Automated disruption management for urban railway systems under the considerations of energy efficiency

J. Adamek & R. Schönemann
Berlin Institute of Technology, Germany

Abstract

While energy-efficient train control allows for the optimal driving for a single train, there is much more potential in saving energy when the traffic on the whole network is examined. A combination of energy-efficient train control with methods of automatic rescheduling opens up possibilities for a smoother traffic and the prospect of reducing energy consumption in railway networks. In order to attain the minimal energy consumption in every operating status, the research project "Energy-Efficient Disruption Management" has been set up to investigate the potential of making dispatch decisions regarding the aspect of energy reduction. In the event of an unplanned interruption of service, the system automatically proposes several solutions. Various criteria, such as delays, but also the energy consumption are evaluated in order to support the decision making process at the control centre in case of disruptions.
Keywords: energy-efficient train driving, disruption management, rescheduling, dispatching, decision support system.

1 Introduction to automatic railway scheduling and dispatching

Control and scheduling of railway operations are performed by dispatchers located in the operations control centre. Due to its high complexity, railway control is characterized by manual work that requires most experienced dispatchers. In case of a disruption, rescheduling is too complex so that a manual solving is time consuming and does not guarantee for the optimal solution. Hence, algorithmic support and a higher degree of automation in railway scheduling is required to allow for a better usage of infrastructure as well as a faster and optimal dispatching.

The research analysing automatic rescheduling in railway traffic is extensive. Different approaches have been considered since the late 1990's, using various optimisation strategies. A boost of research in this field has been initiated by the Arrival project (c.f. [1]). Among others, the strategies mostly made use of in the past were knowledge based systems, genetic algorithms [2], or linear optimisation [3]. For a good introduction to railway scheduling and dispatching theory we refer for example to [4]. However, efforts to implement such systems are rare. The main reasons for this are the challenge to describe the complex context of rescheduling in a sufficient manner and the lack of strategies for integration in existing Computer Aided Dispatch/Automatic Vehicle Location (CAD/AVL) systems. Also, the optimal solution is required in real-time which cannot be ensured in most large networks. The research that analyses optimal train control has begun already in the 1980's but goes on until today. The most famous implementation was certainly the Metromiser (see [5]). Knowledge in optimal train control is applied at some railway systems but not in combination with automatic disruption management. A first approach combining these two fields can be found in [6] analysing traffic on bi-directional single passenger railway line in the UK.

An approach for urban and suburban railway transit systems is currently being developed in the project Energy-optimal Operation Control and Disruption Management (in German Energieoptimiertes Leit- und Störfallmanagement – ELS). The general project idea is briefly explained in chapter 2. Chapter 3 describes the technology of the DataWarehouse and in chapter 4 the development and assessment of dispatching strategies is discussed. Conclusions and outlook are comprised in chapter 5.

2 Idea and aims of energy-efficient disruption management

Reducing energy consumption in all conditions, in normal and in disturbed operation, was the basic idea of the project Energy-optimal Operation Control and Disruption Management (in German Energieoptimiertes Leit- und Störfallmanagement - ELS). The core of the development are two interrelated assistance systems:

- an on-board driving aid, called ecoM, to provide an optimal driving strategy, depending on the current traffic situation
- a decision support system, called ecoC, for the traffic controller that provides rescheduling suggestions in disrupted situations

Both systems interact with each other to enable an optimal traffic flow based on certain criteria such as the locations of other vehicles, connection relationships, track conflicts, etc. The whole system is illustrated in Fig. 1.

During a normal traffic situation, the system's only function is that of calculating driving and dwelling times for each train in order to achieve the energy-optimal driving performance. This is realised by the ecoM on-board unit on each train separately. In a disrupted stage, however, the system calculates three possible rescheduling suggestions for dispatching the situation and provides them to the traffic controller. Depending on the location and the type of disruption, new

Figure 1: Overall structure of the ELS system with the sub-systems ecoO (Offline-simulation and disruption management), ecoC (central unit), and ecoM (on board unit).

timetables for all affected trains are calculated. Furthermore, several key figures, e.g., punctuality, energy consumption, or timetable stability are calculated in order to assess and compare the three proposals. The final decision about the actual countermeasures will then be made by the traffic controller. The information about the new timetable is communicated to the trains where the new energy-optimal driving strategy is calculated. The ELS system uses a self-learning knowledge database which is filled with historic incidences and their dispatch solutions. When a new event occurs, historic data will be checked against the current traffic situation and recalculated for the production of dispatching strategies. Modern data mining algorithms and pattern recognition are used for a quick computation of dispatch strategies.

The project is funded by the TSB Innovationsagentur Berlin and is executed by research institutions and companies of the greater Berlin area and two sub-contractors. A public transport operator participates as an industry partner providing necessary data and allowing for prototype tests.

3 Learning of the past based on historical data

3.1 Incidents and dispatching methods

Incidences in railway operation are manifold and distinguishable. From broken vehicles or infrastructure elements to connection delays to fire-fighting operations all occurrences are possible. However, for disruption management it is essential to generalise or classify occurrences in order to generate optimal dispatch methods. Details of disruption types and dispatching rules are shown in Fig. 2. The classification is hierarchical ordered in three different levels with an increasing granularity, and every class of incidents, i.e. the most granular level, yields an approximated duration time based on empirical studies.

Rather than its actual type, the duration of an incident is important from the operations point of view. Through the generalisation process, each incident can be reduced to an infrastructure element that is temporarily not available. Information about the incident's duration and with it the length of the infrastructure blockade is especially required for choosing suitable dispatch methods. Its duration is not known in advance, and since conditions vary in every case, the duration prediction is difficult.

The classification of incidents in the ELS project is geared to realistic scenarios in underground systems. The prediction of incident's durations has been enabled by analysing historical incident records of an operation control centre. They have been subjected to a cluster analysis. Thus, three incident length groups (3, 35, and 75 minutes) have been identified and possible incident types have been mapped to these groups. This statistical analysis provided the basics for the formal description of incidents which has been developed as the next step. Moreover, the duration is essential to our allocation of all in this interval affected train numbers.

3.2 Data warehousing

Nowadays, novel database systems in the field of Computer Science need to handle a huge amount of uncompressed data up to 2 PetaBytes (e.g., Facebook with Hive [8]), which makes it more difficult to find certain patterns or to analyse these enormous volumes of data. Gaining knowledge from the data, e.g., in order to develop new opportunities for an organization, is a key task of Business Intelligence (BI) and led to an increase of interest in the research of novel database structures. Data warehouses assure the consideration of multiple dimensions to be analysed by focussing on query efficiency. Classical OLTP databases are not optimized for querying and BI analysis. For our approach, we modelled and implemented a data warehouse containing the data of historical incidents in a specific structure format, called Snowflake. We use PostgreSQL 8.4 as an open-source database management system. The attributes of the fact relation in the centre describe quantitative data, whereas the connected dimensions symbolize qualitative data, e.g., time, location, etc. As a data warehouse is optimized for analysing and generating data, short-term updates and inserts of data are realized
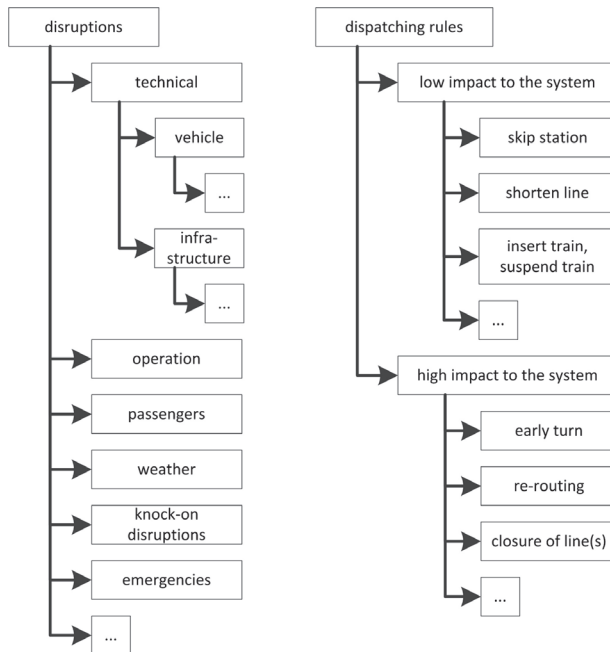
Figure 2: Disruption types and dispatching rules in ELS (selection, only 1st and 2nd order), partly based on [7].

in an operational database (OLTP database). In our case, new entries, after being evaluated in the online-simulation, are inserted in the operational database. To assure a certain timeliness, the data warehouse is as well updated in an initially defined interval. This actualisation by using the interface between OLTP- and multidimensional database, depicted in Fig. 3, is defined through an ETL-process (Extract, Transform, Load).
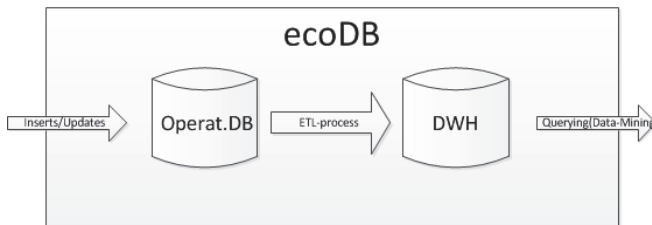
Figure 3: ETL-process between operat.DB and Data Warehouse (DWH).

One entry of each incident can have a maximum of three variations of dispatch decisions. In a following online-simulation step, every variation is evaluated concerning timeliness, energy-efficiency and stability. Beside our further

processing of our data, stored in the data warehouse, the framework could easily be modified and extended to a BI suite like Mondrian [9]. We will focus on Data Mining algorithms, which will be presented in the following section 3.3.

3.3 Pattern detection and evaluation of incidents

Based on a huge amount of historical data, our approach searches for similar pattern in the past. The user, in our case an operator at the control centre, needs to indicate the class of incident and the location, where the problem has occurred. These parameters are automatically taken into account in order to query all entries in the Data Warehouse and to reduce the set of incidents that correspond to the initial circumstances. By having the possibility of learning from past decisions of disruption clearances (in the case of an occurring incident), our approach generates the most frequently selected combination of dispatch decisions. We also consider the aforementioned evaluation criteria (i.e., punctuality, energy consumption, and timetable stability), which act as a weighting factor for choosing the optimal combination of dispatch decisions. One main component of the simulation process is a Data Mining algorithm, written in Java and integrated in a Tomcat-environment. Our algorithm to compute the optimal dispatch decisions is an extension of the tree-based TD-FPGrowth algorithm [10], published in the research field of Association Rule Miming (ARM). In comparison to the algorithm a-priori [11], the results of our evaluation show a high efficiency concerning runtime and accuracy. A transaction in our case is a set of dispatch decisions (set of items). The idea of Weighted Association Rules is adapted to the recursive TD-FPGrowth algorithm. Instead of allocating a value to every single item, every transaction has a weighting value.

Based on the existing itemsets, i.e., the historical data, it is necessary to compute probabilities about how frequent items and combinations of items can occur. The probability value has to meet criterions a threshold called minimal support, otherwise the item or itemset is not considered frequent. We will briefly describe

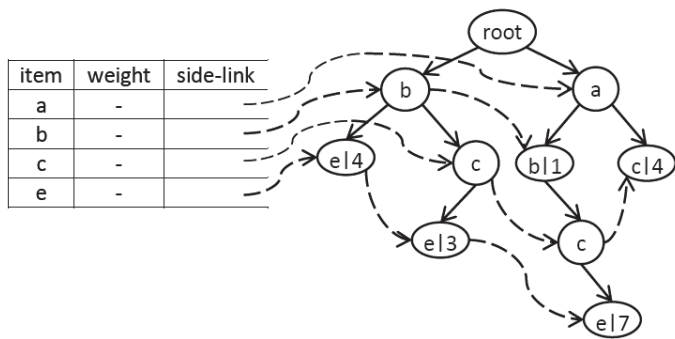| item	weight	side-link
a	-	
b	-	
c	-	
e	-	

Figure 4: Tree-structure of wTD-FPGrowth.

the functionality of our algorithm as follows: In the first step, only frequent items are considered and stored within the tree. After computing the sums of the weights for every frequent item, the itemsets are extended step by step by the next item. The traversal of the tree describes a combination of items, i.e., of dispatch decisions. To simplify the illustration of our algorithm, Fig. 4 is reduced to characters, each representing a dispatch decision and the corresponding weight. The table on the left side of Fig. 4 shows the initial header table, before computing the sums of the weights.

3.4 Priorization of incidents

The huge amount of data makes it necessary for users to distinguish between minor and major incidents, especially for a conclusive analysis. Another component of our framework improves the usability and should make it possible to emphasize serious disruption situations, respectively their entries in our database. In order to cluster entries of incidents in a meaningful way, we implemented a classification algorithm to assign all entries, beside their disposition to different priorization categories. Furthermore, the mathematical concept of our approach called cTD-FPGrowth is the aforementioned algorithm TD-FPGrowth [10]. Instead of computing itemsets, in the same step rules are generated.

4 Simulation and verification of dispatching strategies

The simulation of energy efficient train control and of dispatching strategies has been realised with the help of the microscopic railway simulation OpenTrack [12] and additional individual software. Infrastructure, timetable and vehicles have been modelled in OpenTrack. Thus, driving times and energy consumption have been retrieved. Later, the data have been exported to other tools for further processing. The so called offline-simulation (see Fig. 1) was the basic for testing and developing dispatching strategies.

4.1 Formal description of incidents

For the exchange of information between the sub-systems of ELS it was necessary to develop several exchange formats. Incidents and dispatching methods which have been simulated need to be stored in a database. This has been done with the help of an XML scheme, created to describe a disrupted situation and one or more appropriate countermeasures.

In detail, simulation results which have been obtained as CSV files were converted easily into the new XML format with the help of several scripts. The XML scheme (see Fig. 5) allows describing an incident in a formal way, amongst others with its specific type, the location and time. Thus, all information required for rescheduling actions are provided. The scheme also includes the possibility of delivering rescheduling actions. This is needed for the basic set of rescheduling actions for a knowledge based system.
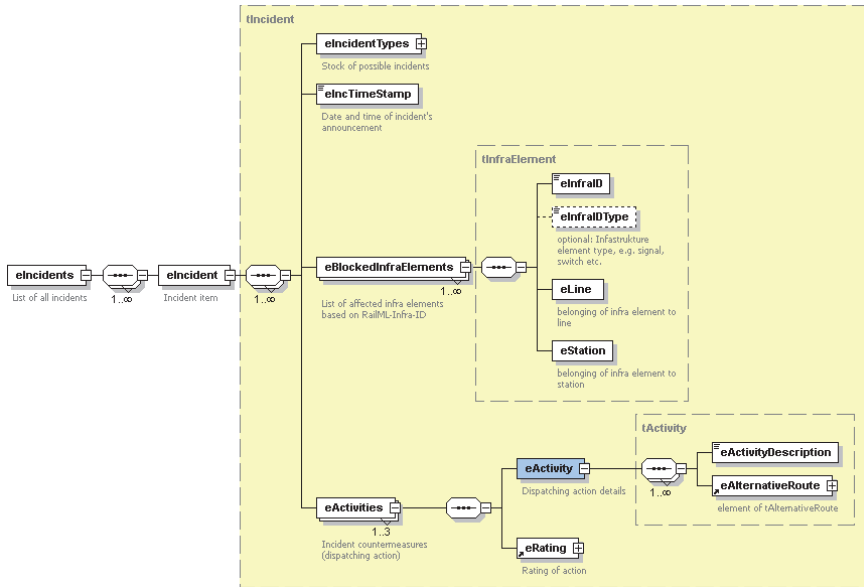
Figure 5: Detail of the Rail Incidence XML scheme to exchange rescheduling information between simulation and online system.

The fully formal description of incidents allows defining disrupted situations in railway operations and related dispatching methods in a precise manner. It provides the base input data for automatic rescheduling and the basis for further work on this topic.

4.2 Assessment of dispatching methods

For nearly all incidences the dispatching can be executed in more than one way. For an efficient disruption management it is necessary to find the optimal strategy depending on given objectives. The ELS project pursues different, partly conflicting optimisation criteria. Thus multiobjective optimisation methods are used.

In [13] for example, optimisation objectives are the minimisation of delays and delay costs. In ELS two more criteria are considered:

- Recover to timetable: time until the traffic has returned to a stable state where the level of service is as expected by passengers. The criterion is met when the service frequency is as scheduled; vehicle rotation may be divergent.
- Minimise energy consumption of dispatching method: The overall energy consumption during the disruption and the recovering phase is calculated for each dispatching strategy.

Other criteria, e.g. the efficient utilisation of infrastructure resources, are certainly considerable. During the test phase it will become evident whether other parameters are needed. For now, the three criteria delay, energy and recovery are considered. The evaluation of a dispatching strategy $d \in D$ provides the three cardinally scaled values:

- P_d (punctuality of strategy d)
- E_d (energy consumption of strategy d)
- R_d (recovery time of strategy d)

For each criterion P, E, R exists one optimal strategy among all the strategies D where $P = P_{opt}$, $E = E_{opt}$, and $R = R_{opt}$. To enable flexible assessment, weighting factors $g_x, x \in \{P, E, R\}$ are introduced. According to the company-specific weighting, these factors can have a value between 0 and 1. This results in the following term to assess the quality Q of dispatching method d:

$$\frac{P_d}{P_{opt}} \times g_P + \frac{E_d}{E_{opt}} \times g_E + \frac{R_d}{R_{opt}} \times g_R = Q_d$$

with $g_P, g_E, g_R \in [0; 1]$ and $g_P + g_E + g_R = 1$.

By means of the previously described evaluation system, a scale is set up that allows the dispatcher to grasp differences between various dispatching alternatives instantly. Thus, it speeds up the decision making process for rescheduling.

5 Discussion and conclusions

In this paper, we present a decision support system for energy-efficient automatic railway disruption management. In particular, approaches for energy-efficient railway simulation, the generation of automatic dispatching rules, and their assessment have been evaluated. The system is going to be tested under real conditions by summer 2012. Although tests have not been carried out yet, further steps of research became apparent.

First of all, the automatic dispatching routines will be developed further. There is high potential in building a system that is not only dependent on historic data. Furthermore, the assessment of dispatching alternatives will be updated when new decision criteria, e.g. those which are contradictory, evolve. Finally, there is a need of integrating the ELS system in existing ITCS. The prototype will be set up as an add-on system which results in partial double work load for the traffic operator. A smart integration in ITCS however leads to more automation (e.g. conflict identification) and less work load for the operator.

References

[1] Ahuja, R.K., Möhring, R.H. and Zaroliagis, C.D., (eds.) *Robust and online large-scale optimization*. Springer: Berlin and New York, 2009.

[2] Wegele, S. and Schnieder, E., Automated dispatching of train operations using genetic algorithms. *Computers in railways IX*, ed. J.J. Allan, WIT Press: Southampton, England and Boston, 2004.

[3] D'Ariano, A., *Improving Real-Time Train Dispatching: Models, Algorithms and Applications*. Ph.D. thesis, Technische Universiteit Delft, 2008.

[4] Törnquist, J., Computer-based decision support for railway traffic scheduling and dispatching: A review of models and algorithms. *5th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS'05)*, eds. L. Kroon and R.H. Möhring, Schloss Dagstuhl - Leibniz-Zentrum für Informatik GmbH: Wadern, 2006.

[5] Howlett, P.G. and Pudney, P.J., *Energy-efficient train control*. Springer: London and New York, 1995.

[6] Albrecht, A., Koelewijn, J. and Pudney, P.J., Energy-efficient recovery of delays in a rail network. *Australasian Transport Research Forum 2011 Proceedings*, ed. University of South Australia, Adelaide, 2011.

[7] Jespersen-Groth, J., Potthoff, D., Clausen, J., Huisman, D., Kroon, L., Maróti, G. and Nyhave Nielsen, M., Disruption management in passenger railway transportation. *Robust and online large-scale optimization*, eds. R.K. Ahuja, R.H. Möhring and C.D. Zaroliagis, Springer: Berlin and New York, 2009.

[8] Thusoo, A., Shao, Z., Anthony, S., Borthakur, D., Jain, N., Sen Sarma, J., Murthy, R. and Liu, H., Data warehousing and analytics infrastructure at facebook. *Proceedings of the 2010 international conference on Management of data*, ACM: New York, NY, USA, SIGMOD '10, pp. 1013–1020, 2010.

[9] Pentaho, Pentaho analysis services (mondrian). online, http://mondrian.pentaho.com, 2012-06-09.

[10] Wang, K., Tang, L., Han, J. and Liu, J., Top down fp-growth for association rule mining. *Proceedings of the 6th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, Springer-Verlag: London, UK, UK, PAKDD '02, pp. 334–340, 2002.

[11] Agrawal, R. and Srikant, R., Fast algorithms for mining association rules in large databases. *Proceedings of the 20th International Conference on Very Large Data Bases*, Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, VLDB '94, pp. 487–499, 1994.

[12] Hürlimann, D. and Nash, A., *OpenTrack: User Manual Version 1.6*, 2010.

[13] Törnquist, J., Railway traffic disturbance management—an experimental analysis of disturbance complexity, management objectives and limitations in planning horizon. *Transportation Research Part A: Policy and Practice*, **41(3)**, pp. 249–266, 2007.