the loss in solution quality caused by decomposition. An example of the benefits is presented in figure 1.

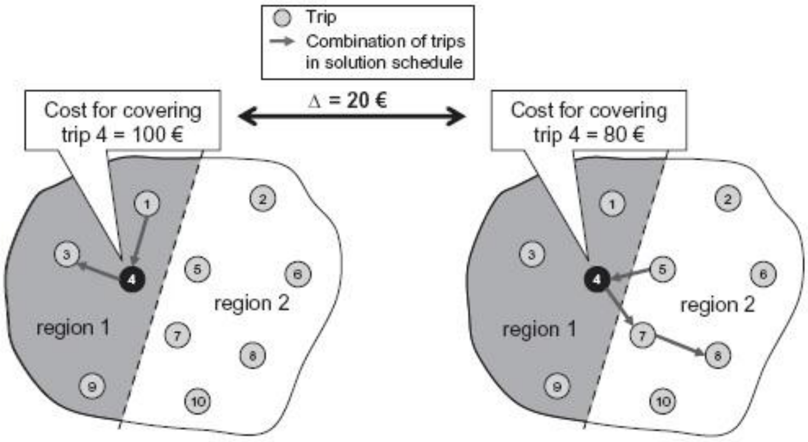

Figure 1: An example of the usage of divide-and-price algorithm (Jütte and Thonemann 2012).

Execution of divide-and-price is fast and it gets close to optimal solutions, so it can be used in short term crew scheduling. Decompostion is really beneficial in reducing solution times. In an example case of 28 000 trips, solution time is reduced to only 8 % of the original problem, when the problem is split to 4 instances with 7000 trips. A primary region is given to each trip and also possibly one or more secondary regions. Secondary regions can be considered when a trip is close to the border(s) of other region(s). A trip will be assigned to the secondary region, if the cost of covering it in that region is lower than in its primary region. In addition to geographical decomposition, chronological decomposition according to trip departure times is tested. Chronologial decomposition with 4 regions is the best compromise for weekly problems and Geographical decomposition performed better on daily problems.

Kwan (2011) presents examples of implementing an automatic crew scheduling system TrainTRACS in UK. The UK rail industry is privatized, which has forced railway operators to optimize their operations. Train crews account for 20-25% of total operating costs, so using automatic crew scheduling can generate substantial savings. TrainTRACS has duty generation and duty selection phases. Constraints of duty and schedule level are integrated in the process. The system uses set covering and ILP techniques in the duty selection phase. First a relaxed LP is solved using the revised simplex method and column generation. After that an integer solution is searched using branch-and-bound method.

Kwan (2011) distinguishes between three example cases of the benefits of automatic crew scheduling for railway operators. The first one is optimizing capability: Virgin West Coast managed to increase amount of weekly trains by 32 %, while amount of crew needed increased by only 0.37 %. Second is solution operability: Southern Railway adopted TrainTRACS in 2006 and the system was able to make fully operable solutions, although the network was one of the largest and most difficult in the UK. Third example is incremental accumulation of domain knowledge: as operators face new problems, the solutions are implemented as a part of the TrainTRACS and distributed to every operator using it, accumulating domain knowledge incrementally. The adopting of TrainTRACS has relieved the stress of planners facing tight deadlines and enabled them to focus on improving the schedule rather than making it feasible.

Multiple other solution methods for Crew Scheduling Problems are also presented in literature. Nishi et al. (2011) use column generation with dual inequalities to reduce solution time of the problem compared to ordinary column generation, and prove that using dual inequalities does not cut of optimal solution. Guillermo and José (2009) present a hybrid algorithm which uses tabu search and integer programming. The algorithm is tested with input data of Chilean operator MERVAL. Shen et al. (2013) present an adaptive evolutionary crew scheduling approach (AECS), which uses a hybrid genetic algorithm. Chromosome lengths vary adaptively during iterations, whereas in most of genetic algorithm –based approaches chromosome lengths are fixed. AECS is tested using real data from Chinese crew scheduling problems. AECS produces efficient solutions in short computing times. Compared to previous results produced by fuzzy genetic algorithms, better solutions were found for the test cases.

4. Solving a Crew Scheduling Problem

4.1. Planning software

In this chapter planning software of the Finnish railway operator is used to find solutions for real world crew scheduling problems. The solver of the planning software splits crew scheduling problem into two parts, duty generation and duty selection, each of which have their own parameters. There is also a possibility to add positioning trips, but the solver cannot incorporate shunting work into the solution. Tasks that need to be included in the solution must be placed to candidate tasks. The candidates can be sequenced using different sequencing methods. In the following test cases method “IP Blocks” was used. It couples most of the tasks to pairs; coupled tasks will not be broken apart in the solving process. Sequencing increases quality of solutions and makes solution times faster by reducing the possible amount of duties.

The planning software presents duties as shown in figure 2. Numbers indicate the number of the train, lines below numbers are tasks (positioning trips are marked differently, see the task under 69111 in the last duty), text below task is ending station of the task and * indicates a meal break.

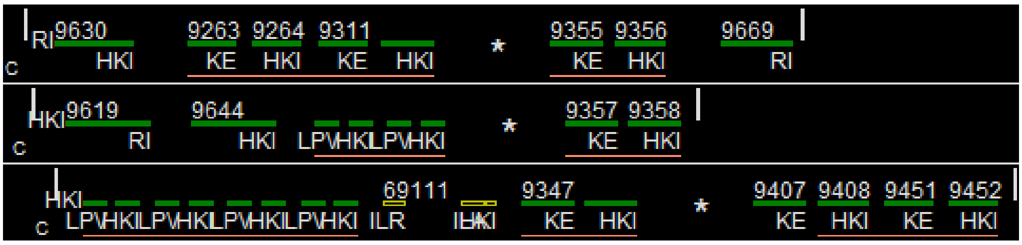

Figure 2: Duties in the planning software.

4.2. Optimization parameters

The solver of the planning software includes multiple optimization parameters. The objective in the test cases is to find combinations of parameters that minimize total working hours and amount of duties in the final solution, while amount of conflicts is as low as possible. First a relatively small crew scheduling problem is solved. Parameters altered while solving this problem were:

- *MinEff* is the minimum efficiency of duties, effective working time divided by duty duration. Increasing MinEff decreases the number of generated duties.
- *MinDur* is the minimum duration of duties. Labor rules constrain the smallest possible value for this to 7 hours. Increasing MinDur decreases the number of generated duties.
- *Ptrip* indicates if positioning trips are allowed, allowing is marked with X in Table 1.
- *Duty EC* is the extra cost per duty in final solution, increasing this reduces the number of duties in the solution.
- *CCWH EC* is the extra cost per calculated working hour in the final solution.

The second problem was larger, so two additional parameters were altered in order to reduce amount of generated duties:

- *Max RSC* is the maximum amount of rolling stock changes (driver switching from one rolling stock unit to another) allowed inside a duty.
- *Max PNB* is the maximum percentage of relieving time allowed inside a duty, the value is entered as a decimal.

Parameters MinEff, MinDur, Max RSC and Max PNB are parameters, which alter duty level constraints. Paramater Ptrip is related to formulation. Parameters DUTY EC and CCWH EC are weighting coefficients for the objective function; the total cost to be minimized is a linear combination of the number of duties and working hours. The software includes multiple other parameters also, but the ones altered in these experiments are the most important in the case of regional trains.

After optimization, the software provides detailed results. In this thesis the focus is on the amount of duties and total working hours. Abbreviations used in the result tables (tables 2 and 4) are:

- *Duties* is the total number of duties in the solution.
- *Total Hours* is the total number of working hours in the solution.
- *TOC* is the total number of overcovered tasks. A task is overcovered, if it is included in more than one duty that is included in the final solution.
- *HOC* is the number of overcovers in the middle of duties
- *BEOC* is the number of overcovers in the beginning/end of duties.

4.3. Small problem

In order to reduce the solution time, the set of tasks included in the optimization was first limited to four regional train lines. Solution times of approximately 1 hour were achieved for problems without positioning trips and solution times of approximately 2 hours were achieved for problems, in which positioning trips were included. This allowed testing multiple combinations of optimization parameters in a reasonable amount of time. The initial values for parameters (used in sub-problem 1) were given by an experienced planner. In total 13 different combinations of optimization parameters were used, which are presented in table 1.

Table 1: Parameters used in optimization.

SubProblem	MinEff	MinDur	Ptrip	Duty EC	CCWH EC
1	0.33	7:30		0:30	1:00
1P	0.33	7:30	X	0:30	1:00
2	0.40	7:30		0:30	1:00
3	0.33	7:00		0:30	1:00
3P	0.33	7:00	X	0:30	1:00
4	0.40	7:00		0:30	1:00
5	0.45	7:00		0:30	1:00
5P	0.45	7:00	X	0:30	1:00
6	0.50	7:00		0:30	1:00
7	0.40	7:00		0:00	1:00
8	0.40	7:30		0:00	1:00
9	0.33	7:30		1:30	1:00
10	0.33	7:30		0:30	10:00

When the parameters have been chosen, the software generates duties, taking into account all the labor rules and additional parameter constraints. Thereafter the duty selection phase begins. Every task must be covered at least once in the final solution. The software performs multiple iterations and produces a solution for each. The amount of solutions in different problems varied between 15 and 58. The best solution for each problem was determined by choosing the one with the smallest amount of overcovers and the smallest possible cost. Table 2 presents the best results of optimization of duties for each problem.

Table 2: Results of optimization.

SubProblem	Duties	Total Hours	TOC	HOC	BEOC
1	49	466:29:00	1	1	0
1P	49	467:09:40	2	2	0
2	49	465:53:40	1	1	0
3	49	465:44:40	1	1	0
3P	49	463:12:00	1	1	0
4	50	464:21:40	1	1	0
5	50	463:07:40	1	1	0
5P	50	465:27:00	1	1	0
6	49	464:30:40	1	1	0
7	52	461:57:40	1	1	0
8	50	469:06:20	2	1	1
9	49	468:21:20	2	2	0
10	50	468:24:00	2	1	1

In problems 1-6 Duty EC and CCWH EC were kept the same. With the initial values for parameters total working time of 466 hours and 29 minutes was achieved with one overcover. Next the optimization was run with same parameters, but enabling the use of positioning trips. This increased the amount of overcovers by 1 and also increased total hours. Adding positioning trips had a different effect in the case of problems 3 and 3P. Total hours decreased and amount of overcovers stayed the same. In problem 5P the enabling of positioning trips increased total hours compared to problem 5. Decreasing minimum duty duration from initial value of 7:30 to 7:00 lowered total working hours in solution (see problem 1 compared to problem 3 and problem 2 compared to problem 4). Effect of minimum efficiency was tested using four different values: when comparing problems 3, 4, 5 and 6, value of 0.45 produces best results. When positioning trips are added, problem 3P produces better results than problem 5P, even though the results of problem 3 are worse than results of problem 5. Problems 3P and 5 produced the best results. Problem 5 the has least total hours, and problem 3P has only four minutes more working time and the amount of duties is smaller by 1.

In problems 7-10 the effect of altering Duty EC and CCWH EC was tested. Extra cost per duty was lowered to 0 in problem 7. This helped to achieve the least total hours, but number of duties increased to 52. Results of problems 8-

10 were clearly inferior, when compared to problems 3P and 5, including more overcovers and more total working hours.

4.4. Large problem

Finally some runs using all the regional train lines were made. The size of the problem is over two times bigger, then the previous small problem. Solution times of the solver grow exponentially, as the problem size increases. The solver was left running overnight for each run and solution times of approximately 6-10 hours were achieved. As the planning software is 32-bit, this raises another problem in addition to long solution times. When the size of the data file including generated duties is over 2 GB, the planning software cannot read it and proceed to duty selection phase. Because of this limitation some additional parameters compared to previous small problem had to be altered and positioning trips could not be added in any of the sub-problems. Multiple test runs had to be made in order to find combinations of parameters, which leave the file size of generated duties below 2 GB. Table 3 presents the parameters used in successful optimization runs.

Table 3: Parameters used in optimization, large problem

SubProblem	MinEff	MinDur	Duty EC	CCWH EC	Max RSC	Max PNB
Big1	0.4	7:00	0:30	1:00	2	0.5
Big2	0.4	7:00	0:30	1:00	3	0.35
Big3	0.4	7:30	0:30	1:00	3	0.35
Big4	0.4	7:00	0:30	1:00	3	0.4
Big5	0.45	7:00	0:30	1:00	3	0.4
Big6	0.4	7:00	0:30	1:00	3	0.3

Best results of optimization with each parameter combination are presented in table 4. Small enough size of generated duties file was first achieved by lowering MaxRSC to 2. This produced a bad solution, which had too many duties and overcovers. Lowering Max PNB turned out to be a better way to decrease the size of generated duties file. Additional way to decrease the file size is increasing minimum duty duration, but that produced results with increased amount of overcovers and total hours (problem Big3 compared to Big2). Two values for minimum efficiency were tested, 0.4 used in problem Big4 produced results with less total hours and overcovers than 0.45 used in problem Big5. The effect of altering Max PNB was tested with three different values in problems Big2, Big5 and Big6. The lowest value 0.3 was used in problem 6, and the solution had least total hours. In problem 2, using Max PNB of 0.35 resulted in little more total hours, but amount of overcovers was decreased by one. Allowing positioning trips could have made the amount of overcovers smaller, but due to the limitation in generated duties-file size positioning trips could not be included in solutions.

Table 4: Results of optimization, large problem

SubProblem	Duties	Total Hours	TOC	HOC	BEOC
Big1	125	1095:55:00	24	18	7
Big2	111	1029:22:40	12	8	4
Big3	109	1039:19:00	20	11	9
Big4	111	1036:38:20	12	8	4
Big5	112	1037:40:00	14	10	4
Big6	111	1028:50:40	13	8	5

5. Conclusions and Discussion

This thesis discusses crew scheduling in railway transportation, using a case of Finnish railway operator as an example. Traditionally experienced planners made duties by hand, but demands of modern world have made it almost impossible to produce manually sufficient solutions in available timeframes. In the recent years many railway companies have adopted different kinds of planning software, which use techniques of operations research to create solutions for crew scheduling problems. Computer aided methods have provided feasible solutions that are more efficient and acceptable, than the ones created by hand. Automation of the process has alleviated the stress and workload of planning personnel. This has enabled them to move their focus from making the solution feasible to improving feasible solutions.

For the Finnish operator fundamental objective of crew scheduling is to form a set of duties, which includes all the tasks that need to be performed by drivers. Crew scheduling includes two phases, gathering base data and constructing duties. Main goals of the data gathering phase are making the base data error free and gathering the data in time. The most important issues that must be concerned while duty generation phase are economical – and efficiency targets, quality of work and constraints related to duties and schedule of planning process. The constraints related to duties include some simple restrictions such as "minimum duty length is 4 hours", but most of the constraints are more complex. After crew scheduling is finished, planners form rosters, which determine working days for each driver.

Two real world crew scheduling problems based on Finnish data are solved with the planning software of the operator. The effect of different optimization parameters on final solution is tested. Objective is to find solutions that minimize total working hours and amount of duties, while amount of conflicts is as low as possible. Multiple runs with different parameter combinations were made for both problems. One limitation of the planning software is that it can't process generated duties, if file size was over 2 GB. Because of this the selection of parameters in the large problem had to be done so, that amount of duties generated was reduced and possibly some good solutions were cut out. In some cases the planning software did not improve solution quality, when amount of generated duties was raised. For example generated duties of sub-problem Big4 include all the duties of sub-problem Big2 and additional duties with PNB of 0.35-0.4. This should make the solution quality of sub-problem Big4 better than in Big2, or in the worst case the quality should stay the same. Actual result was that Big2 produced better results. Adding positioning trips to the problem should also produce better results than without positioning trips, but this did not happen in all cases.

Nonetheless some clearly bad values for parameters were found, but in both problems there was little difference between the sub-problems, that produced best results. Further test runs will be needed to get more accurate results. Best combination of parameters seems to vary between different problems, so finding a perfect combination for every situation is impossible. Better approach would be to define intervals for the values of each parameter, which would produce good solutions in all (or at least most) real world situations.

References

[1] Abbink, E.J.W., Albino, L., Dollevoet, T., Huisman, D., Roussado, J. & Saldanha, R.L. 2010, "Solving Large Scale Crew Scheduling Problems in Practice", *Report / Econometric Institute, Erasmus University Rotterdam,* vol. 1, pp. 1-19.

[2] Abbink, E.J.W., Albino, L., Dollevoet, T., Huisman, D., Roussado, J. & Saldanha, R.L. 2007, "Solving large scale crew scheduling problems by using iterative partitioning", *OASIcs,* vol. 7, pp. 96-106.

[3] Caprara, A., Kroon, L., Monaci, M., Peeters, M. & Toth, P. 2007, "Chapter 3 Passenger Railway Optimization", *Handbooks in Operations Research and Management Science,* vol. 14, no. C, pp. 129-187.

[4] Fischetti, M., Kroon, L.G., Timmer, G., Vromans, M.J.C.M. & Abbink, E.J.W. 2004, "Reinventing Crew Scheduling at Netherlands Railways", *Interfaces,* vol. 35, no. 5, pp. 393-401.

[5] Guillermo, C.G. & José, M.R.L. 2009, "Hybrid algorithm of tabu search and integer programming for the railway crew scheduling problem", *PACIIA 2009 - 2009 2nd Asia-Pacific Conference on Computational Intelligence and Industrial Applications,* vol. 2, pp. 413-416.

[6] Hartog, A., Huisman, D., Abbink, E. & Kroon, L. 2009, "Decision support for crew rostering at NS", *Public Transport,* vol. 1, no. 2, pp. 121-133.

[7] Huisman, D., Kroon, L.G., Lentink, R.M. & Vromans, M.J.C.M. 2005, "Operations Research in Passenger Railway Transportation", *ERIM report series research in management Erasmus Research Institute of Management,* vol. 1566.

[8] Jütte, S., Albers, M., Thonemann, U.W. & Haase, K. 2011, "Optimizing Railway Crew Scheduling at DB Schenker", *Interfaces,* vol. 41, no. 2, pp. 109-122.

[9] Jütte, S. & Thonemann, U.W. 2012, "Divide-and-price: A decomposition algorithm for solving large railway crew scheduling problems", *European Journal of Operational Research,* vol. 219, no. 2, pp. 214-223.

[10] Kroon, L.G., Huisman, D., Abbink, E.J.W., Fioole, P., Fischetti, M., Maroti, G., Schrijver, A., Steenbeek, A.A. & Ybema, R.R. 2008, "The new Dutch timetable: The OR revolution", *Report / Econometric Institute, Erasmus University Rotterdam,* vol. 1, pp. 1-18.

[11] Kwan, R.S.K. 2011, "Case studies of successful train crew scheduling optimisation", *Journal of Scheduling*, vol. 14, no. 5, pp. 423-434.

[12] Narayanaswami, S. & Rangaraj, N. 2011, "Scheduling and Rescheduling of Railway Operations: A Review and Expository Analysis", *Technology Operation Management*, vol. 2, no. 2, pp. 102-122.

[13] Nishi, T., Muroi, Y. & Inuiguchi, M. 2011, "Column generation with dual inequalities for railway crew scheduling problems", *Public Transport*, vol. 3, no. 1, pp. 25-42.

[14] Shen, Y., Peng, K., Chen, K. & Li, J. 2013, "Evolutionary crew scheduling with adaptive chromosomes.", *Transportation research. Part B: methodological*, pp. 174-185.

A. Summary in Finnish

Junaliikenteen työvuorojen suunnitteluongelma on laajempi ja monimutkaisempi, kuin vastaavat ongelmat lento- tai linja-autoliikenteessä. Poiketen näistä toimialoista veturinkuljettajien tulee junien ajamisen lisäksi suorittaa muun muassa erilaisia vaihtotöitä (esimerkiksi vaunujen kytkentä) ja työvuorot sisältävät usein matkustamista. Lisäksi veturinkuljettaja voi vaihtaa ajettavaa junaa useaan kertaan työvuoron aikana ja junan koostumus voi muuttua tietyillä asemilla, vaikka kuljettaja pysyisikin samana.

Perinteisesti kokeneet suunnittelijat ovat tehneet työvuorot käsin, mutta nykyajan yhä tiukkenevat tehokkuus-, täsmällisyys- ja työtyytyväisyysvaatimukset ovat tehneet tästä lähes mahdotonta. Monissa maissa on otettu käyttöön suunnitteluohjelmistoja, jotka käyttävät operaatiotutkimuksen menetelmiä ratkaisujen tuottamiseen. Ohjelmistot ovat auttaneet tuottamaan tehokkaampia ja laadukkaampia käypiä ratkaisuja, kuin aiemmat käsin muodostetut ratkaisut. Tämä on pienentänyt suunnitteluhenkilöstön työtaakkaa ja mahdollistanut keskittymisen siirtämisen käyvän ratkaisun tekemisestä ratkaisujen hiomiseen.

Kandidaatintyössä käydään läpi suomalaisen rautatieoperaattorin työvuorojen suunnitteluprosessi ja perehdytään kahteen oikeaan dataan perustuvaan suunnitteluongelmaan. Työvuoroihin liittyviä rajoitteita käydään läpi ja suunnitteluongelma formuloidaan matemaattisesti. Kirjallisuuskatsauksessa esitellään muissa maissa käytettyjä ratkaisumenetelmiä. Lopuksi kaksi oikeaan dataan perustuvaa työvuorojen suunnitteluongelmaa ratkaistaan käyttäen suomalaisen operaattorin suunnitteluohjelmistoa.

Työvuorojen suunnittelun perustarkoitus on muodostaa työvuorot siten, että ne sisältävät kaikki tehtävät, jotka kuljettajien tulee suorittaa. Tehtävät ovat yleisimmin kahden aseman välisiä ajotehtäviä. Muita tehtävätyyppejä ovat esimerkiksi vaihtotöiden tekeminen ja matkustaminen seuraavan ajotehtävän lähtöasemalle. Ensimmäinen kahdesta suunnitteluprosessin vaiheesta on pohjatietojen kerääminen. Kaikki suunnitteluohjelmiston tarvitsemat tiedot, kuten junien aikataulut ja käytettävissä olevan työvoiman määrä, kerätään ja tuodaan ohjelmistoon. Tämän vaiheen tärkeimmät tavoitteet ovat datan virheiden korjaaminen ja datan kerääminen ajoissa suunnitteluprosessin aikataulun mukaisesti. Suunnitteluprosessin toinen vaihe on työvuorojen tekeminen. Työvuoroja muodostettaessa on huomioitava monia asioita, kuten tehokkuustavoitteet, työvuorojen laatu ja työvuoroihin liittyvät rajoitukset. Työvuoroista pyritään tekemään mahdollisimman kustannustehokkaita ja hyväksyttäviä (eli kuljettajien mielestä hyviä). Muita tavoitteita tässä vaiheessa ovat esimerkiksi töiden kokonaismäärän, yötyötuntien ja pyhätyötuntien jakaminen tasaisesti eri miehistöpaikkojen työntekijöille.

Työvuorojen muodostamiseen liittyy useita rajoitteita. Jotkut rajoituksista ovat yksinkertaisia, esimerkiksi työvuorojen minimipituus on 4 tuntia kaukoliikenteessä ja 7 tuntia lähiliikenteessä. Työvuoron maksimipituus taas on 10–15 tuntia riippuen työvuoron sisällöstä. Lisäksi työvuoroja muodostettaessa tulee ottaa huomioon esimerkiksi ruokataukojen pituudet, ajoaikojen pituudet, yövuoroihin liittyvät rajoitukset ja elpymisaikoihin liittyvät rajoitukset (elpymisajaksi lasketaan vähintään 12 minuutin tauko työtehtävissä).

Käsitelty ongelma voidaan formuloida matemaattisesti melko yksinkertaisella tavalla, kun oletetaan, että yksittäisiin työvuoroihin liittyvät rajoitteet otetaan huomioon ratkaisussa käytettäviä työvuoroja luodessa. Ensin siis tulee luoda joukko työvuoroja, joista valitaan lopulliseen ratkaisuun osajoukko, joka tuottaa optimiratkaisun. Kohdefunktio on kokonaiskustannusten minimointi. Rajoitusehdolla pakotetaan ratkaisu sellaiseksi, että kaikki tarvittavat tehtävät sisältyvät johonkin lopulliseen ratkaisuun valittuun työvuoroon. Lisäksi päätösmuuttujat rajoitetaan binäärisiksi. Formulointiin voidaan lisätä myös miehistöpaikka-tasoisia rajoitteita, joiden avulla voidaan tasata esimerkiksi työtunnit ja yötöiden määrä eri miehistöpaikkojen välillä.

Työvuorojen muodostamisen jälkeen suunnittelijat sijoittavat vuorot matriisimuotoisiin vuorotauluihin, joissa kukin sarake vastaa tietyn työpäivän töitä ja kukin rivi annetaan yksittäiselle kuljettajalle. Kun kuljettaja on tehnyt tietyn rivin työt loppuun asti, hän siirtyy seuraavalle riville. Kuljettajien kiertäminen vuorotaulun rivien läpi lisää työhön vaihtelua ja nostaa työtyytyväisyyttä.

Kirjallisuudessa työvuorojen suunnitteluongelmaa mallinnetaan yleensä joukon peittämisongelmana. Työvuorojen suunnitteluongelma on NP-kova, ja lisäksi ongelman koko on yleensä erittäin suuri. Tämä tekee kaikkien ratkaisujen generoinnista ja tarkan ratkaisun etsimisestä mahdotonta, joten ratkaisutavat perustuvat sarakkeiden generointiin ja erilaisiin heuristiikkoihin. Sarakkeiden generoinnissa käytetään lähtökohtana käypää ratkaisua, ja luodaan uusia työvuoroja, jotka voivat parantaa ratkaisua. Jokaisen generointikierroksen jälkeen ongelma ratkaistaan uudelleen.

Kaksi oikeaan dataan perustuvaa työvuorojen suunnitteluongelmaa ratkaistiin käyttäen suomalaisen rautatieoperaattorin suunnitteluohjelmistoa. Suunnitteluohjelmiston ratkaisin jakaa ongelmanratkaisun kahteen osaan, työvuorojen generointiin ja työvuorojen valintaan. Molempiin osiin voidaan vaikuttaa erilaisilla optimointiparametreilla. Ratkaisin osaa lisätä työvuoroihin matkustamista, mutta vaihtotöiden lisääminen ei ole mahdollista. Tehtävät, jotka halutaan sisällyttää ratkaisuun, tulee siirtää kandidaateiksi. Ennen ratkaisemisen aloittamista kandidaatit voidaan järjestellä käyttäen erilaisia menetelmiä, näissä tapauksissa käytettiin menetelmää *IP-Blocks*, joka teki useimmista tehtävistä pareja. Ratkaisin ei hajota näitä pareja missään vaiheessa, vaan käsittelee niitä kuten yksittäisiä tehtäviä. Tämä parantaa ratkaisujen laatua. Samalla laskenta-aika pienenee, sillä tehtävien järjesteleminen pienentää generoitavien työvuorojen määrää.

Ensimmäinen ongelma sisälsi neljä lähiliikennelinjaa ja oli suhteellisen pienikokoinen. Ratkaisuajat olivat noin 1–2 tuntia. Tämä mahdollisti useiden erilaisten parametrien kombinaatioiden testaamisen kohtuullisessa ajassa. Muutettaviksi parametreiksi valittiin työvuorojen minimitehokkuus *MinEff*, työvuorojen minimikesto *MinDur* ja matkustamisen lisääminen työvuoroihin (kyllä vai ei) *Ptrip*. Lisäksi säädettiin kustannusta per työvuoro (*Duty EC*) sekä kustannusta per työtunti (*CCWH EC*) lopullisessa ratkaisussa. Yhteensä tehtiin 13 ajoa erilaisilla parametrien arvoilla. Tavoitteena oli saada ratkaisuja, joissa olisi mahdollisimman vähän työtunteja ja mahdollisimman pieni määrä konflikteja, eli tilanteita, joissa tehtävää käytetään useammassa kuin yhdessä työvuorossa. Joitain selvästi huonoja parametrien arvoja löydettiin, mutta parhaat vaihtoehdot tuottivat tuloksia, jotka olivat lähellä toisiaan. Työvuoron minimipituus kannatti selvästi pudottaa pienimpään sallittuun eli 7 tuntiin. Minimitehokkuudelle ei löydetty selvää parasta arvoa, sillä paras arvo näytti riippuvan siitä, sallitaanko matkustamisen sisällyttäminen työvuoroihin. Työvuorojen ja työtuntien kustannusten muuttamisesta ei saatu merkittävää hyötyä.

Toiseen ongelmaan otettiin mukaan kaikki lähiliikennelinjat. Tämä kasvatti ongelman koon yli kaksi kertaa isommaksi, ja samalla ratkaisuajat nousivat 6–10 tuntiin. Suurin työvuorotiedoston koko, jota suunnitteluohjelmisto pystyi käsittelemään, oli 2 GB. Tämän vuoksi käyttöön otettiin kaksi uutta parametria, joiden avulla generoitavien työvuorojen määrää saatiin pienennettyä. Nämä olivat maksimimäärä kaluston vaihtamisia työvuoron aikana *Max RSC* ja maksimiosuus elpymisajalle työvoron pituudesta *Max PNB*. Matkustamista ei lisätty mukaan mihinkään ajoon, sillä tällöin ongelman koko olisi kasvanut liian suureksi. Yhteensä tehtiin 6 ajoa erilaisilla parametrien arvoilla. Tavoitteena oli jälleen saada ratkaisuja, joissa olisi mahdollisimman vähän työtunteja ja mahdollisimman pieni määrä konflikteja. Max RSC:n pienentäminen osoittautui huonoksi tavaksi rajoittaa ongelman kokoa, sillä ratkaisun laatu heikkeni huomattavasti. Sen sijaan Max PNB:n pienentäminen osoittautui hyväksi tavaksi. Ratkaisut näyttivät tulevan paremmiksi, kun Max PNB:tä pienennettiin, tosin kahdella pienimmällä arvolla saatiin lähes yhtä hyvät tulokset. Parametrille MinEff pienempi testatuista kahdesta arvosta tuotti parempia ratkaisuja.

Suurempikokoisessa ongelmassa olisi ehkä saatu parempia tuloksia, jos ohjelmistossa ei olisi ollut 2 GB:n rajoitetta työvuorotiedoston koolle. Rajoituksen vuoksi generoitavien työvuorojen määrä piti rajoittaa, mikä saattoi jättää optimoinnin ulkopuolelle hyviä ratkaisuja. Suunnitteluohjelmisto ei kuitenkaan aina onnistunut parantamaan ratkaisuja, vaikka generoitavien työvuorojen määrää nostettiin. Esimerkiksi kun Max PNB:tä kasvatetaan, niin generoitavien työvuorojen määrä ja mahdollisten ratkaisujen määrä kasvaa. Koska mahdollisia ratkaisuja tulee enemmän, lopullisen ratkaisun pitäisi olla parempi tai vähintään yhtä hyvä kuin pienemmän Max PNB:n ongelmassa. Ohjelmisto kuitenkin tuotti huonompia ratkaisuja, kun tämän parametrin arvoa kasvatettiin. Joka tapauksessa joitakin selvästi huonoja parametrien arvoja löydettiin. Molemmissa testitapauksissa parhaat parametrien arvot tuottivat ratkaisuja, joiden laatu oli hyvin lähellä toisiaan. Tarkempien tuloksien saamiseksi tulisi tehdä lisää testiajoja. Parhaat parametrien arvot näyttivät vaihtelevan eri ongelmien välillä, joten yksittäisiä parhaita parametrien arvoja ei todennäköisesti ole mahdollista löytää. Parempi lähestymistapa olisi määrittää jokaiselle parametrille tietty väli, joka tuottaisi hyviä ratkaisuja ainakin lähes kaikissa käytännössä esiintyvissä suunnitteluongelmissa.