A multi-component closed-loop control framework for rail traffic networks

Raimond Wüst, IDP, ZHAW
Albert Steiner, IDP, ZHAW

Conference paper STRC 2012

STRC | **12th Swiss Transport Research Conference**
Monte Verità / Ascona, May 2 – 4, 2012

A multi-component closed-loop control framework for rail traffic networks

Raimond Wüst
IDP
School of Engineering
ZHAW

Rosenstrasse 3
8401 Winterthur
Switzerland

Phone: +41 (0)58 934 65 81
Fax: +41 (0)58 935 65 81
Email: raimond.wuest@zhaw.ch

Albert Steiner
IDP
School of Engineering
ZHAW

Rosenstrasse 3
8401 Winterthur
Switzerland

Phone: +41 (0)58 934 78 01
Fax: +41 (0)58 935 78 01
Email: albert.steiner@zhaw.ch

May 2012

Abstract

The operation of rail transport systems has become an increasingly challenging task over the last decades. One of the main reasons is the conflict between the surge in passenger and freight demand and the capacity constraints of the railway system (on railway lines and at railway stations). To allow for a better understanding of the system dynamics in different operational states (including disruptions) and to have improved control strategies at hand, a multi-component simulation framework, representing an entire closed-loop operation environment for railway networks is currently under way. Based on this framework, railway network operators shall be enabled to investigate timetable and scenario generation, railway traffic operations, operational decision support, and operational performance criteria. In addition, the architecture of the framework shall allow for investigating the effects of moving from a distributed, event driven train control system (where involved operational parties are only weakly interacting) towards an increasingly integrated, time controlled and automated operational concept. This concept of tight time control requires all operational processes to be continuously monitored with respect to the production schedule. Deviations exceeding some pre-determined tolerance thresholds will results in a re-adjustment of the production plan in real-time. A dedicated rescheduling algorithm will be implemented to achieve this goal. This algorithm is based on a resource-constrained multi-commodity flow model for conflict-free train scheduling recently developed at ETH Zurich.

To explicitly take into account the various parties (agents) involved, the system can be configured such that the agents behave according to existing or virtual profiles. For instance, train drivers might be technically enabled to follow new operational targets like re-adjusted train speeds while approaching conflict points. With this framework, scenarios for different operational and technical conditions as well as for internal and/or external disturbances can be investigated and the differences in performance with respect to customer satisfaction can be evaluated (represented for example by the overall train delay, cumulated passenger delays or timely availability of customer information). The framework, called *Rail Transport Service Environment* (RTSE), consists of three main modules: (i) a traffic simulation environment, (ii) a system state monitoring module, and (iii) the rescheduling algorithm mentioned above. The modules are interconnected through standard communication interfaces so that each module can be exchanged easily depending on the user environment. Railway traffic simulations are carried out using the dedicated railway simulation tool OpenTrack. The tool is used by various train operation and train infrastructure companies in order to test, amongst others, the feasibility of timetables or signalling scenarios. The simulated traffic scenarios will be interpreted by an automated monitoring module (including some threshold detection mechanism, where the thresholds depend on the operational policies, etc.), which compares actual and planned process states and induces rescheduling actions executed by the rescheduling algorithm, if required. Those rescheduling actions result in new (adjusted) state space plans, which take into account changes in process states and the eventually reduced availability of resources.

Keywords

Rail traffic – Service intention – Rail traffic state monitoring – Real-time dispatching – Network performance – Rail traffic simulation

1. Introduction

Due to a continuously increasing demand, public transport networks became highly utilized in many countries over the last years. Travel and transport requirements often grow at a much faster pace than financial budgets, network services or space available for network extensions. As a consequence, several problems arise for both train operating carriers (TOC's) and infrastructure operating companies (IOC's), which are difficult to solve.

1.1 Motivation

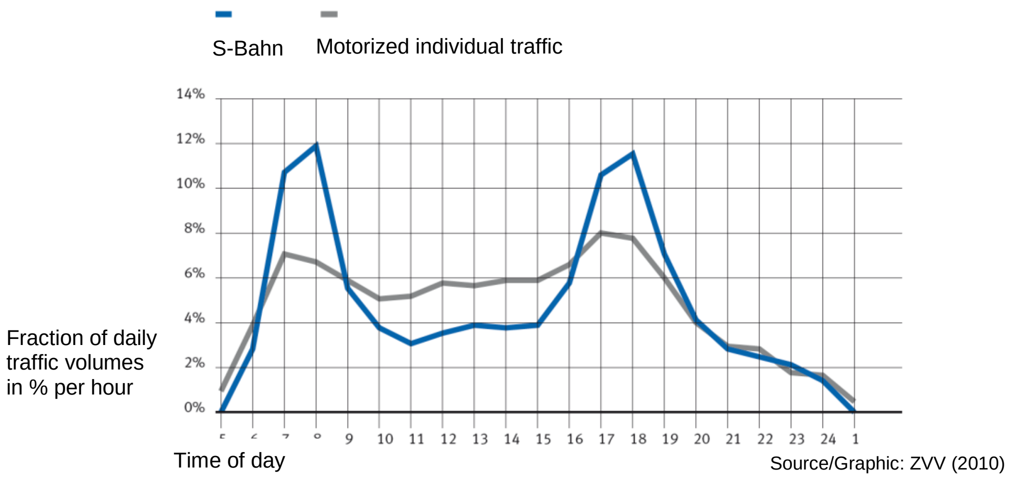
The identification of increasing capacity problems that ask for a redesign of railway operations was the main motivation of the project participants for developing the proposed framework. In order to derive the elements of the proposed approach, we will first have a closer look on the major problems.

Problems in service delivery

Especially during peak hours, operating staff and control centres are facing enormous challenges to make sure that reliable transport services can be provided.

Figure 1 Illustration of some of the main problems in public transport: (a) Very high peak demand (e.g., in S-Bahn Zurich), (b) Operating staff confronted with decreasing service reliability, each with different responsibilities within the service process chain.

(a)



(b)

The empirical facts on average public transport demand in the course of the day for S-Bahn Zurich (see ZVV, 2010) as shown in Figure 1 (a) together with some typical situations the operating staff is facing (see Figure 1 (b)) shall provide a view on some problems: (a) the conflict between demand and supply and (b) the lack of timely information on the system provided to both to operating staff as well as to the customers.

Following the above explanations, the major service delivery problems can be summarized as follows:

- Tight regular timetable due to an increasing gap between peak hour and off peak hour demand.
- Decreasing service reliability due to operational volatility and technical disturbances.
- Limited usability of public transport due to communication problems.
- Significant total passenger delays due to local dispatching decisions.

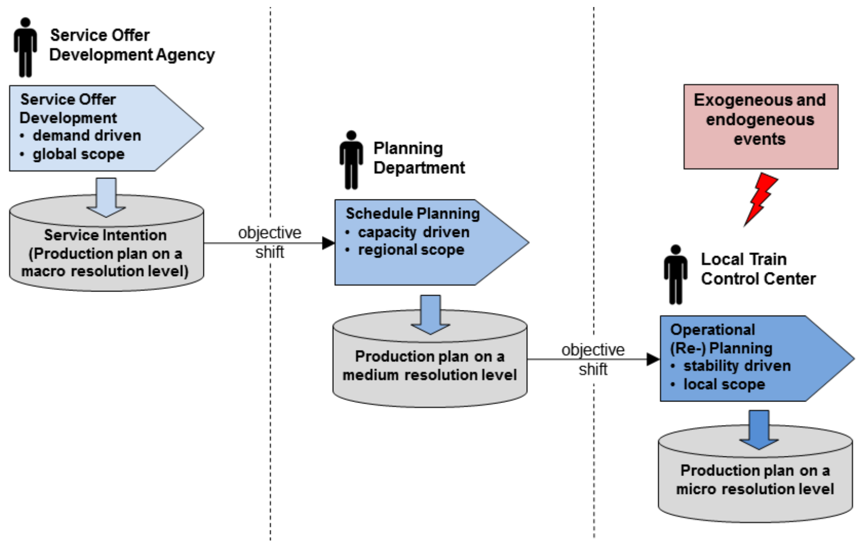
If we analyse specific cases and try to figure out what happened before major network delays, we observe similar patterns. We see, that in most cases operational disturbances lead to blocked resource assignments and hence unusable production plans. As the production plan is the technical basis for operating the timetable, it is the main task of operational control to reassign resources such that the normal timetable will be restored as soon as possible. This is a complex task that requires, even in the case of small delays, the consideration of numerous operational and technical constraints.

Problems in service development

Usually timetable development is an extensive, iterative planning process performed years ahead of the beginning of the actual timetable period. The main objective is to find train runs with departure, arrival and dwell times that simultaneously meet functional requirements (Service Intention, SI) connecting each origin and destination in the network with required travel times, frequencies, service levels and technical requirements (the utilization of available resources, such as track topology, rolling stock and operation staff). This is a strongly iterative process, which usually requires a lot of feasibility checking. It is considered to be unrealistic to try implementing such a process in operational environments. The main problem of this service delivery process is its sheer complexity (see Figure 2):

- The development process is sub-divided into three phases: 1. Service offer development, 2. Timetable development, and 3. Operations and Dispatching.
- Organisation and tools involved in each of the development phase are specific for each phase.
- Each phase transition is related to a certain shift of objectives (phase 1: demand, phase 2: capacity, phase 3: stability)
- Taking dispatching decisions during operation, it is very difficult to reach the original service intention again in real-time.

Figure 2 Service offer value chain: Development process in three phases, distributed over many years. From this it becomes obvious, that it is very difficult to refer in all phases to the same objective (service intention).

<table>
<tr><td>Planning level</td><td>Long term planning (Service offer and resources)</td><td>Mid-term planning (Timetable)</td><td>Short term planning (Operations and dispatching)</td></tr>
<tr><td>Time horizon</td><td>20 to 5 years</td><td>5 years to 24 hours</td><td>24 to 0 hours</td></tr>
<tr><td>Interrelations</td><td colspan="3">
</td></tr>
<tr><td>Level of detail</td><td colspan="3">
low level of detail … high level of detail</td></tr>
</table>

Probably this is the reason why in case of disruptions dispatching, decisions are commonly determined by the attempt to keep existing timetable components untouched as long as possible. In fact this has been the most common way in operation so far since there were no appropriate planning tools available in the past, which enabled dispatchers to allocate resources to functional requirements in real-time. Besides the circumstance of missing tools, the skills and information available for solving this planning task are usually located in the planning departments responsible for phase 1 and 2 (Stalder, (2005)) or even other organisations than in the operational control centres. Nevertheless it has been stated in several studies that the implementation of real time rescheduling is a promising way to increase the efficiency of heavily used railway networks (Lüthi 2008, 2009, Wegele et al., 2007, van den Boom et al., 2011).

1.2 Contribution of this work

Most of these problems turned out to originate in a lack of formal transparency of operational process objectives. In order to eliminate this drawback, a couple of years ago, the Swiss Federal Railways (SBB CFF FFS) together with the Institute for Operations Research (IFOR) at ETH Zurich, decided to develop a semi-automated planning procedure that supports planning staff such that functional requirements together with the available set of operational resources are converted into a concrete production plan. This production plan should be based on processes that are fully described by process durations and timed events. Procedures were developed in order to generate these production plans almost in real-time. Once generated, a production plan has to be formatted appropriately either for timetable publication or for the purpose of instructing operational staff such as train drivers, train conductors or dispatchers (Wüst et al., 2008). The result of this initiative, i.e. the concept of the planning procedure, was published in Burkolter (2005) and Caimi (2009). Only recently an approach for utilizing this evidence for real-time train dispatching in the area of the railway station of Berne has been published by Fuchsberger (2012).

This paper proposes a simulation framework that is based on the concepts mentioned, which evolved during the past collaboration. The framework will be jointly developed by several partners from research and industry in a project which is funded by Swiss CTI-Organisation (contributing partners: IDP (ZHAW) Winterthur, InIT (ZHAW) Winterthur, IBM Research GmbH Rüschlikon, OpenTrack Railway Technology Zürich, Emch + Berger AG Bern). The project started in autumn 2011, and hence some of the concepts presented shall be considered as work in progress as they will be further refined in the course of time. Nonetheless, the concept described in this paper represents the common view of the research partners on how railway network operations should be configured and how the framework should be constructed.

1.3 Organisation of the paper

The rest of the document is organised as follows: In chapter 2 we introduce the fundamental elements required for the proposed concept. This includes, amongst others, the definition of some important terms (e.g., service intention) and concepts (e.g., time controlled and event controlled railway operation). Furthermore, special network performance effects that are specific for systems with clock-face timetables, are introduced. In chapter 3 we present an approach for solving the problems described in chapter 2. This solution consists of a process (3.1) and a system view (3.2). Subsequently, the components of the proposed framework are described at a functional level. Section 3.3 describes the rail control problem solution in a state space approach. Finally we discuss the proposed performance measure with respect to this state space model. Section 3.4 describes common practice to estimate and provide service quality in different railway system environments. Our framework is designed to allow for comparing network operations in different environments. In chapter 4 we explain the intended practical application of the framework with respect to the Danish rail infrastructure implementation project. Finally, in chapter 5 we summarize the main points achieved by the project so far and provide an outlook on the next steps as well as on future research.

2. Fundamental elements of the concept

There are three major conceptual elements for implementing the framework which will be described in the following sections: (i) the service intention, (ii) time controlled production processes and (iii) network performance measures. None of these three conceptual elements is really new in the railway environment. However, for real-time control of railway networks these elements have not been combined so far. In the following, we discuss why combining these elements is highly beneficial for railway operations.

2.1 Service intention as objective of optimized rescheduling in case of operational disturbances

The central conceptual element of the proposed approach is the assumption that any timetable is one out of several possible technical realizations of an underlying service intention (SI). Because the SI can be regarded as a functional requirement for the scheduling task, it is defined in terms of frequencies between origins and destinations, travel times and service levels rather than of exact departure and arrival times, rolling stock utilization or even trains numbers. Although these attributes are information typically provided by a public transport timetable, it is nothing else than the result of an assignment of resources (consisting of track topology, rolling stock and operational staff) to these functional requirements, while considering numerous spatio-temporal constraints. As a consequence, operational irregularities or disruptions, which, for a certain period of time prohibit further operation of the timetable planned, require a new assignment of available resources to the functional requirements. The partial periodical service intention is defined as follows: "*The concept of a partial periodic service intention (short ppSI) is introduced here as an interface between commercial offer and technical process planning. It consists of all services a railway company would like to offer during a day. Each train service is specified by its line, stopping stations, interconnection possibilities, periodicity, and the time frame. The ppSI is not the representation of a technical timetable, but it describes only the commercial offer and contains therefore only the customer-relevant information.*" (Caimi (2009), p. 52).

The following explanations focus on a pure periodical service intention, which is a special case of the partial periodical service intention. According to Caimi (2009), a periodical service intention $\tilde{G}$ is defined as follows:

$$\tilde{G} = \left(Z, C, D, \bar{\rho}\right), \tag{1}$$

where Z denotes the set of all train runs, C is the set of all connections, D denotes the set of all technical and operational dependencies and $\bar{\rho}$ is a time period.

Now, there exists a value T with the following properties: (i) For all train runs $z \in Z$ there exists a value $n \in \mathbb{N}$ with $\rho(z) \cdot n = T$. This means that train run z is repeated exactly n times during the period. (ii) There is a time interval of length T where n repetitions of the train run are completely included, either with n full train runs or with $n-1$ full train runs, plus one leaving the time interval and one entering it in the same place. These two fractional train runs can be seen as one complete train run *modulo* T.

In this case, the service intention $\tilde{G}$ is said to have periodicity T. It could be described in an equivalent and efficient way using $\bar{\rho} = T$.

A single train run z (with $z \in Z$) is defined as the run over $K+1$ stations in the railway network, repeated R times with periodicity ρ (minutes). It is formally specified by

$$z = \left(\tilde{z}, \left(v_k, t_{\text{dwell}}^{k-}, t_{\text{dwell}}^{k+}, t_{\text{trip}}^{k-}, t_{\text{trip}}^{k+}, \omega_k^-, \omega_k^+ \right)_{k=0}^{K}, \rho, R \right), \qquad (2)$$

where $\tilde{z}$ denotes the train type used for train run z, with $\tilde{z} \in \tilde{Z}$ and $\tilde{Z}$ being the set of all train types considered, k denotes the number of a station with $K+1$ being the overall number of stations considered (with $k \in K$), $\tilde{z}$ is the t_{dwell}^{k-} and t_{dwell}^{k+} denote the minimal and maximal dwell time at station k, respectively, t_{trip}^{k-} and t_{trip}^{k+} denote the minimal and maximal trip time between vertex v_{k-1} and vertex v_k, ω_k^- and ω_k^+ represent the minimal and maximal connection time at station k. For a detailed discussion of the formalism we refer again to Caimi (2009), section 3.5 ff.

The SI is not a representation of a technical timetable, but rather a description of the relevant train services and therefore contains only information, which in some way is relevant to the service offered to the customers.

2.2 Time controlled versus event controlled operations

In the domain of public transport and especially in rail traffic management the production plan plays a central role. It serves as a basis for communicating the concrete realization of the SI to the customer (timetable format) as well as for scheduling and instructing operations staff. The production plan tells exactly when resources are going to be allocated to which part of the transport offer. And it can be used to provide recommendations to train drivers, instruct conductors when to close the doors or to supply customers with individual trip information in real-time.

However, due to numerous dependencies between individual technical and operational details of the production plan on the one side, and the strong variability of process times on the other side, process control has evolved to be typically event driven (e.g. waiting (for how long?) until event x occurs to execute process a; if event x occurs execute process a, else execute process b). If the operational environment is characterized by unpredictable events and process durations this event based conditioning provides a lot of temporal flexibility. As such, this has led to the fact, that interlockings are still configured using train route allocation sequences rather than train route allocation times. This allows trains to request route allocation at flexible points in time if at least the planned sequence is maintained. Especially in capacity constrained traffic networks that require frequent and quick reassignments, this fact represents one of the major obstacles to keep time based production plans on track (Roos, 2006).

Tests carried out by SBB and presented in Lüthi et al. (2007) have shown that changing operational processes to be time controlled and rescheduling them in case of disturbances, increases operational stability significantly while at the same time reduces the variability of process times. This so-called ‘six sigma-effect’ is well known from service management literature (e.g. Fitzsimmons, 2008, p.157).

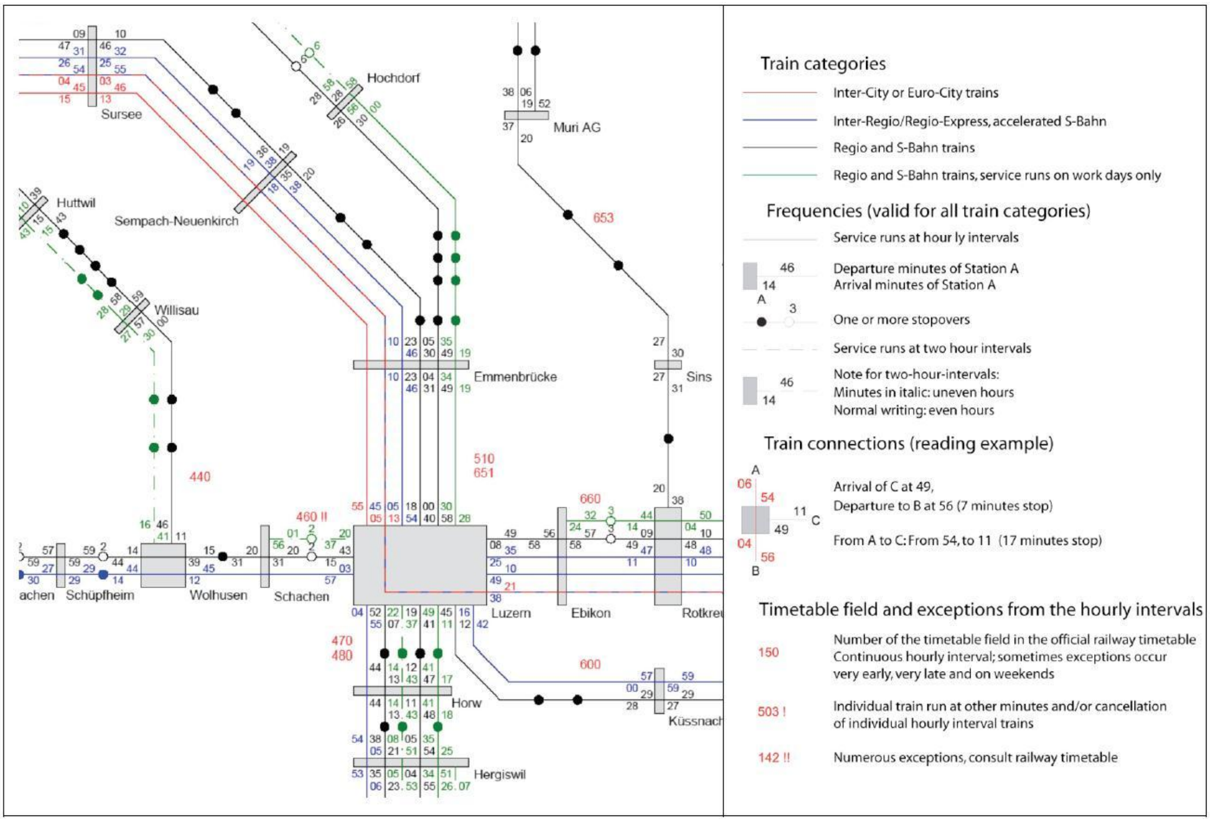
A consistent time-based design of operation processes has the following five important advantages:

1. Time control accelerates process chains and enables service benchmarking: operation processes, which are exclusively coupled by logical constraints can only be executed sequentially. Timed processes can easily be executed in parallel. In the context of train operations this helps increasing capacity.
2. If resource allocations are time based problems and their causal relations can be located faster, it is even possible to automatically calculate complex delay propagation effects and estimate network impacts (Goverde, 2005).
3. In heavily used railway networks efficiency can be improved by modelling operation processes with control loops (Lüthi, 2007). Quantitative performance measures require certain system dynamics. Planning and monitoring process times is mandatory for achieving the required system dynamics.
4. Natural fluctuations of process running times can be monitored effectively with respect to tolerance intervals. Incidents of exceeded tolerance thresholds can be used for triggering the rescheduling action (Fitzsimmons, 2008, p.125).
5. Strictly time based resource allocation enables the quantitative definition of capacity utilization given a certain timetable. Quantitative capacity utilization measures facilitate strategic infrastructure planning decisions.

2.3 Network based in favour of line or station based performance measures

The Swiss railway network and the services running on it can be characterized as a typical hub and spoke network with integrated clock-face timetable and it is well known to be strongly interconnected. This means that there are lots of point to point services that require one or more train transitions (see for instance the Swiss timetable: Figure 3 shows a detail of the timetable 2005 around Lucerne as network graph). In these cases individual delays have very often impacts on a large part of the network. This is the case, if train dispatchers decide to hold back the connecting trains affected as well as if they decide not to keep communicated connections. In the first case, the initial delay of the feeding train not only causes a delay for the passengers of this train but also for the passengers that are not connecting to other trains. In the latter case in which connections are broken, all connecting passengers will suffer a much higher delay as they have to wait for the next connecting train to their intended destination.

Figure 3 Clock-face timetable for the area of Lucerne, Source: SMA and Partner AG Zurich, Timetable 2005.



The high connectivity of services as illustrated in Figure 3 for the area of Lucerne in 2005 provides a good idea that assessing the service quality in highly utilized train traffic networks requires network performance measures, which take temporal and spatial delay propagation into account. Pure line or station oriented views of train punctuality do not capture this delay propagation effect. Next we will demonstrate the impact of a network based view on total train delay by means of a simple example.

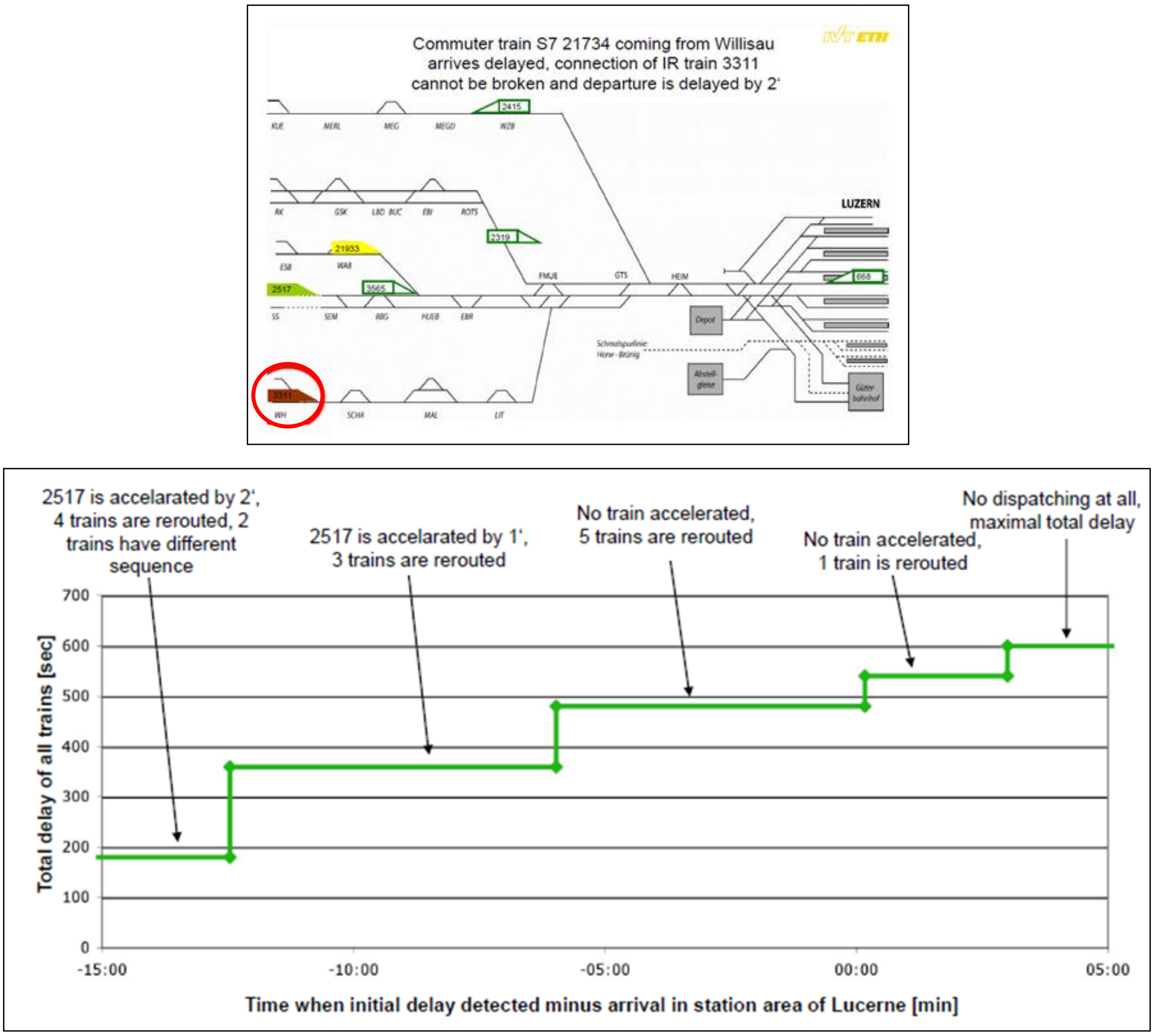
Network effects

The influence of different operation control types in terms of the total delay of train movements within a regional control area was investigated by the operations department of SBB infrastructure and the IVT (Institute for Transport Planning and Systems, ETH Zurich) in a case study (see Lüthi, 2007).

In this study the delay propagation of a delayed commuter train in the area of Lucerne station has been simulated with different dispatching scenarios in OpenTrack (see Figure 4). For each scenario it was assumed that the operations control centre had a specific delay detection mechanism and different options of interacting with affected trains. In the best scenario, delay could be detected immediately after a tolerance threshold of 15 seconds for the arrival of the delayed commuter train was exceeded. In this case there was time enough to accomplish all actions required in order to reduce the total delay of all trains including accelerating other trains approaching Lucerne (this requires an appropriate way to communicate optimal target speed to train drivers), rerouting approaching and departing trains and changing route sequences. In this case the total delay (including outgoing trains) could be reduced to 3 minutes, which is not much more than the original delay. In the worst case, the delay was not detected at all and no dispatching actions were taken. This resulted in a total train delay of 10 minutes.

One should note, that operation control technology in Lucerne today allows rerouting trains and changing route sequences. Hence, with the installed delay detection technology, a reduction of the total train delay of around 8 to 9 minutes would be feasible. This shows that there is significant potential for improvement, provided that the following requirements are met: an appropriate technology for delay detection, a fast and high-performing rescheduling algorithm and an efficient and reliable implementation of communication equipment available for operating staff.

Figure 4 Example of simulated rail traffic scenarios using OpenTrack for two different dispatching strategies after detection of an initial delay of two minutes in the area of Lucerne.



3. Proposed solution

Recent research shows that it is possible to allow for a functional division into subtasks, which are required for the reassignment of resources (restricted or accessible with delay) to functional requirements (SI) in real-time (Fuchsberger, 2012; van den Boom et al. 2011). Therefore we propose an approach, which technically integrates all three phases of the service offer value chain according to the aims listed in the preceding section.

3.1 Service Intention based integration of development processes

From the explanations and arguments provided in sections 1.1 and 2.1, it becomes clear that a solution needs to focus on service intention. To achieve this, the proposed integrated platform solution requires the following:

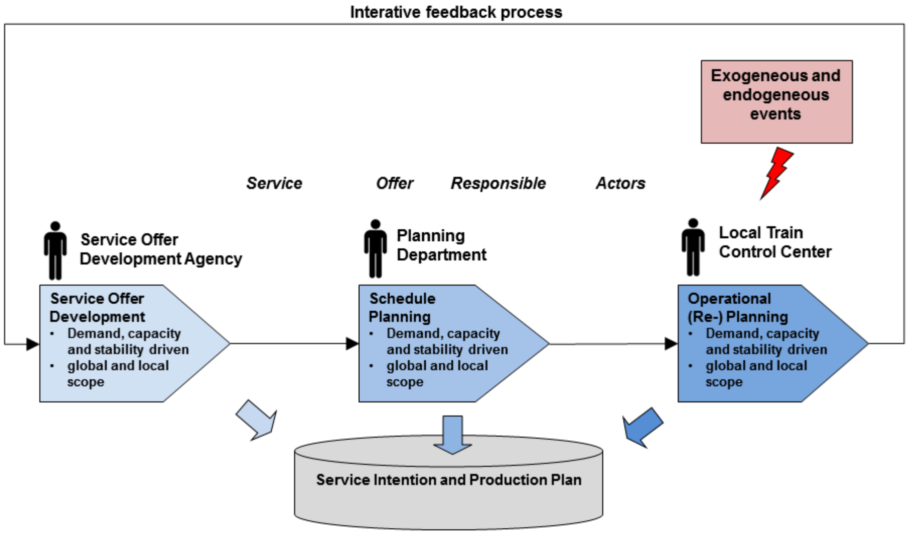
- A global data model (e.g. train definition).
- Global rules (e.g. resource conflict conditions).
- Scalability of both traffic network and system architecture (to insure optimality of production plan with a global network scope).
- The assessment of the process state (in case of operations), the handling of technical conditions, and the recalculation of an SI-based production plan in real time.

If we sketch this proposed solution in the process development scheme of Figure 2, the service offer value chain now looks as illustrated in Figure 5.

Figure 5 The development of a production plan with a particular focus on service intention enables the integration of processes, objectives and scopes. The time horizons of the three planning levels are the same as those shown in Figure 2.

Planning level	Long term planning (Service offer and resources)	Mid-term planning (Time table)	Short term planning (Operations and dispatching)
Time horizon	20 to 5 years	5 years to 24 hours	24 to 0 hours

Process



Level of detail

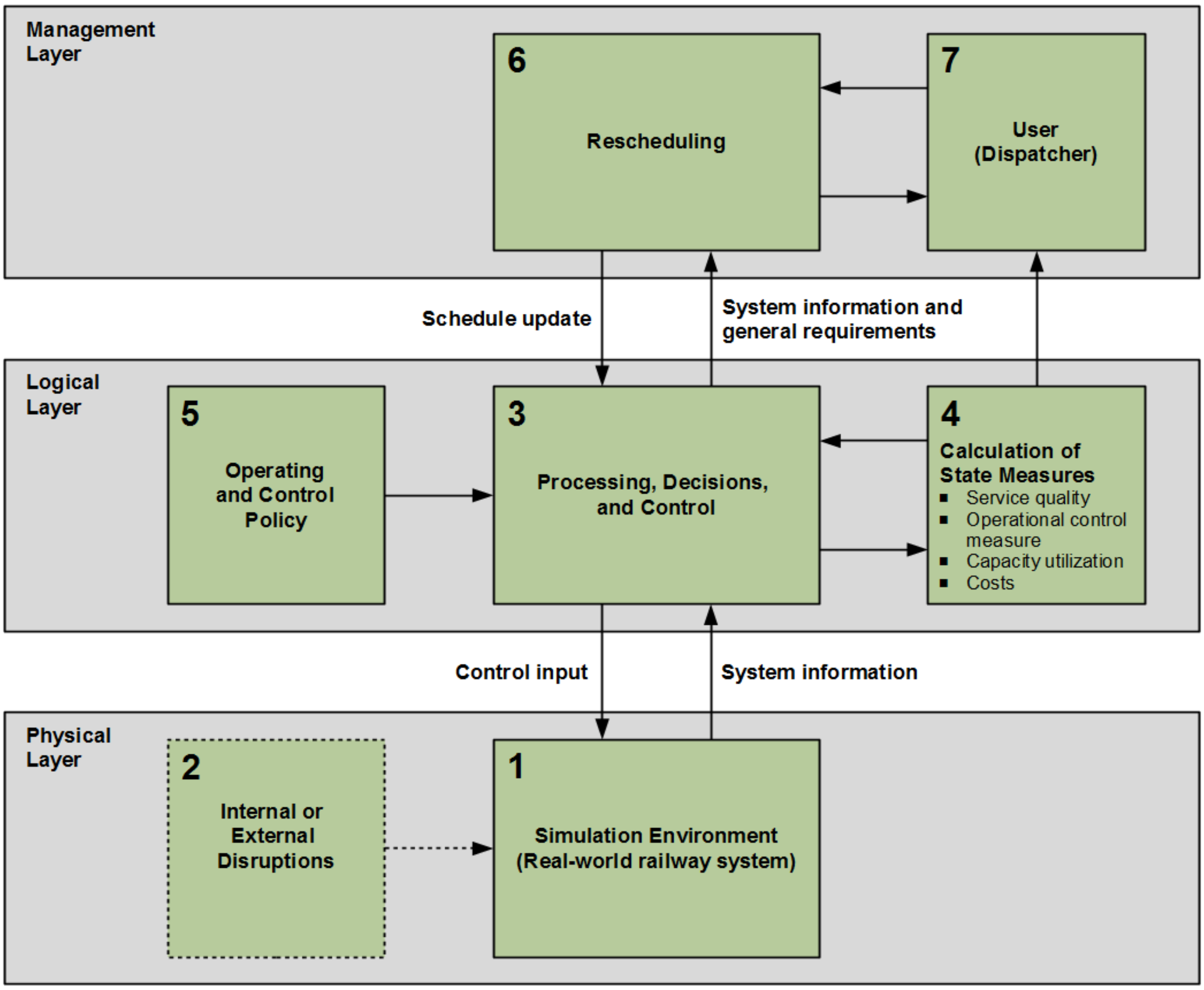
all levels of detail at each planning level

3.2 System overview

3.2.1 General framework

To address the problems outlined in chapter 1 and to incorporate the fundamental conceptual elements described in detail in chapter 2, we developed the general framework shown in Figure 6. The system consists of three layers, each with dedicated functionalities and each with one or more building blocks (labelled by numbers 1 to 7). In this section we keep the explanations pretty general, as detailed information on the building blocks will be provided in the next sub-section. Starting at the bottom, the *Physical Layer* consists of the operational environment (1), represented either by the simulation environment (in our case the railway simulation tool OpenTrack) or by the real-world system.

Figure 6 The overall system of the RTSE consists of three layers and seven building blocks.



The overall system as depicted in Figure 6 consists of four main functional components (blocks 1 to 4), a block capturing the operating policy (5), and a block representing any sort of disruptions (6). In the following we describe each building block and their interrelations.

As outlined in chapter 1, the goal of this project is, amongst others, to improve the stability of the railway system under normal conditions as well as after severe disruptions. In other words, the resilience of the system shall be improved in case of exceptional events. The system under investigation (1) comprehends of all physical and operational components of a railway system, i.e. rail infrastructure (network, railway fleets, etc.) and timetable (schedule) together with other information required for reliable operation. To allow for assessing (i) effects of disruptions on the stability of the railway system and (ii) of strategies to improve the resilience of the system, a specific block (2) is included to generate such causal events.

These events may be caused internally (e.g. hauling engine failure or railway control centre disruption) or externally (e.g. meteorological perturbations or severe disturbances in other transport systems leading to substantially higher rail traffic demand and hence higher utilization of the railway system).

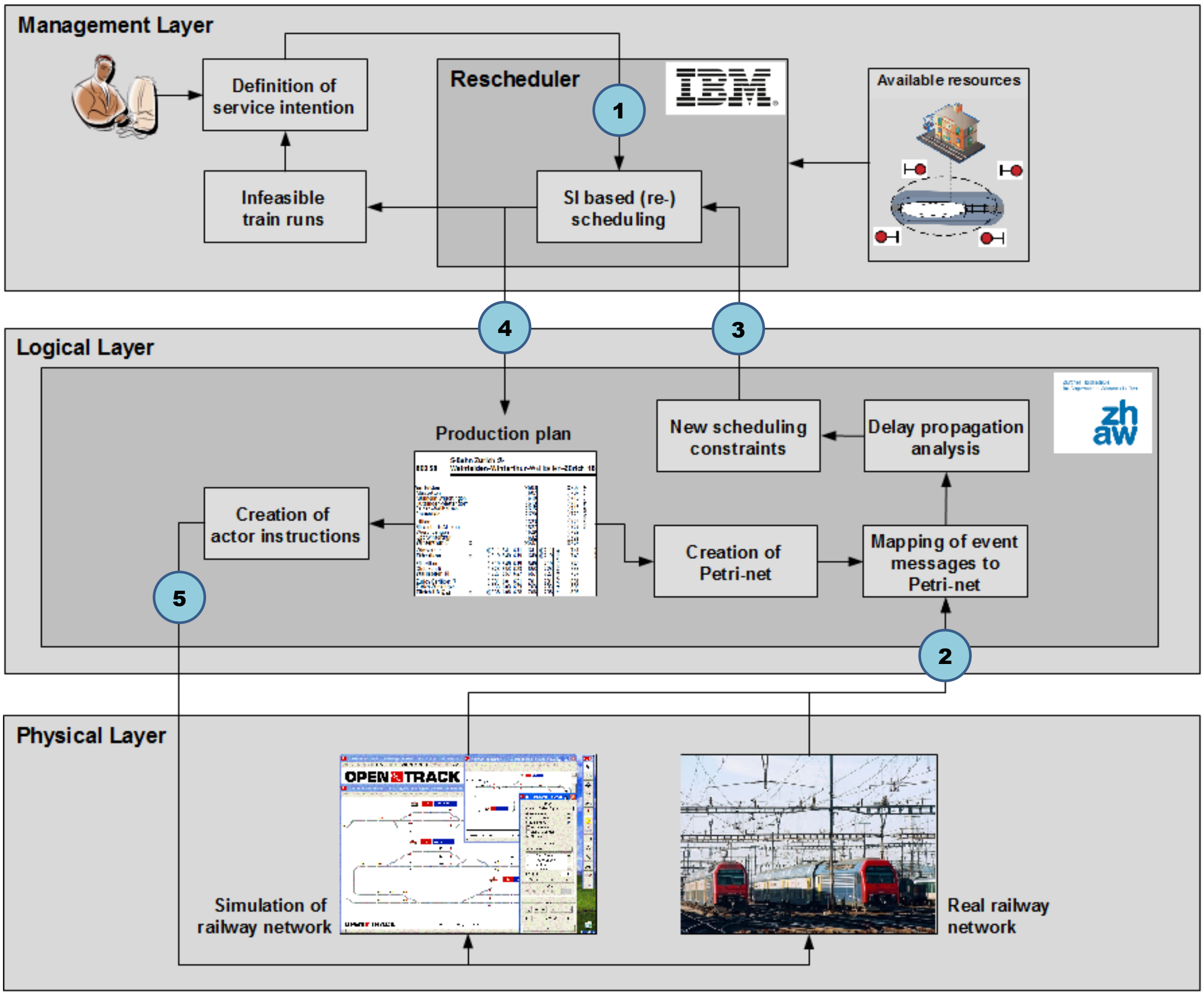
The *Logical Layer* in the middle consists of three blocks (3, 4 and 5): Block (3) is the central component of the system with the following main functions: (i) Reading information provided by the physical layer (e.g. train messages), the block calculating the state measures (4), and the block providing the operating and control policy (5); (ii) Determine, if a rescheduling is required, given the operating policy and the state of the system and start reschedule (6), if required; (iii) Read schedule information and determine new control values; (iv) send new control input to the operational environment (1), e.g. new speed profiles for engine drivers or new departure time are provided to passengers waiting on a particular platform. Finally, block (4) estimates the state of the system, which provides the basis for rescheduling decisions.

Finally, on top we have the *Management Layer*, which comprehends of the rescheduling algorithm (6) and an interface for the dispatcher (7) to allow for direct interaction. Details on the idea of the rescheduling algorithm can be found in the next section.

3.2.2 System with functional components as developed in this project

Based on the discussion of the structure of the overall system provided above, we now have a closer look on the functional components as used within the RTSE-framework. The simulation environment (which is part of the *Physical Layer*) consists either of the real-world railway network or the simulated railway network. It is important to note, that the interfaces of the *Logical Layer* (interfaces number 2, 3 and 4) are standardized. On the one hand, this ensures that conclusions, which can be drawn from the simulation environment are transferable to real world conditions. On the other hand, different rescheduling modules (*Management Layer*) can be used without changing anything in the *Logical Layer*.

Figure 7 The functional components of the system, the feedback loop and the corresponding interfaces (interface numbers are shown in blue circles): (1) service intention, (2) position and process state messages, (3) new scheduling constraints related to the resources available, (4) production plan and unfeasible train runs (if any), (5) timetable and actor instructions for real-time process configuration of the railway network.

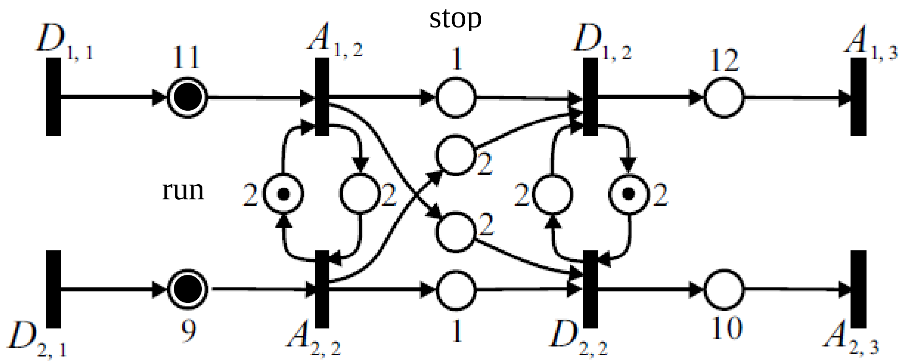


Among the functional components there is the role of the dispatcher. He represents the highest decision level in the *Management Layer*. The dispatcher's task is to manage the functional requirements, i.e. the SI. The SI is entered into the *Management Layer* component rescheduler (interface number 1) as a basis for the calculation of the normal production plan. In case of an operational rescheduling requirement, which cannot satisfy original SI-requirements, the dispatcher has to relax the SI such that the rescheduler can find a feasible solution for a new temporary production plan (interface 4). This interactive process will eventually need several iterations. After actor specific formatting, the production plan is used by the *Logical Layer* for (re-)configuration of the railway network. During the run time of the railway network there are position and process state messages (interface 2) sent permanently to the operation control part of the *Management Layer*.

The two main tasks of the *Logical Layer* are to first translate the production plan into actor instructions (descending branch of the feedback loop) and secondly to monitor the actual process state with respect to the current production plan. Actual process states are allowed to deviate from the intended process states (according to the production plan) within a certain tolerance range. As soon as a production plan is made available by the *Management Layer*, it is translated into a corresponding process model.

Figure 8 Example of a DEDS (Discrete Event Dynamic System) description that comprehends two train runs, each with a stop at a station and with transferring passengers (from Goverde, 2005). In addition both trains use the same switch for entering and exiting the station. (a) The model illustrated as a Petri net graph. Numbers indicate process durations. The D- and A-coefficients at transitions indicate timed events according to the production plan. In (b) to (e) the model constraints are expressed as inequalities, which can be evaluated to extract the delay propagation.

(a)



(b) Constraints of train runs

$$A_{1,2}(k) \geq D_{1,1}(k) + 11$$
$$D_{1,2}(k) \geq A_{1,2}(k) + 1$$
$$A_{1,3}(k) \geq D_{1,2}(k) + 12$$
$$A_{2,2}(k) \geq D_{1,2}(k) + 9$$
$$D_{2,2}(k) \geq A_{2,2}(k) + 1$$
$$A_{2,3}(k) \geq D_{2,2}(k) + 10$$

(c) Constraints of Connections

$$D_{1,2}(k) \geq A_{2,2}(k) + 2$$
$$D_{2,2}(k) \geq A_{1,2}(k) + 2$$

(d) Constraints of Headways

$$A_{1,2}(k) \geq A_{2,2}(k-1) + 2$$
$$A_{2,2}(k) \geq A_{1,2}(k) + 2$$
$$D_{2,2}(k) \geq D_{2,2}(k) + 2$$
$$D_{2,2}(k) \geq D_{1,2}(k-1) + 2$$

(e) Translate constraints to event domain nonlinear equation system of the DEDS example

$$A_{1,2}(k) = \max(D_{1,1}(k) + 11, A_{2,2}(k-1) + 2)$$
$$D_{1,2}(k) = \max(A_{1,2}(k) + 1, A_{2,2}(k) + 2, D_{2,2}(k) + 2)$$
$$A_{1,3}(k) = D_{1,2}(k) + 12$$
$$A_{2,2}(k) = \max(D_{1,2}(k) + 9, A_{1,2}(k) + 2)$$
$$D_{2,2}(k) = \max(A_{2,2}(k) + 1, A_{1,2}(k) + 2, D_{1,2}(k-1) + 2$$
$$A_{2,3}(k) = D_{2,2}(k) + 10$$

The process model is formally represented as a discrete event dynamic system (DEDS) as proposed by Goverde (2005) and Goverde (2010). The formulation of the process constraints given by train running times, resource allocations and passenger transfers is illustrated in Figure 8, where a description of a small example is provided using the DEDS formalism.

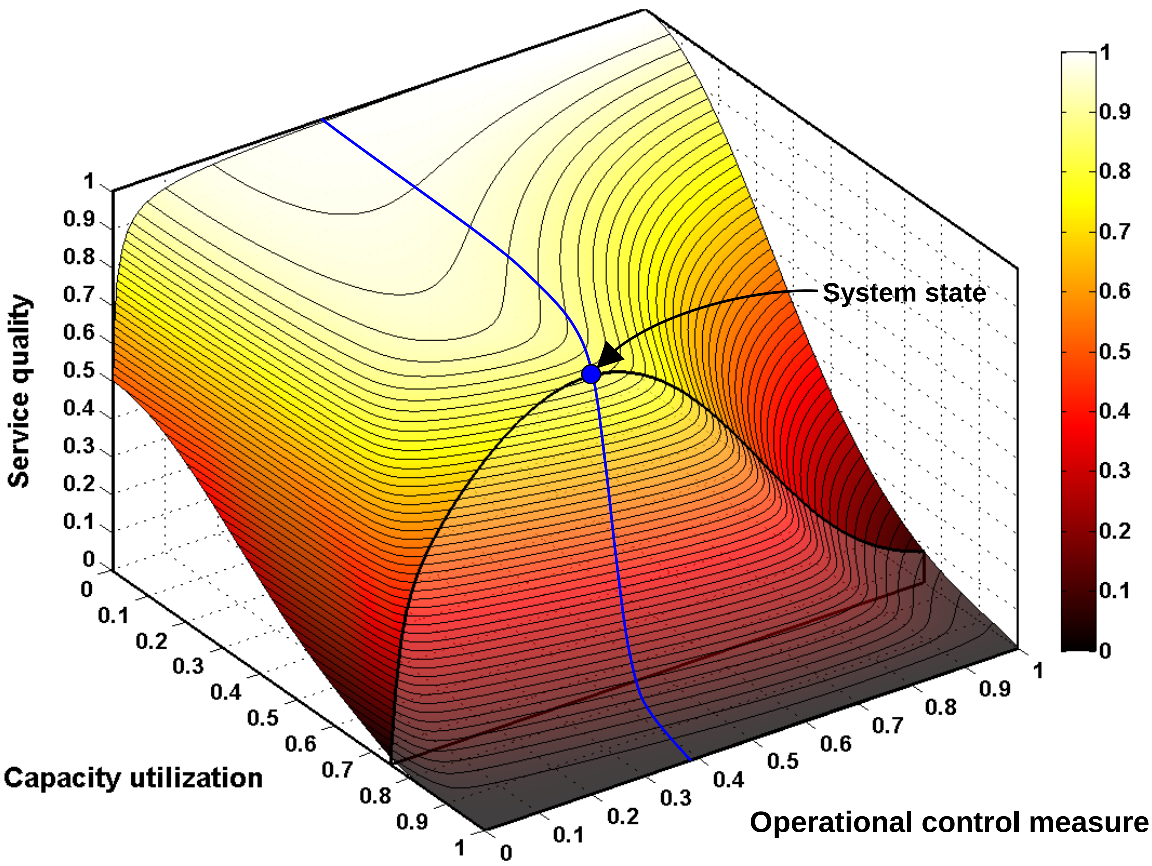
The Max-Plus linear equations of the DEDS are used to calculate delay propagation (Goverde, 2010) in case of exceeding deviation thresholds. With the help of this delay propagation, the spatio-temporal range of train runs that should be considered for rescheduling is identified. In the last functional step the related resource constraints are extracted from the process model and are supplied to the rescheduler (interface 3). Using this input, the rescheduler calculates a new production plan and with this step the control loop is closed.

3.3 State estimation and performance measurement

3.3.1 State estimation

In section 3.2.2 we have seen that one central component to decide if a rescheduling is required is to have a reliable estimate of the current state of the system. Hence, the development of measures that provide these estimates is a crucial task. In order to capture all relevant aspects of the system, we defined the following three components to characterize the state of the system on a network level: (i) the utilization of the capacity available (conflict between supply and demand), (ii) a measure that quantifies the type/amount of operational control, and (iii) the quality of the service provided to the travellers.

Figure 9 Example of the dependency of service quality from capacity utilization and operational control. The capacity utilization reflects the competition between supply and demand. The operational control is represented by a corresponding measure, defined in a range of zero (fully decentralized control; no coordination) to one (full central, real-time control). The slice along the y-axis (operational control measure) represents the theoretical case where the utilization of the system is constant while only the operational control varies. The maximum service quality in this case is around 0.85 and denoted by the blue point, i.e. the point where the thick blue line crosses the thick black line (with an operational control measure of around 0.38).



It is important to stress here, that we distinguish between the state on a microscopic (local) level and the one that describes the system on a macroscopic level (global, network-level). This distinction is justified since: (i) The microscopic view is appropriate for a fast detection of disruptions within the network, which allows for immediate decisions on any required action (one of which is performing a rescheduling that leads to an adjusted time table); (ii) The macroscopic level provides a global view on the system under investigation. Measures like the system performance are calculated based on macroscopic state information.

Figure 9 provides an example of the system characteristics in state space at a macroscopic level. The shape of the surface can be considered as being characteristic for a particular railway system as it directly reflects the impact of different control strategies and system utilizations on the service quality provided by the system. *The idea is that both the system utilization as well as the operational control affect the state of the system, which finally leads to some level of service quality.*

Each axis represents one of the three components mentioned above. Hence, the state of the system at time t is defined by $\mathbf{x}(t)=\left[u(t),c(t),q(t)\right]^{\mathrm{T}}$, where $u(t)$, $c(t)$ and $q(t)$ are random variables and denote the capacity utilization, the operational control measure and the service quality, respectively, and $[\cdots]^{\mathrm{T}}$ denotes the transpose of a vector or matrix. Unless otherwise stated, in this section time t is used as a discrete variable, with an interval between two successive steps Δt. In the following, we discuss some approaches to estimate these three variables.

However, as the project started quite recently and refinements are likely to be required, the methods presented below shall be considered as preliminary. This holds in particular for the derivation of the capacity utilization and the operational control measure, both on a network level.

Capacity utilization

The capacity utilization in railway systems is usually calculated for a single line (see UIC, 2004; Landex et al., 2006 or Höllmüller and Klahn, 2005). The standard document provided by UIC (UIC, 2004) allows for a good estimation of the capacity in various cases. However, when having a macroscopic view on the system, that is, if the capacity of a whole network needs to be estimated, a new approach extending the single-line case needs to be developed to finally get an estimate for $u(t)$. This will be part of the next project steps, where we also consider recent approaches, like for example discussed in Abril et al. (2008).

Operational control measure

The operational control, as discussed above, can be anywhere in the range of fully uncoordinated to fully coordinated (in real-time). As running a railway system involves various players (actors) with a multitude of interdependencies, we model the network of players and their interaction with methods from complexity theory. Good introductions can be found for example in Barrat et al. (2008), Albert and Barabási (2002) and Barthélemy (2011).

Various measures are available to characterize the structure and of such networks (e.g. centrality, betweenness, clustering, degree distribution or average path length, to mention just a few) together with their dynamical processes. Many transport networks have been studied over the last years (see e.g. Porta and Scheurer, 2006; Chen et al., 2009; Shi et al., 2009), ranging from air traffic to road networks or describing the connectivity of large transport hubs.

The evaluation of optimal measures to capture the interconnectivity within the control network together with their practical implementation will be performed later on in the project.

Service quality

A very common measure to quantify the service quality is to calculate the punctuality at different nodes in the system (see e.g. Vromans, 2005). Although easy to calculate, using punctuality has some drawbacks. The most important one is that punctuality does only provide information (together with a distribution) on the frequency of deviations of arrivals and departures relative to the time table. Furthermore, it does not take into account the number of passengers affected by delays nor does it cover the case where passengers are missing connecting trains.

The following formula overcomes the above drawbacks, as it accounts for passenger delays (unit: passenger times minutes) instead of only delays (unit: minutes). Moreover, both local as well as transfer passengers affected by disruptions are considered:

$$d(t) \equiv d\left(t,T \middle| s,u(t),c(t)\right) = \sum_{k\in K} \sum_{i\in I_k(t,T)} \left\{ \begin{aligned} & \max\left[t_{ki}^{\text{arr}}(s) - \bar{t}_{ki}^{\text{arr}}(s), 0 \right] F_i \\ & + \sum_{j\in J_{ki}(t,T)} H\left[t_{ki}^{\text{arr}}(s) - t_{kj}^{\text{dep}}(s) + \omega_k^- \right] W_{ij'} F_{ij} \end{aligned} \right\}, \quad (3)$$

where k and K represent a particular node and the set of all nodes considered, respectively, T defines the time interval considered, $i \in I_k(t,T)$ denotes the set of trains accounted for at node k during period $[t,t+T)$, s represents the number of the scenario investigated, $t_{ki}^{\text{arr}}(s)$ and $\bar{t}_{ki}^{\text{arr}}(s)$ denote the actual and planned arrival of train i at node k, $t_{ki}^{\text{dep}}(s)$ and $\bar{t}_{ki}^{\text{dep}}(s)$ denote the actual and planned departure time, respectively, F_i denotes the number of passengers (estimated form empirical data) from train i with destination at node k, F_{ij} is the number of passengers transferring to connecting train j (again estimated form empirical data), with $J_{ki}(t,T)$ being the set of all planned connecting trains for train i, again during period $[t,t+T)$, ω_k^- is the minimum connecting time for railway station k, $H(x)$ denotes the Heaviside function, with $H(x) = \{0 : x < 0;\ 1 : x \geq 0\}$, and $W_{ij'} \equiv W_{ij'}(s) = \bar{t}_{kj'}^{\text{dep}}(s) - \bar{t}_{kj}^{\text{dep}}(s)$ finally is the waiting time to catch the next train j' to the same destination as train j.

It is important to note that $d(t)$ is a random variable, as the measured arrival and departure times, i.e. $t_{ki}^{\text{arr}}(s)$ and $t_{ki}^{\text{dep}}(s)$, and passenger flows F_i and F_{ij}, are random variables as well.

Based on cumulative delay $d(t)$ (unit: passenger times minutes) we next define some requirements for the derivation of the measure for service quality, i.e. $q(t)$. A requirement for the functional dependency of $q(t)$ from $d(t)$ is that it shall be monotonically decreasing, i.e. the higher the delay, the lower the service quality.

Furthermore, we know that the higher the utilization of a system, the higher the average delay (see Figure 9: once the utilization approaches a value of one, the service quality tends to zero, e.g. blue line).

For the moment, we define the service quality at time t with time interval T (see above), assuming that the system is in a certain state defined by scenario s and the two random variables $u(t)$ and $c(t)$, both introduced above, in a quite general way as:

$$q(t) \equiv q\left(t,T \middle| s,u(t),c(t)\right) = f\left(d\left(t,T \middle| s,u(t),c(t)\right),\underline{\theta_d}\right), \tag{4}$$

where $\underline{\theta_d}$ denotes a vector with parameters required to determine function $f(\cdots)$. These parameters together with the functional relationship $q(t) = f\left(d(t)\right)$ will be explored in detail in the next steps of the project. Again, $q(t)$ does not have a direct analytical dependency on $u(t)$ and $c(t)$ but rather represents changes in the systems state caused by these two variables.

3.3.2 Performance measurement

The performance measurement in railway networks is by far not a simple task (see for example Rietveld et al., 2001, Vromans et al., 2005 or Lin, 2012).

Once we will have determined $q(t)$, i.e. the measure representing the service quality, a simple yet reasonable approach to derive the average performance of the system within a certain time period Γ (i.e. within a set of subsequent discrete time steps), is as follows:

$$\begin{aligned} p(t,\Gamma) &= 1 - \frac{1}{|\Gamma|}\sum_{t' \in \Gamma}\left(1 - q(t')\right) \\ &= \frac{1}{|\Gamma|}\sum_{t' \in \Gamma} q(t') \end{aligned} \quad , \tag{5}$$

where $|\cdot|$ denotes the cardinality of a set. However, further approaches to determine the overall performance of the system will be evaluated.

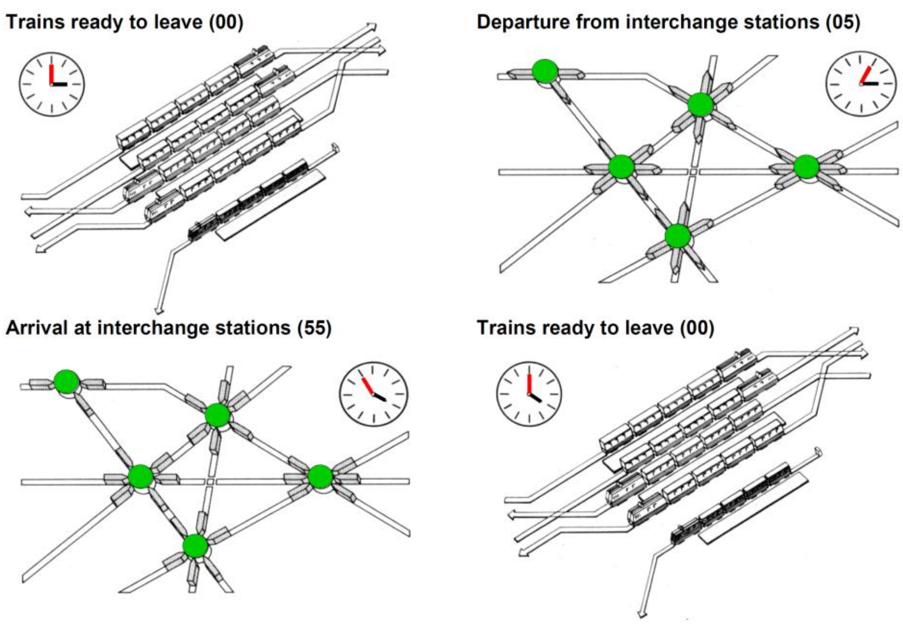
3.4 Common practice to secure service quality

An alternative measure commonly used for qualifying public transport services is timetable robustness. Timetable robustness describes the ability of a timetable to absorb small operational disturbances (van den Boom, 2011). Because of the high timetable density in Swiss rail traffic, there is not much buffer time for absorbing operational irregularities. So robustness is quite low and it doesn't really tell us, what can be done operationally. A consequence might be to eliminate critical trains from the timetable in order to increase timetable robustness. This has been successfully applied in the Dutch railway network (Goverde, 2002). Another measure is timetable stability, which describes rather the operating company's ability to compensate for disturbances and relates to the time period needed for return to the normal schedule based operations after the occurrence of an event (Goverde, 2007). Hence, the quick restoration of the regular timetable in case of disruptions, which is an extremely difficult task in itself, has been traditionally the common objective for dispatching. On the other hand, disturbances cause rather quick delay propagations within the network and it usually takes a considerable amount of time to return to the normal timetable situation. During such a time period of reduced service due to troubleshooting a lot can be done however, to keep the supplied service as good as possible (i.e., the processes could actually be improved significantly). *This is why we suggest to use the network based overall passenger delay as a quantitative estimate of the service quality of the system rather than measuring timetable robustness*.

Instead of putting the focus on the timetable, we suggest to investigate the way a given train traffic network is operated. In other words, we focus on examining and quantifying the influence of the operation process control with the goal to secure a high and stable service quality. The way a train company operates its traffic network varies strongly depending on the nature of the schedule on the one side and the network structure on the other. If a timetable has a strong periodicity, operational staff is used to highly repetitive procedures and rather periodic event occurrences and is trained to make appropriate dispatching decisions (Stohler, 2003; Stalder, 2006). If the network structure is characterized by rather independent lines, operated by different TOC's and each operating at high frequencies like for instance in Japanese or Dutch railway systems, there are typically not so many transport connections published and hence there has not much operational effort to be done in taking care, that train line interconnections are ensured. Train routes are firmly assigned to certain trains and dispatching focuses on the decision, whether to change the train sequence of route utilization or not. As individual lines are operated at high frequencies, like in the Japanese rail traffic network, this approach can provide a rich transport service as well.

On the other side rail network operation in Switzerland has to secure multiple connections at major hubs, as illustrated in Figure 10. Here train runs have to be synchronized, which requires the control of running times on the one hand and flexible route assignment in major shunting areas on the other hand.

Figure 10 Clock-face timetables require synchronisation of trains in hubs. In this way multiple connections can be guaranteed. (Figure from Stohler, 2003.)



As a consequence the different schedule and network structures have established different types of communication between central dispatching staff and local operation staff. In general one can say that if there is less interdependence between single train runs there is less intense communication required between central network control and local operating staff. If, however, there are strong regular patterns, which depend on multiple operational process cycles (maintaining connections between train lines, rolling stock and staff circulations) there is efficient decision making and communication required for maintaining service quality under conditions of operational disturbances.

Our framework in combination with the general state model illustrated in Figure 9, is designed for comparing the different schedule and traffic network conditions described in this chapter.

Our hypothesis is that for low capacity utilizations, the service quality is rather high, even if the operational control is low. However, under these operational conditions, the service quality rapidly drops, if one increases the capacity utilization. In situations of high capacity utilization, the structure and hence the resulting dynamics of the system play an important role as many events happen simultaneously. Performance might rise substantially with increasing capabilities to speed up operational control. If, however, the rate of disruptive events exceeds a certain level, we assume that the performance will drop. We assume that for dense network conditions it will be better to give local operational actors more freedom to manage their local situation and act in a self-responsible manner. Of course, under ideal conditions, this tolerance range of self-responsible acting will be limited and in case of exceeding thresholds the operating centre needs to be informed to initiate control actions. In the future we will further investigate the influence of the operational process control on service quality under various conditions of capacity utilization.

4. Pilot application

Denmark is currently re-implementing its entire rail road infrastructure. The main objective is to replace the existing safety system, which is based on optical signalling by the modern European standard safety system ETCS (European Train Control System). The responsible rail service provider BDK (Banedanmark) uses this opportunity for redesigning all operational processes and for installing a leading edge train control system. In addition to the purpose of reducing operational costs, the investment shall enable BDK to offer improved train services with more connections and higher train frequencies.

RTSE as benchmarking framework for Banedanmark (BDK) train control system offers

In order to achieve these goals, BDK is redesigning operation processes and has launched calls for tenders for a new train control system. Swiss rail engineering companies support BDK in defining the functional requirements (see Banedanmark, 2009). Therefore, some of these companies (with the contributions of ZHAW) are currently developing a simulation framework that allows for investigating effects of the intended new operation processes on system performance and service quality. BDK staff and other project members create and investigate certain test cases create benchmarks and evaluate system offers. The RTSE framework is designed to facilitate these tasks. Therefore the following use cases have been identified together with the consulting companies for using RTSE as pilot application:

- Quantitative performance measurement of scheduling systems.
- Quantitative performance measurement of new operation paradigm (time controlled operation processes in contrast to event controlled operation processes).
- System calibration of train departure and speed control.
- Network effects of real time rescheduling.

Developing this pilot application is one of the primary goals of this CTI-funded development project.

5. Conclusions and outlook

In this paper we proposed an integrated scheduling and dispatching approach in order to enable TOC's and IOC's to investigate, which technical and operational measures can be taken in order to optimize the system's performance (e.g. minimize total train and/or accumulated passenger delay) at a network level. Therefore we propose to use the concept of service intention (SI) as functional requirement for rescheduling in case of operational disturbances. In order to be able to execute SI-based rescheduling we introduced the RTSE framework for simulating a network-wide operational control, which integrates (i) train simulation, (ii) process modelling and (iii) rescheduling in one system. The proposed modular approach allows for exchanging single components in order to achieve independence from specific system providers.

The modelling approach is specifically designed to investigate the key relationship of service quality, capacity utilization and the operational control. In order to allow TOC's and IOC's for implementing and monitoring their key processes, investments in new processes and technologies need to be made. Therefore it is our aim, to provide a powerful and meaningful simulation framework that allows for comparing costs and benefits of different types of operational control before they are implemented in a real-world environment.

One of the next steps will be to further refine the definition of the measures as discussed in section 3.3.1. It is our primary goal to further improve the proposed framework such that it optimally supports operating staff and performance managers. Further activities will include the development, implementation and testing of the rescheduling algorithm, as well as the overall software framework, which combines all components discussed above.

The proposed modelling approach and system design is developed as part of an on-going joint research collaboration between ZHAW, OpenTrack Railway Technology GmbH, IBM Research GmbH, and Emch+Berger AG.

Further results shall be presented at the IT13.rail International Railway Conference 2013 in Zurich.

Acknowledgements

We would like to thank our project partners, Edi Mumprecht and Jonas Looser (InIT, ZHAW, Winterthur), Martin Hess and Bernhard Seybold (IDP, ZHAW, Winterthur), Marco Laumanns (IBM Research GmbH, Rüschlikon), Daniel Hürlimann (OpenTrack Railway Technology GmbH, Zürich), and Samuel Roos (Emch+Berger AG, Bern) for the inspiring and fruitful collaboration. Furthermore we would like to thank the CTI for financial support of this project.

References

Abril, M., Barber, F., Ingolotti, L., Salido, M.A., Tormos, P. and Lova, A. (2008) An Assessment of Railway Capacity. *Transportation Research Part E: Logistics and Transportation Review*, vol. 44, issue 5, pp. 774–806.

Albert, R. and Barabási, A.-L. (2002) Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, vol. 74, issue 1, pp. 47-97.

Banedanmark (2009) *The Banedanmark Signalling Programme*. For details see BDK article http://uk.bane.dk/visArtikel_eng.asp?artikelID=6090, Accessed April 28, 2012.

Barrat, A., Barthélemy, M. and Vespignani, A. (2009) *Dynamical Processes on Complex Networks*. Cambridge University Press, Cambridge, UK.

Barthélemy, M. (2011) Spatial networks. *Physics Reports*, vol. 499, issues 1-3, pp. 1-101.

BAV (2007) *Faktenblatt ETCS (European Train Control System)*, Bundesamt für Verkehr, http://www.bav.admin.ch/dokumentation/publikationen/00475/00476/00728/, Accessed April 28, 2012.

Burkolter, D. (2005) *Capacity of railways in station areas using Petri Nets*. PhD Thesis, Institute for Operations Research, ETH Zurich, Zurich.

Caimi, G.C. (2009) *Algorithmic decision support for train scheduling in a large and highly utilised railway network.* PhD Thesis, Institute for Operations Research, ETH Zurich, Zurich.

Chen, J., Liu, J. and Jiang, Z. (2009) Reliability of Urban Rail Networks Based on the Network Connectivity, In: *Proceedings of the 2nd International Conference on Transportation Engineering 2009*. ASCE Conf. Proc. (doi:http://dx.doi.org/10.1061/41039(345)333)

Fitzsimmons, J.A. and Fitzsimmons, M.J. (2008) *Service Management – Operations, Strategy, Information Technology,* 6th edition. McGraw-Hill.

Fuchsberger, M. (2012) *Algorithms for adaptive train control in main stations*. PhD Thesis, Institute for Operations Research, ETH Zurich, Zurich.

Goverde, R.M.P. and Odijk, M.A. (2002) Performance Evaluation of Network Timetables Using PETER, In: Allan et al. (eds.) *Computers in Railways VIII*, pp. 731–740.

Goverde, R.M.P. (2005) *Punctuality of Railway Operations and Timetable Stability Analysis,* PhD thesis, Technical University of Delft, Delft.

Goverde, R.M.P. (2007) Railway timetable stability analysis using max-plus system theory. *Transportation Research Part B: Methodological*, vol. 41, issue 2, pp. 179-201

Goverde, R.M.P. (2010) A delay propagation algorithm for large-scale railway traffic networks. *Transportation Research Part C: Emerging Technologies*, vol. 18, issue 3, pp. 269–287.

Höllmüller, J. and Klahn, V. (2005) Implementation of the UIC 406 capacity calculation at Austrian railways (ÖBB), In: I.A. Hansen, F.M. Dekking, R.M.P. Goverde, B. Hindergott and L.E. Meester (eds.) *Proceedings of the 1st International Seminar on Railway Operations Modelling and Analysis*.

Landex, A., Kaas, A. H., Schittenhelm, B. and Schneider-Tilli, J. (2006) Practical use of the UIC 406 capacity leaflet by including timetable tools in the investigations, In: J. Allen et al. (eds.) *10th International Conference on Computers in Railways COMPRAIL X*, pp. 643–652.

Lin, Y.-K. (2012) Performance evaluation for a transportation system in stochastic case.

Lüthi, M., Nash, A., Weidmann, U., Laube, F. and Wüst, R. (2007) Increasing Railway Capacity and Reliability through Integrated Real-Time Rescheduling. In: *Proceedings of the 11th World Conference on Transport Research, Berkeley*.

Lüthi, M. (2008) Optimizing traffic flow in heavily used railway networks: influence factors and potential strategies. Paper presented at the *8th Swiss Transport Research Conference,* Monte Verità, Ascona, October 15-17, 2008.

Lüthi, M. (2009) *Improving the Efficiency of Heavily Used Railway Networks through Integrated Real-Time Rescheduling*. PhD Thesis, Institute for Transport Planning and Systems, ETH Zurich.

OpenTrack (2012) OpenTrack Railway Technology GmbH: http://www.opentrack.ch/, Accessed April 29, 2012.

Porta, S. and Scheurer, J. (2006) Centrality and connectivity in public transport networks and their significance for transport sustainability in cities, In: *World Planning Schools Congress, Global Planning Association Education Network,* 13-16 July 2006, Mexico.

Rietveld, P., Bruinsma, F.R. and van Vuuren, D.J. (2001) Coping with unreliability in public transport chains: a case study for the Netherlands. *Transportation Research Part A: Policy and Practice*, vol. 35, issue 6, pp. 539–559.

Roos, S. (2006) *Bewertung von Knotenmanagement-Methoden für Eisenbahnen*. Master’s thesis, Institute for Transport Planning and Systems, ETH Zurich.

Shi, Y., Lu, H., Nie, C. and Yoshitsugu, H. (2009) Analysis of Public Transport Networks in Nagoya Based on Complex Network Theory, In: R. Liu, J. Zhang and Ch. Guan (eds.) *Logistics: The Emerging Frontiers of Transportation and Development in China. Proceedings of the 8th International Conference of Chinese Logistics and Transportation Professionals (ICCLTP)*, Chengdu, China from October 8-10, 2008.

Stalder, O. (2006) Competitive advantages gained by redesigning processes: The case of Rail 2000 and beyond, In: *Proceedings of the World Congress on Railway Research*, Montreal, Canada.

Stohler, W. (2003) *Why is an integrated clockface- driven railway system more efficient than a divided competition-oriented system?* For details see homepage of SMA: http://www.sma-partner.ch > Veröffentlichungen > Publikationen, Accessed April 29, 2012.

UIC (2004) Capacity, 406 R, 1st edition, Union Internationale des Chemines de fer UIC, June 2004.

van den Boom, T.J.J., Weiss, N., Leune, W., Goverde, R.M.P. and de Schutter, B. (2011) A permutation-based algorithm to optimally reschedule trains in a railway traffic network, In: *Proceedings of the 18th IFAC World Congress*, Milano, Italy, pp. 9537–9542

Vromans, M. (2005) *Reliability of Railway Systems*. PhD Thesis Erasmus University of Rotterdam at Research Institute of Management

Vromans M.J.C.M., Dekker R. and Kroon, L.G. (2006) Reliability and Heterogeneity of Railway Services. *European Journal of Operational Research*, vol. 172, issue 2, pp. 647–665.

Wegele, S., Slovák, R. and Schnieder, E. (2007) Real-time decision support for optimal dispatching of train operation, In: I.A. Hansen, A. Radtke, J. Pachl and E. Wendler (eds.) *Proceedings of the 2nd International Seminar on Railway Operations Modelling and Analysis (RailHannover 2007)*, Hannover, Germany.

Wüst, R., Laube, F., Roos S., Caimi G. (2008) Sustainable Global Service Intention as objective for Controlling Railway Network Operations in Real Time, In: *8th World Congress of Railway Research (WCRR)*. Seoul.

ZVV (2010) *Strategie 2012–2016: Grundsätze über die Entwicklung von Angebot und Tarif im öffentlichen Personenverkehr. Erläuternder Bericht*. Zürcher Verkehrsverbund, Zürich, Juli 2010.