Optimizing traffic flow in heavily used railway networks: influence factors and potential strategies

Marco Lüthi, ETH Zürich

Conference paper STRC 2008

STRC

8th Swiss Transport Research Conference
Monte Verità / Ascona, October 15-17, 2008

Optimizing traffic flow in heavily used railway networks: influence factors and potential strategies

Marco Lüthi
Institute for Transport Planning and Systems
ETH Zürich
Switzerland

Phone: +41 44 633 24 15
Fax: +41 44 633 10 57
email: Luethi@ivt.baug.ethz.ch

October 2008

Abstract

Rail travel demand in Switzerland has grown strongly since introduction of the "Bahn 2000" program and growth is projected to continue in the coming years. As part of the "Bahn 2000" program many routes are operated every 30-minutes in a network-wide regular interval timetable. This integrated clock-face timetable provides an optimal nationwide timed transfer system resulting in high accessibility and relatively short passenger travel times. However, it also means that many trains arrive at and depart from stations in a short interval of time. As the number of trains increases, the number of potential conflicts also increases; this means that a small initial disturbance at one location can significantly impact the entire network. Today many network locations are operating at their effective capacity, therefore adding more trains could reduce service quality.

The paper presents an overview of an ongoing research project designed to develop methods for increasing capacity and reliability for railway networks. These methods aim to strengthen the competitiveness of railway systems and supports rail as an important player in sustainable transportation. The proposed approach combines a real-time rescheduling traffic control system with new methods for precise production in an attempt to minimize schedule reserve times without reducing stability and thereby increases network capacity accompanied by reduced energy consumption. After illustrating the past development of rail modal spilt in Switzerland, the paper gives an overview of the new approach, discusses the rescheduling process and presents issues to be addressed in order to effectively use the new approach. In addition, the paper outlines how the network should be divided for planning and operating, and presents how the approach can minimize energy consumption.

Keywords

Rail traffic operation, Real-time rescheduling, Rail network capacity management

1. Introduction

Transport plays an important role in supporting economic, environmental and social objectives. Over the years, passengers and freights rail traffic has become less important whereas the amount of road traffic rapidly increased. This fact, combined with the rapid worldwide traffic growth, leads to a deteriorating environment and diminishing resources. Thereby, not only ecological and resource constraints are consequences of increased road traffic. Limited capacity, especially in dense urban areas causing traffic congestions, frequent and serious accidents or restricted quality of life because of noise and pollution are other examples. Accordingly, a conflict exists between the traffic participant's behavior and the requirements for a sustainable development. Public transportation and in particular rail traffic could play a key role to ensure sustainable mobility. Strengthening their position, railways must be more competitive, more efficient and deliver a safe, fast, reliable, cheap and comfortable service meeting the customer's needs.

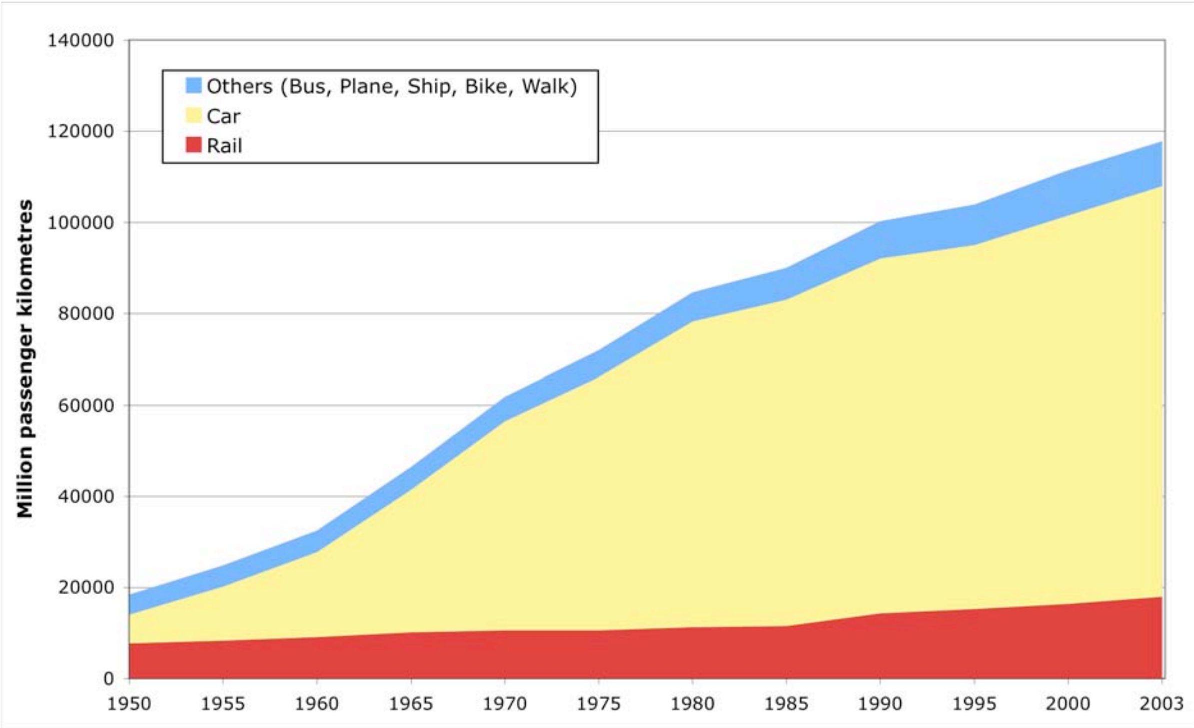
1.1 Development of mobility and transport in Switzerland within the last 50 years

The triumph of private transport over public and rail transport since 1950 in Switzerland is parallel to the development of income and wealth. Starting with a modal share of 52.1% in 1950, rail has lost market share dramatically until 1985 having only 13.2% of all passenger kilometers. During the same period, the modal share for private road transport increased up to the maximum of 81.7% in 1985 (see Figure 1). From that time on, modal share for public rail transport slowly increased up to 15.7% in 2003.

By contrast, the negative trend of the modal split for rail freight transport, starting with 69.7% for the sum of inland and transit transport in 1950, still sustains. In 2003, only 29.6% of the overall ton kilometers in Switzerland's freight were transported on railways. Thereby, the modal split decrease for rail freight ton kilometers is observed for both transit and inland traffic. In 2003, inland rail freight traffic had a modal spilt of only 17.7% whereas in international transit, rail had a modal split of 69.7%.

In 1982, Swiss Federal Railways (SBB) launched the cyclic clock-face timetable offering a nationwide sophisticated and harmonized service concept for passenger rail transport. So, passenger requirements were enhanced and the negative development of passenger rail transport could be stopped. The ongoing growth of the modal split since 1985, although it is rather limited, shows that public rail operators improved their performance and could accomplish passenger needs and demands.

Figure 1 The development of personal mobility in Switzerland since 1950. Source: Adapted from LITRA



Beside rail and road traffic, the other modes play a minor part for both freight and passenger traffic in Switzerland. Effectively, the amount of freight traffic by ship and air is in summation less than 1 percent. Since 1995, the modal share for passenger kilometers by airplanes is stable with about 2.5%. Interestingly, the modal split for public road transportation (busses) is roughly stable since 1960 and varies between 2.9 and 3.4%.

However, to improve the competitiveness of railways against road traffic in passenger and freight transport, a lot of challenges have to be solved, namely:

- Increase efficiency and reliability (quality) by reduced and controlled costs;
- Offer fast connections close to the market demand;
- Allow disposal of free slots within shortest time;
- Hold or increase the safety level;
- Handle the predicted growing demand;
- Respond to the developments in other transport modes by reducing emissions.

To address these challenges and especially to improve the railway network efficiency, a new and innovative framework is developed by the SBB in cooperation with the Swiss Federal Institute of Technology ETH. The project's objective is to identify methods for increasing capacity and stability of railway networks without making significant infrastructure investments. Thereby, the rescheduling and train control processes are central elements. The framework is called integrated real-time rescheduling since the infrastructure operator (who determines the new schedule after a delay or incident in real-time) cooperates with the train operator (who controls/runs the trains) in the transportation production process. In addition to improve capacity and stability, this new framework can also help to address demands for adding slots for freight operations within shortest time and minimize the energy consumption.

1.2 Paper organization

Section 2 of this paper gives an introduction in Switzerland's rail network and shows some fundamental coherences and recent trends to improve railway capacity. Section 3 describes the rescheduling process; a system underlying process that is subject to enormous changes from the today's manual, experience-based dispatching to an almost fully automated service. Section 4 describes different approaches to classify and manage railway networks. Section 5 describes the integrated real-time rescheduling framework. Finally, Section 6 presents conclusions and an outline of future research.

2. Railways in Switzerland and recent development trends

2.1 Switzerland's railway network

Switzerland has one of the worlds most heavily used national rail networks. During the 1990s, SBB built new infrastructure designed to increase capacity and improve service as part of the Bahn 2000 program. However, demand for rail service continues to increase and the SBB must develop new capacity while minimizing costs. One of the key strategies being investigated by the SBB for increasing capacity is using the infrastructure more efficiently using dynamic rescheduling.

Switzerland's passenger rail service is based on the concept of an integrated clock-face timetable (described by Maxwell 1999). This essentially consists of a timed-transfer system for the entire country. Using this approach, the Bahn 2000 program expanded this system to more cities and increased the number of operated trains.

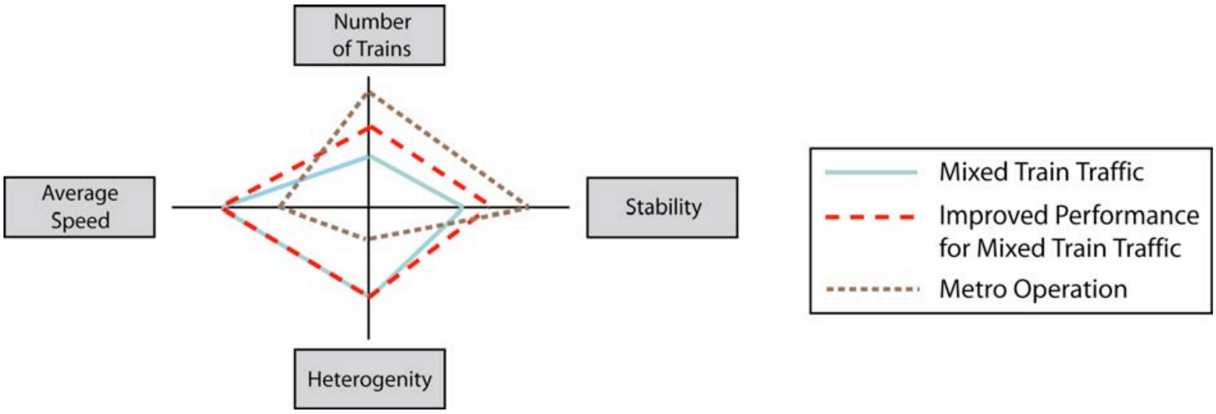
Initial results, presented by Ullius 2005 and in the SBB's Annual Report 2006, show that the Bahn 2000 program has been successful and that, even with the large amount of service operated, the rail network still satisfies the SBB's strict delay quality standard (96.2 % of all passenger trains had an arrival delay of 5 minutes or less in 2006). However, it is also clear that the network is operating at the edge of stability (up to 3000 train movements through one station area in a single regular day). In order to increase service and maintain high service quality new strategies for increasing capacity are needed.

Building new infrastructure is the most obvious possibility for increasing capacity, however this is expensive and, particularly in bottleneck and station areas, often no longer possible. Therefore, production-based strategies for increasing capacity must be developed (i.e. based on how trains are operated). These strategies essentially allow more trains to be operated on the same infrastructure.

2.2 Production-based strategies to increase rail network capacity

In order to increase the number of trains operated, the headway time between trains must be decreased. The headway between trains is determined by two components: blocking times and buffer times. The blocking time is a time interval in which a given track section is exclusively allocated to one train. This ensures the safety and prevents collisions between running trains. The schedule reliability component is designed to provide reserve (or buffer) time necessary to ensure that trains remain on schedule (i.e. it reduces the impact of delays on system-wide operations). Therefore, optimizing both stability and capacity together is not possible with common given methods. The relation capacity and stability is illustrated in Figure 2.

Figure 2 The capacity balance based on the UIC-Codex 406 (2004).



The lowest possible headway is determined based on the blocking time. There are many strategies for reducing the minimum headway between trains, for example with shorter block sections. Therewith, more signals would be needed. Another strategy is based on communicating "stop" or speed instructions to trains more quickly and independently of the train's position (e.g. moving block signals, or the new European Train Control System ETCS). Eichenberger presented in 2007 the strategy how ETCS can be used to increase capacity in detail. Shorter block sections and other developments require cost-intensive improvements to signaling systems and on-board equipment.

Railway operations depend on controlling trains to ensure safety and efficient operations. The term train control can be used to describe two different activities: the actual control of a particular train (this is normally done by the train driver/operator based on instructions from signal systems but can also be automated to various degrees) and the management of many trains operating on the network (this is done by developing schedules and operating plans). In this paper the term "train control" will be used to describe activities for an individual train and "traffic management" will be used to describe the higher-level control activities.

Increasing levels of train control and traffic management can provide improved safety and reliability in a railway network, thereby allowing headways to be reduced and capacity to be increased Revising railway schedules to reflect actual network status in real-time is an example of increasing traffic management. This research focuses on developing strategies for improving railway traffic management, but these strategies can only be implemented using train control on specific trains. Therefore both train control and traffic management are considered in the research.

There are three aspects of control (in the general sense): knowing what needs to be done, communicating what needs to be done and doing what needs to be done. Translated to the railway environment this means:

1. Developing a schedule that specifies where each train should be at all times;
2. Communicating the timetable to all affected parties (e.g. train operators, infrastructure operators); and
3. Operating trains and infrastructure according to the timetable.

Under normal conditions timetables are developed and communicated to involved parties well in advance and the train operators drive their trains accordingly. In these cases schedule planners have time to optimize timetables and use track capacity efficiently. However, when there is an incident or disturbance a new timetable must be developed in real-time, communicated to involved parties as quickly as possible and these parties must take appropriate actions immediately (which are often different from the planned timetable actions).

The process of developing new timetables and communicating them to all involved parties is complex and time consuming. This is why schedule planners add reserve time to timetables. Thereby, three types of reserves were distinguished:

- Reserves for the driving accuracy
- Reserves for the running times
- Operational Reserves (reserves between trains) to handle larger delays.

Therefore, reserve time allows both reducing the need for rescheduling and simplifying the process of rescheduling. However, reserve time wastes capacity. If it were possible to reduce the amount of reserve time network capacity could be increased without constructing new infrastructure. Research about optimizing the amount and distribution of reserves in railway networks is thus a key factor of success.

Once the new schedule has been developed and communicated to the involved parties, it must be implemented. This is the train control process. At its simplest level the train control process can simply mean that the train operator (locomotive driver) drives the train based on the new schedule; when more precise control is required, various levels of automatic control can be introduced.

2.3 Real-time rescheduling in complex railway networks

Rapid improvements in information and communications technologies have made it possible to imagine development of a real-time railway rescheduling process. In fact, many modern rapid transit railways currently have automated rescheduling systems. This is possible since they operate systems with limited network complexity, uniform vehicle types, a dense train position detection system and a comparatively small number of external influences. In contrast, developing an automated rescheduling system for a mixed traffic railway network is very difficult given network complexity and size, the variation in train types, the (relative) lack of train detection and control equipment and the different train operating companies, to name several of the most obvious reasons.

Several research projects have been launched which aim to efficiently realize real-time rescheduling systems for railways (for example the COMBINE 2 project described by Giannettoni and Savio in 2004 or results by Jacobs 2004). So far, the results of the rescheduling process are described on an abstract level. Mazzarello and Ottaviani take this a step further, explaining a rescheduling process for a real-time traffic management system and illustrate it for the ETCS Level 3 case. But none of these projects anticipated the influence of the train driver behavior, the variations at stations and problems along the entire production process chain.

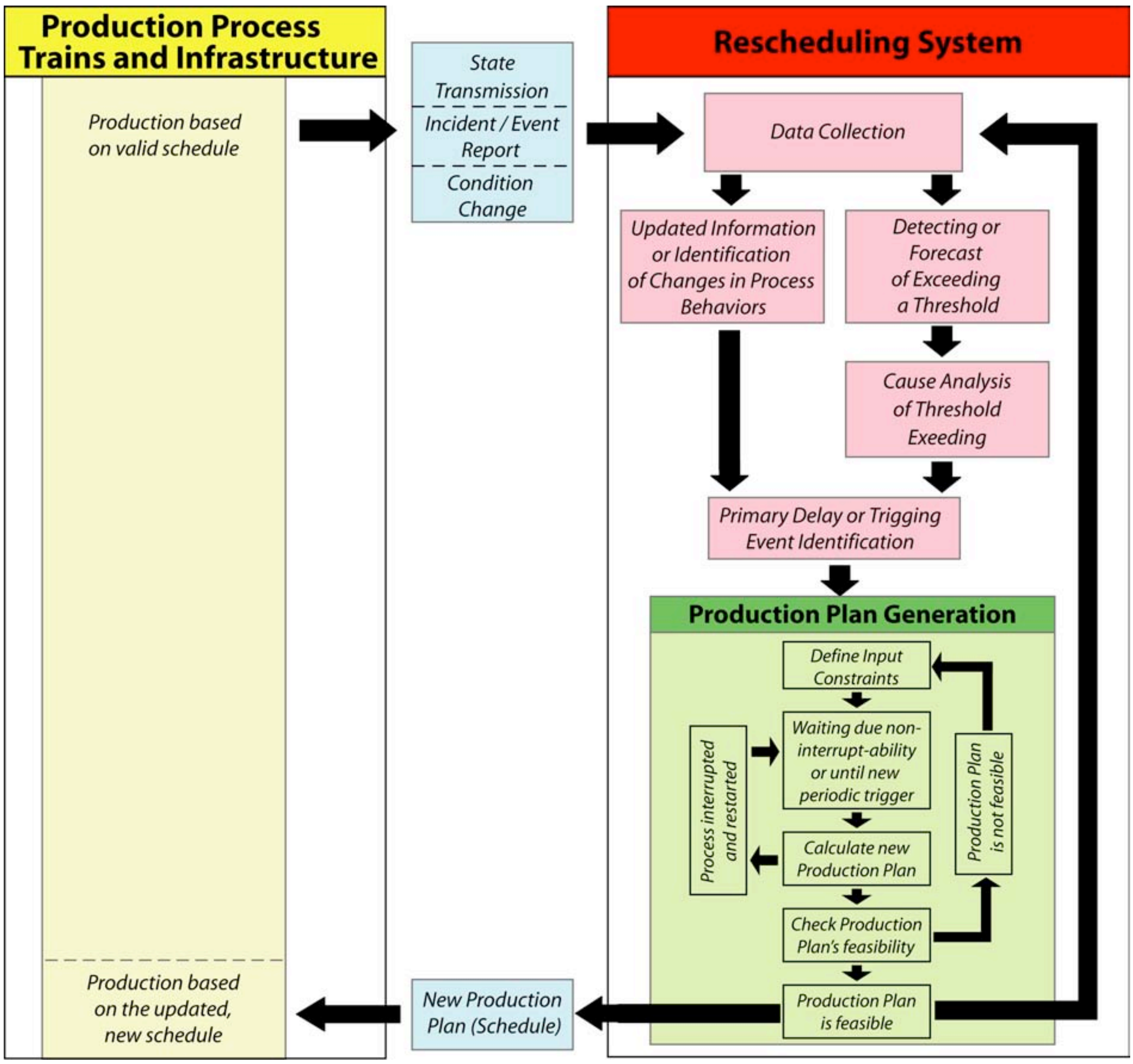
The following terms have precise meanings in the context of this research and are therefore defined:

- *Production plan* – for each resource participating in the production, a plan is specified including beginning and ending times as well as a detailed description for each task of the given resources. The production plan contains the timetable, operating instructions, route definitions, etc. For example, the production plan for a locomotive driver consists of the schedule he must follow.
- *Rescheduling* – the process of updating an existing production plan based on the system's current state and predicted behavior.
- *Integrated Real-Time Rescheduling* – the combined process of updating an existing production plan (schedule) in real-time, and executing the new plan with the assistance of IT tools. In other words, a new schedule would be developed based on the current system state; then, this schedule would be implemented by all system actors (e.g. drivers, infrastructure operators, conductors) with the help of technical devices (i.e. man-machine interfaces and/or fully automated systems).

3. Railway network rescheduling process

Railway network rescheduling is a complex multi-stage process. Figure 3 illustrates the process and tasks on a conceptual level. The process is based on information regarding network conditions (e.g. infrastructure status, train positions). This information is compared to pre-defined thresholds to determine if it is necessary to begin the rescheduling process. If the rescheduling process is triggered, algorithms are used to generate new schedules. These schedules are then transmitted to all relevant actors and implemented.

Figure 3 Railway network rescheduling process.



This section describes these rescheduling sub-processes in detail and issues related to their implementation. The following sections outline how they relate to the integrated real-time rescheduling framework.

3.1 Train detection and threshold exceedence determination

The first step in the rescheduling process is determining if a train has exceeded a pre-determined threshold. This section describes two aspects of this process, the types of threshold exceedences (i.e. reasons for rescheduling) and second, the specific techniques used to determine whether a threshold has been exceeded.

3.1.1 Reasons for rescheduling

There are four basic reasons for starting the rescheduling process:

- Deviation – The most common type of deviation is a time deviation, specifically exceeding a pre-defined tolerance bandwidth in a production plan (e.g. a train is late or early). Other types of deviations include a train using a different route than planned or operating a different combination of trains. Deviations can be the result of an incident, a disturbance, or may also originate in a creeping process. A deviation can be identified both when the deviation occurs or when a deviation can be predicted.
- Disturbance – A disturbance means that due to reduced availability or productivity of a technical component, or of an actor participating in the production, production cannot be continued as planned. After the disturbance is eliminated (and the system regains productivity), the production plan can be adapted.
- Incident – An incident interrupts or delays production on a short-term basis. After an incident all resources are fully available again and production can be continued as planned. Incidents often lead to schedule deviations. Incorrect inputs or other human errors are also classified as incidents.
- Service Change – A service change consists of adding or changing trains in the existing schedule (e.g. changes in a service by adding supplementary stops or adding a new freight train). These types of changes may also impact other lines and services and therefore a new schedule is needed.

A disturbance can be distinguished from an incident or deviation by the fact that after a disturbance, new plan conditions (e.g. new vehicle characteristics or new infrastructure characteristics) must be defined, while in the case of an incident or deviation, in most cases only a change of the time conditions for the next reference points is required.

Table 1 summarizes the rescheduling process goals and time restrictions for completing the rescheduling process for different problems (reasons for rescheduling). As shown in Table 1, the goals, priorities and time restrictions differ significantly depending on the type of problem, location, number of affected trains and the cause.

Table 1 Primary rescheduling goals and time demands by problem type.

Type of problem	Primary rescheduling goal	Time demand
Exceed tolerance bandwidth or incident	- Maximize flow/maximize productivity - Minimize total network delay - Ensure connections with connecting trains	High
Reduced availability of vehicle or infrastructure (small disturbance)	- Limit delay propagation to a geographical area or to a number of trains - Maintain circulation plan	Medium
Interruption of infrastructure or vehicle defect (large disturbance) or service change	- Ensure the flow of the transport chain - Ensure that all stations are served - Minimize the number of replacement trains and additional trains needed	Low

Finally, while considering the goals presented in Table 1, it must be emphasized that the passenger is of highest importance in the rescheduling process, regardless of what additional actions must be taken. It is always essential, however, to determine if the extraordinary costs and personnel expenses that are incurred by these additional actions are reasonable and if they are applied systematically.

3.1.2 Time deviation

The most common reason for triggering the rescheduling process is a time deviation, in other words a train that is either late or early. Basis for determining a threshold exceedence is the precondition that each train (each actor) receives a timetable (production plan) and a bandwidth within which he must operate. This production plan must always be available, without contradictions, and it must be feasible. The bandwidth may be constant or differ in size, depending on route, train type, daytime and function.

The detection of train locations and in particular their concentration or frequency is of utmost importance for the rescheduling system. In the case of deviation from the schedule, if the pre-defined tolerance bandwidth is exceeded, detection density is decisive as well as how fast and exactly the exceeding is identified and how accurately the future behavior can be predicted. Thus, this is a crucial factor, which influences the performance of the rescheduling system. The aim of detecting trains is to determine as soon as possible an exceeding of the pre-defined tolerance bandwidth that has already been taken place or that is about to happen.

The three methods for determining if a time deviation threshold has been exceeded are:

- Infrastructure Train Location Detection – In this method permanently installed infrastructure elements transmit information on train status to the network operator.
- Periodic Train Location Transmission – In this method trains automatically transmit their location on a regular basis (with respect to time) to the network operator (e.g. using radio).
- Participant Transmission – In this method participants (people) directly inform the network operator on train location.

The advantages and disadvantages of the three methods are listed in Table 2.

Table 2 Overview of train detection methods.

Method	Advantages	Disadvantages
Infrastructure-based Train Location	- No additional technical equipment or infrastructure extensions are needed. - Very accurate position.	- No information in a case of delay until a location point is passed. - Signification loss of time until deviation is detected.
Periodic Train Location	- Periodic Information about train state and position is available. - Additional information is easily transmittable.	- Investments for train locations (e.g. GPS) and communication are needed. - Detection is not everywhere possible.
Participant Train Location	- Detailed information about delay reason is transmittable - Is possible in addition to other detection strategy	- Immense time lag until delay is transmitted to the rescheduling system.

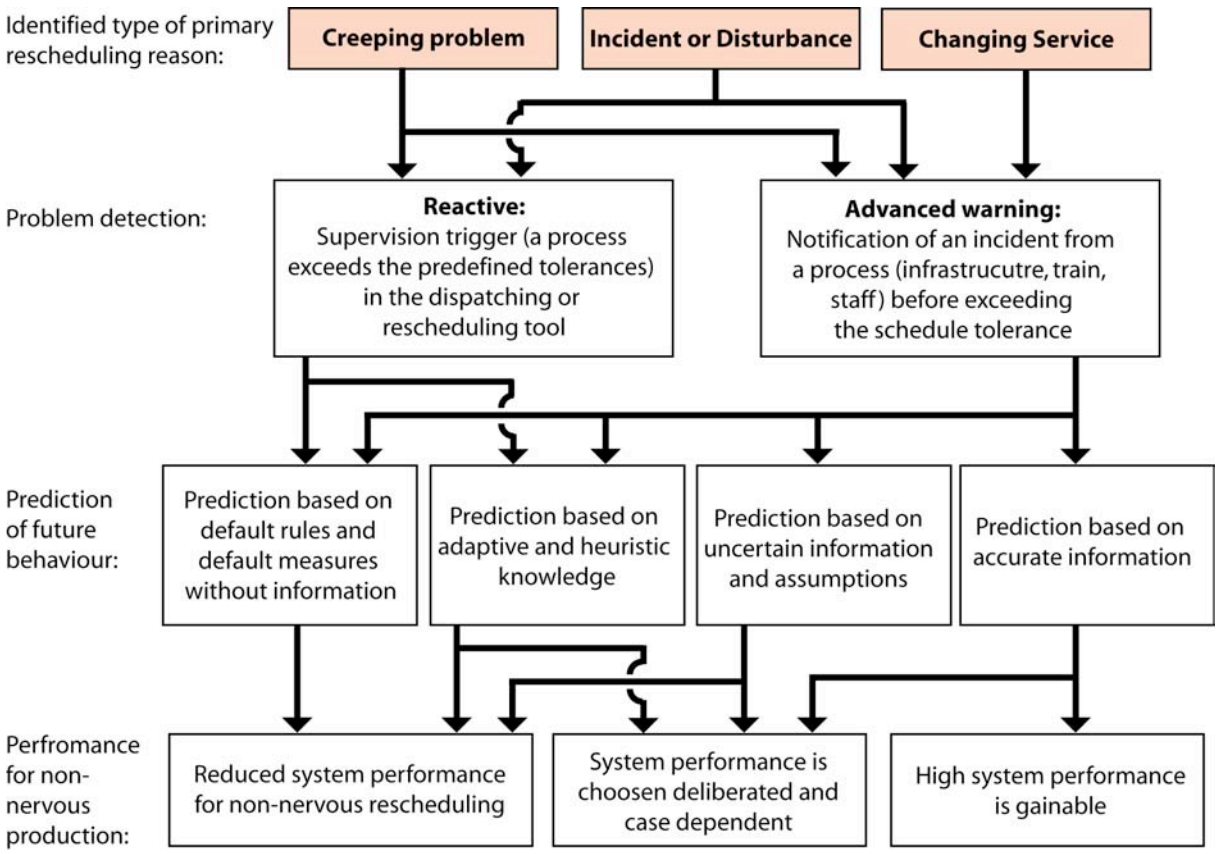
3.2 Generate stable prediction and new production plan

Once it has been determined that a threshold has been exceeded and a new production plan should be developed (rescheduling), the infrastructure operator must prepare a new timetable. In order to prepare a new timetable two tasks must be completed. First, the main reason for the threshold exceedence must be determined and based on this, constraints for the future behavior of the actors are defined. With the constraints as input, one or more rescheduling algorithms must be run to actually develop the new production plan.

3.2.1 Process Outline

Figure 4 summarizes the process of generating a new production plan starting from the reason for initiating the rescheduling and ending with the tolerance levels that should be accepted. The rest of this section describes this process.

Figure 4 Rescheduling process and performance summary.



At the top of Figure 4, the reasons for beginning the rescheduling process are listed. The second row shows the two methods of problem detection, based on the rescheduling reason: either the problem is detected when it occurs (reactive) or in advance. The information necessary to predict a problem in advance of it occurring depends on the reason for the rescheduling and the data flow (these vary depending on problem type).

Once a problem requiring rescheduling has been identified, the system must make a prediction about the future behavior of all trains and actors. As shown in Figure 4, there are four different methods for making a prediction of future network conditions. The choice of method depends on the type of problem, data availability and urgency of the need for a new production plan. Section 2.2.2 below describes this in more detail.

The last row in Figure 4, the production plan tolerance, describes the rescheduling system's performance based on the various methods for predicting future system status. As shown, the lower the information accuracy and time available for the rescheduling process (left side of the figure), the lower the performance level; on the other hand, with more accurate information and more time (right side of the figure), system performance can be improved.

When determining whether or not to prepare a new schedule (i.e. to undertake the rescheduling process), a conflict exists between high productivity (i.e. operating many trains) and rescheduling stability. This means that a rescheduling process is more likely to be initiated in the case of schedules with minimal reserve times because the required schedule bandwidth conditions cannot be kept. This leads to a very high level of data exchange and to nervous production behavior (constant exceeding of thresholds which leads to frequent development of new production plans). This should be avoided in all cases.

On the other hand, performance is lowered unnecessarily if predictions over the future behavior of actors are too conservative. This, too, should be avoided. The conflict between stability and performance is therefore a central aspect to consider in the rescheduling process.

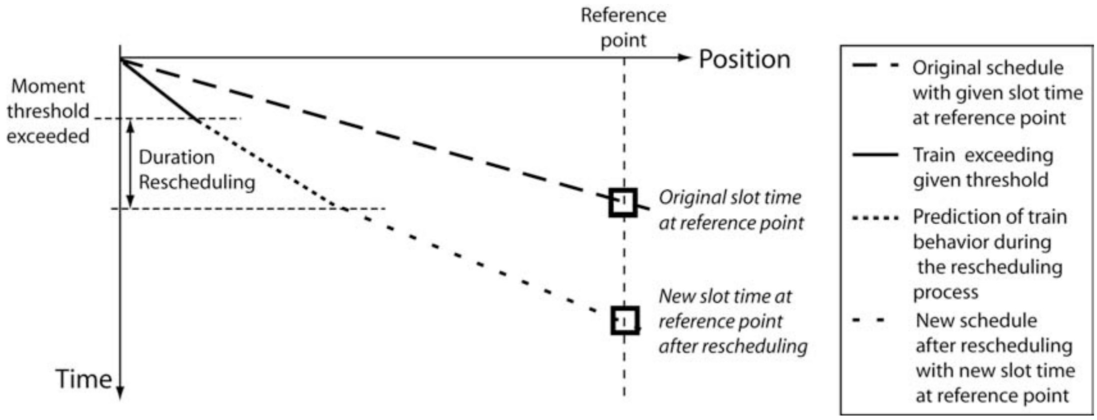
Therefore, it is important to know the time demands for addressing the particular problem (examples are presented above in Table 1) in order to respond correctly: waiting to collect more precise information before starting the rescheduling process, or to adopt pre-defined conditions as fast as possible which could lead to a reduction in productivity.

3.2.2 Prediction of future system state

It takes time for the rescheduling algorithm to prepare a new schedule. Since the system will change during the time between starting the rescheduling algorithm and when the rescheduling is finished, a prediction of the future system state is needed or the new schedule developed in the process will be irrelevant (given the changed system state). The prediction of future system status must be based on data and information on the actors, which are mathematically converted to boundary conditions. In order that this process step is possible, the main reasons for the temporal deviations have to be identified or assumed in the case of missing information.

More specifically, the prediction of future system status consists of two parts (Figure 5 illustrates the two prediction parts in a time-distance diagram). First, the behavior of the trains during the whole rescheduling process is predicted. Thereafter, the second part consists of predicting a sample of new time windows (slots) for trains at specified reference points (time windows). These predicted time windows for all trains are the input (constraints) for the rescheduling system.

Figure 5 Two-level prediction process



3.2.3 Rescheduling algorithms

The actual work of developing the new production plans (schedules) is done in a rescheduling algorithm. This paper focuses on how the new production plans can be most efficiently be produced and implemented rather than on the rescheduling algorithms themselves. However, research on developing new rescheduling algorithms is being completed in another part of this research described by Caimi et al. 2007 or Wuest 2006 and by other researchers (e.g. D'Ariano et al. 2007, Kraft 1987, Jovanovic and Harker 1991, Sahin 1999, or Tornquist 2007). At the present time, rescheduling algorithms are not available for dense mixed-use railway networks since they cannot handle all the following requirements simultaneously:

- Deep level of model detail;
- Mixed rail traffic;
- Accurate prediction of future behavior (especially after disturbances or events);
- Complex network topology/layout (most of the rescheduling research focused on single-line sections and neglected the complexity of train rerouting); and
- Conflict resolution and rescheduling within reasonable time

However, a smart discretization of network and time (called PULS) offers the possibility for fast rescheduling solutions. This approach, developed by Roos 2006, is used for a pilot project in the area of Lucerne.

3.2.4 Problem management process

All railways have sets of procedures that actors use to address problems. These problem management processes, which operate in parallel with the rescheduling process, are designed to eliminate or reduce the impact of disturbances. For example, if a locomotive loses power and stops running, there is a specific set of procedures that the locomotive operator follows in an attempt to regain power. The problem management process is not part of this research, although it is important to note that all activities in this process that help to predict the future behavior, or are information about time, are transmitted immediately to the rescheduling system so that the rescheduling system can accurately predict future conditions and thereby develop production plans best suited for implementation.

3.3 Developing and implementing new production plans

As outlined above, this research project has two components: developing rescheduling algorithms, and analyzing how these new algorithms can be most effectively implemented using the integrated real-time rescheduling framework. There are many questions that still must be answered regarding how, specifically, the rescheduling algorithms should be implemented, and these issues are outlined in this section. Once appropriate rescheduling algorithms are available, system operators will face two key questions:

- When should a new production plan be developed? And
- When should a new production plan be implemented?

The most obvious answer to these questions is that a new production plan should be developed and implemented every time the system detects that a threshold has been exceeded, but this would lead to a very unstable situation where new production plans were constantly being generated (nervous production). Instead, the generation of a new production plan should only be initiated if conflicts arise due to the threshold-exceeding event or if the currently valid production plan can no longer be carried out. This approach combined with the two-stage method helps to avoid a nervous production process.

A second approach is to periodically generate new production plans (e.g. every 2 minutes). This leads to the question of what cycle time should be selected for generating new production plans. Short cycles could lead to nervous production while longer cycles could reduce the effectiveness of the rescheduling process.

In both approaches to generating new production plans (event-driven and periodic), a further question is whether the development of new production plans should be interruptible or not. In other words, should the production plan generation process be interrupted if the system status changes in such a way that the new production plan will no longer be optimal once it is ready to be implemented?

A hybrid approach for producing new production plans seems most reasonable. In this approach a normally periodic rescheduling process could be interrupted and re-initiated due to an event. This works best if a new priority is ascribed to each threshold-exceeding event and only events of a given priority cause the process to be interrupted. For example, events affecting critical trains and occurring in bottleneck areas would be given high priority.

Of course any interruption in the rescheduling process would increase the time needed to develop the new production plan.This leads to the question of whether it makes sense to implement infeasible production plans (timetables). While it sounds illogical to implement infeasible timetables, implementing one may make sense if these timetables can be developed quickly and if they move the system a step closer to a status where it will be possible to implement an optimal timetable that takes longer to develop.

A good example of this problem occurs when the behavior of an actor during the rescheduling process (i.e. while the algorithms are generating the new production plan) is inconsistent with the actor's predicted behavior (which was used as an input to the rescheduling process). In this case it could be impossible for the particular actor to fulfill the new production plan. On the one hand, implementation of the infeasible production plan will generate a threshold exceeding event, which will re-initiate the rescheduling process immediately after transmitting the new production plan and lead to a high rescheduling nervousness (which should be avoided for ergonomic reasons). On the other hand, generally even a infeasible production plan based on "real-time" data is better than the original production plan once an event has occurred. Furthermore, if sub-optimal production plans are not implemented, the rescheduling process could go into a loop during which no valid production plans would be developed over a long time period. These reasons support the idea that it is more important that all actors always have a feasible production plan and are within their limits rather than to produce an optimal production plan, which is very unstable.

3.4 Transmission and Implementation of new production plans

The last step in the rescheduling process is transmitting the new production plan to all affected actors. After receiving this information the actors are responsible for implementing the new plan within the pre-defined limits.

In many cases it is difficult or impossible for actors to implement the new production plans as accurately as necessary without assistance. This is particularly true for train operators who must drive their trains following very precisely defined trajectories. In the case of operators, it is essential to present the trajectory information visually with the aid of user-friendly displays. An iDMI (intelligent Driver Machine Interface), for example, could give information concerning time deviation, maximum speed permitted and planned reference speed. This is similar to an instrument landing system for pilots with the additional constraint that it also includes time constraints. Initial results from an SBB study, presented by Fenix et al. in 2005, show that operators using this type of display cross reference points very accurately (+/- 15 seconds).

4. Network classification

The integrated real-time rescheduling process can be most effectively used if the railway network is divided into areas with excess capacity and areas with capacity bottlenecks. This section outlines how using this network helps optimize the rescheduling process.

4.1 Network classification

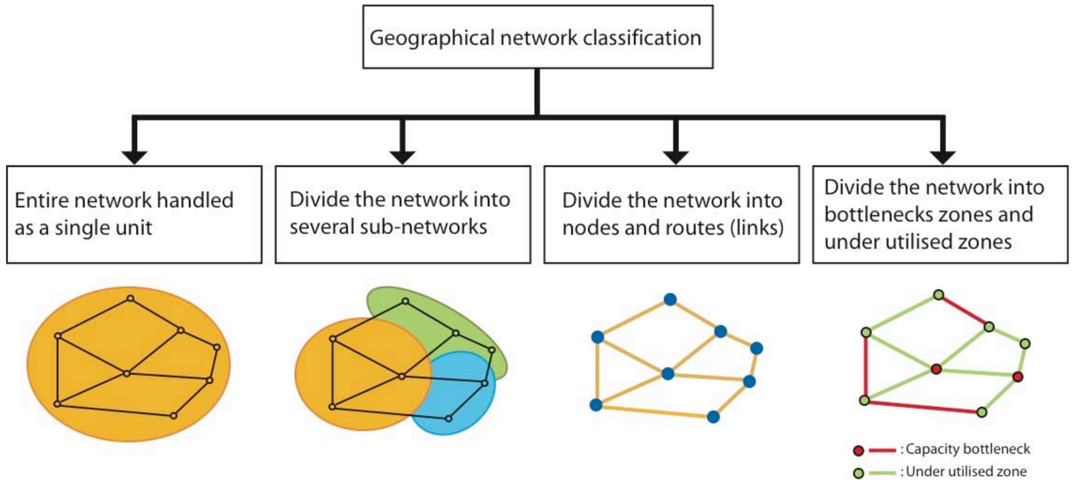
Railway networks can be classified in several different ways depending on the purpose of the ultimate objective (e.g. geographic areas for maintenance management). In many cases network divisions are based on historic developments, which may no longer be optimal for the particular purposes. Figure 6 illustrates four examples for dividing a railway network, these are:

- The entire network is planned and operated as a single unit. This is only possible for small networks and is mainly applied on urban rapid transit systems.
- The network is divided into connected sub-networks. Each sub-network is responsible for itself and there is defined coordination between the different sub-networks. This is the classic method for planning and operating railways.
- The network is divided into nodes and routes. At some points, there are also other regimes between route and station on an operational level. This type of division is mainly used during the planning process.
- The network is divided into capacity bottleneck areas (condensation zones) and areas with excess capacity (compensation zones). This approach, described by Laube et al. in 2007, is used in the planning of timetables and schedules, but is generally not formally defined. The integrated real-time rescheduling framework formally defines and uses this network division to optimize the rescheduling process (outlined below).

In terms of the rescheduling process, the advantage of a large network is that it does not need complicated, multistage processes to generate a new production plan; the disadvantage is that, since it is a large network, developing schedules is a complex and long process. Developing a timetable for a divided network is easier in the sense that the problem is smaller, but it adds the need for coordination between the different areas. This is especially problematic during the rescheduling process since a new schedule affecting trains outside the sub-network must be coordinated with the other sub-networks, adding a second step to the process of developing a new schedule (compared to developing a new schedule for an entire network).

Dividing the network into capacity bottlenecks and areas with excess capacity is a special example of dividing the network into nodes and links. As outlined below, this allows taking advantage of the integrated real-time rescheduling framework.

Figure 6 Network classification schemes.



4.2 Condensation – compensation zones

The concept of condensation and compensation zones is based on the idea that some nodes and links in a railway network have excess capacity (compensation zones) and some have no excess capacity (condensation zones). In condensation zones it is critical that trains be operated extremely precisely or delays will occur that may propagate throughout the entire network. In compensation zones excess capacity provides trains with operational flexibility (i.e. speed control) that allows them to maximize the capacity and schedule stability in condensation zones. More specifically, trains can be operated in zones with excess capacity so that they arrive at exactly the right time and at exactly the right speed at the gateways to the capacity bottleneck zones. Note that arriving at both the correct speed and time is necessary to maximize capacity. Another example is providing an exact departure time for a train from a station platform. The integrated real-time rescheduling framework is designed to provide this type of time and speed information to all affected parties in the network.

The division into condensation and compensation zones facilitates operating capacity bottlenecks optimally and therefore guarantees that a network's current weak spots are always the focus of planning. The integrated real-time rescheduling algorithms must be able to provide new production plans that specify a valid slot time for all trains entering the condensation zone and a specific platform departure time accurate to a tenth-minute.

In summary, the integrated real-time rescheduling framework is based on a systematic, saturated use of network capacity bottlenecks. A data exchange (input constraints for the rescheduling algorithms) between condensation zones coordinates the rail traffic flow within the entire network.

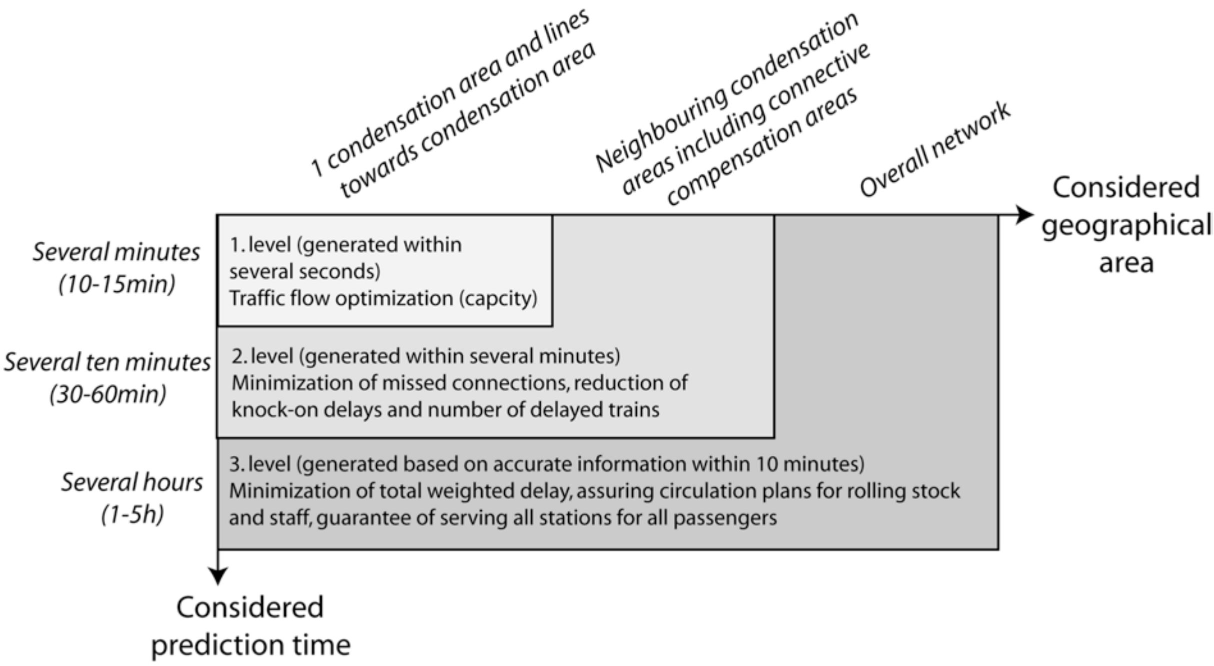
5. Implementing the integral real-time rescheduling framework

The objective of this research project is to develop an approach for increasing rail network capacity at minimum by effectively linking the rescheduling process with train traffic control. The integrated real-time rescheduling approach can be described as a superposition of two control loops.

The external loop is responsible for ensuring that all actors have a valid and conflict-free production plan (including a timetable, rules and routes) available at all times. In the case of disturbances or deviations, a new production plan, based on the current data, will thus be generated immediately.

It is very important that the rescheduling process is carried out within the shortest possible amount of time such that traffic flow and capacity is optimized in the bottleneck areas. Therefore, a chronological multistage method with 3 levels is a good strategy for rescheduling (see figure 7). Thereby, the amount of information, the optimization goal, the temporal requirements, the considered time prediction horizon and the included geographical region are changing.

Figure 7 Configuration and tasks of a multi-level rescheduling process.



The first level consists of developing a 'good' new production plan quickly, although it would be based on limited information. The optimization goal thereby is to maximize the traffic flow within the capacity bottleneck area. Thus, only trains within or approaching the affected condensation area are taken into account and also the prediction horizon is limited to several minutes. One possibility is that new production plans could be developed with pre-defined conditions based on the particular event type.

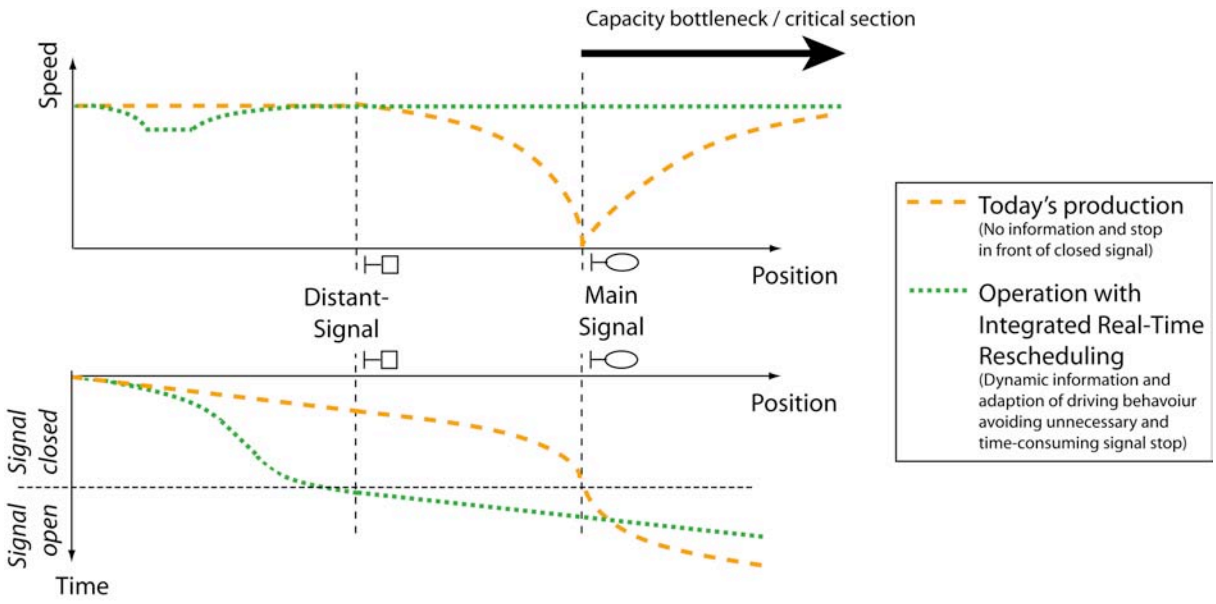
The second level assures the coordination between the neighboring condensation areas such that connections are optimized. Thereby, the time from the first stage is used to collect more information for a more accurate prediction. Also rolling stock and crew roster information is used as input for generating the new production plan. The new production plan then is also valid for a longer temporal horizon.

The last level is to assure global optima as for example minimal amount of weighted delays or missed connections in the network. Also, minimizing changes in circulation plans for rolling stock and staff or assuring a persistent transportation chain are regarded in the rescheduling optimization process. Thereby, the new production plan is generated for all condensation and compensation areas by a global coordination. This task is not time-critical and can be done by gathering all relevant information from all affected actors.

The internal loop is responsible for ensuring that the production is carried out as closely as necessary to the current production plan (schedule). Particularly for running trains, it ensures that the pre-defined tolerance bandwidth (e.g. +/- 15 seconds) around the planned trajectory is not exceeded. In order to realize this approach, the described rescheduling processes must be adopted, and the methods and technologies have to be developed according to the defaults.

The integrated real-time rescheduling framework, in combination with the network division into condensation and compensation zones, allows railways to maximize the utilization of network bottleneck areas. This is achieved by reducing unintended stopping and acceleration (which is very time and energy consuming) in or in front of the condensation areas. Trains are therefore slowed down and 'delayed' or speeded up before reaching the critical section (see Figure 8). Since most of the additional capacity and schedule stability gained through the integrated real-time rescheduling framework can be obtained using the existing infrastructure elements and available technology, the approach is extremely cost effective.

Figure 8 Train control with and without integrated real-time rescheduling.



In addition to the more efficient use of the infrastructure, the integrated real-time rescheduling framework can be used to reduce the overall energy consumption. Slowing down trains for smooth passing of bottleneck areas and thus avoiding unplanned stops allows reducing energy consumption. In the same way, knowing the position and states of all trains helps to coordinate the speed of all trains and thus also unnecessary high speed is avoided. The energy saving can even be improved by combining this approach with algorithms proposed by Franke et al in 2002 or Howlett and Pudney in 1995 for optimal driving strategies including rolling out and taking gradients of the track into account. To summarize, the integrated real-time rescheduling framework helps to optimize the energy usage of all trains in a system together by minimizing their conflicts and thus avoiding energy-intensive stopping and accelerating actions.

As part of the research project, simulations of the integrated real-time rescheduling approach were completed to evaluate its impact on capacity and schedule stability. The simulation was completed using the OpenTrack train simulation program (see Nash and Huerlimann 2004 for an extended description of the tool). OpenTrack is a synchronous, event-driven micro-simulation application that precisely models track topology and train characteristics. Thus, all relevant process elements (infrastructure, rolling stock, timetable) as well as their interactions are simulated very accurately.

The simulation was executed for a specific area around the dead-end station of Lucerne. Thereby, the simulation showed that the integrated real-time rescheduling approach could significantly increase schedule stability, reduce knock-on delays and this in combination with an increase of the capacity by up to 20% (see Luethi et al. 2007 for more details). Rerouting of trains in station areas and speeding-up trains were identified as the most effective measures to cut down consecutive delays. The simulations were also used to evaluate the impacts of certain constraints on the integrated real-time rescheduling process. Specifically, the simulation showed that inaccurate system status data and increasing the length of time between detecting an event (threshold exceedence) and completion of the rescheduling process both reduce the potential impact of the integrated real-time rescheduling approach. An especially important finding was the relationship between rescheduling process duration and total system delay; the results show that total system delay increases stepwise and significantly once the duration reaches a certain point. This shows the importance of developing a coordinated approach to rescheduling and its effective implementation as well as fast and efficient rescheduling algorithms.

6. Conclusions

Public rail traffic has best prerequisites for sustainable transportation. Rail transport is safe, reliable, and saves space, resources and environment. But this is not enough to be competitive against road traffic. To get more popular and play a key role, rail operation has to solve several challenges. Especially, a competitive and close to the demand's need rail service has to be offered and the costs have to be reduced and controlled. Therewith, increasing the efficiency for railways is a key challenge.

The paper has shown that rescheduling, in combination with train control, represents a promising low cost approach for increasing capacity and stability of railway networks and thus can strengthen the position of railways in transportation. The detailed process description has shown that in addition to developing new and fast algorithms for the rescheduling process, it is critical that careful thought be given to how the rescheduling process can be implemented within the whole production process. Only by adjusting the production processes and sub-processes the full benefit of rescheduling can be achieved.

Beside the improvements for capacity and stability, the integrated approach helps to reduce the overall energy consumption by influencing the trains and minimizing unnecessary stops. And finally, the new framework is the basis to provide more precise and personalized information for customers online in the case of a delayed train or after an incident.

The research will continue to use the Lucerne station area as a pilot project area for analysis and evaluation of the integrated real-time rescheduling approach by the SBB. This will help showing the approach's effectiveness for a specific condensation area. Thereafter, several condensation areas will be connected together to evaluate the approach's effectiveness on the network level.

Acknowledgements

This paper was produced as part of the Swiss Federal Railway's PULS 90 research project. The author would like to thank the partners at the SBB Infrastructure Division for their ideas and funding for this project namely Felix Laube, Raimond Wuest, Samuel Roos and Oskar Stalder and also Jeff Kennworthy and Ulrich Weidmann for their input and fruitful discussions.

References

Caimi, G., Burkolter, D., Herrmann, T., Chudak F., Laumanns M. 2007. Design of a railway scheduling model for dense services. In Proceedings of the 2nd International Seminar on Railway Operations Modelling and Analysis. Edited by I.A. Hansen, A. Radtke, J. Pachl, and E. Wendler (eds.), Hannover.

D'Ariano, A., Pranzo, M., Hansen I.A. 2007. Conflict Resolution and Train Speed Coordination for Solving Real-Time Timetable Perturbations. IEEE Transactions on Intelligent Transportation Systems 8(2):208-222.

Eichenberger, P. 2007. Kapazitätssteigerung durch ETCS. Signal und Draht 3:6-14.

Fenix, J., Graffagnino, T., Sagot, J.-C., Valot, C. 2005. User centred design applied to increase timetable stability. Proceedings of the 36th Tagung Moderne Schienenfahrzeuge, Graz, Austria.

Franke, R., Meyer, M., Terwiesch, P. 2002. Optimal Control of the Driving of Trains, Automatisierungstechnik 50(12):606-613.

Giannettoni, M., Savio, S. 2004. The European project COMBINE 2 to improve knowledge on future rail traffic management systems. In Computers in Railways IX: 695-704. Edited by Allen J., Hill R.J., Brebbia C.A., Sciutto G., Sone S. Southampton: WIT Press.

Howlett, P.G., Pudney, P.J. 1995. *Energy-efficient train control*, London: Springer.

Kraft, E. R. 1987. A branch and bound procedure for optimal train dispatching. 1987. Journal of the Transportation Research Forum 28:263-276.

Jacobs, J. 2004. Reducing delays by means of computer-aided 'on-the-spot' rescheduling. In Computers in Railways IX 603-612, Edited by Allen J., Hill R.J., Brebbia C.A., Sciutto G., Sone S. Southampton: WIT Press.

Jovanovic, D., Harker, P.T. 1991. Tactical Scheduling of Railway Operations: The SCAN I System, Transportation Science 25(1): 46-64.

Laube, F., Roos, S., Wuest, R., Luethi, M., Weidmann, U. 2007. PULS 90 – Ein systemumfassender Ansatz zur Leistungssteigerung von Eisenbahnnetzen, ETR Eisenbahntechnische Rundschau 3:104-107.

Luethi, M., Medeossi, G., Nash, A. 2007. Structure and Simulation Evaluation of an Integrated Real-Time Rescheduling System for Railway Networks, In Proceedings of the 2nd International Seminar on Railway Operations Modelling and Analysis. Edited by I.A. Hansen, A. Radtke, J. Pachl, and E. Wendler (eds.), Hannover.

LITRA Information service about Swiss public transport 2004. *Verkehrszahlen - Traffic data 2004*, Bern.

Maxwell, R. 1999. Intercity Rail Fixed Interval Time Transfer System: Applicability of the Integrated Taktfahrplan Strategy to North America, Transportation Research Records 1691:1-11.

Mazzarello, M., Ottaviani E.. 2007. Traffic Management System for Real-Time Traffic Optimisation in Railways. Transportation Research Part B 41(2):246-274.

Nash, A., Huerlimann, D. 2004. Railway simulation using OpenTrack. In Computers in Railways IX 45-54, Edited by Allen J., Hill R.J., Brebbia C.A., Sciutto G., Sone S. Southampton: WIT Press.

Roos, S. 2006. *Bewertung von Knotenmanagement-Methoden für Eisenbahnen*. Master Thesis, Institute for Transport Planning and Systems, Swiss Federal Institute of Technology Zurich.

Sahin, I. 1999. Railway traffic control and train scheduling based on inter-train conflict management, Transportation Research Part B 33(7):511-534.

SBB Swiss Federal Railways. 2006, *Annual Report 2006*, Bern.

Tornquist, J., Persson J.A. 2007. N-tracked railway traffic re-scheduling during disturbances, Transportation research Part B 41(3):342-362.

UIC International Union of Railways. 2004. Capacity, *UIC-Codex 406*, Paris.

Ulius, M. 2005. Delay Analysis of Rail 2005 1st Phase Using OpenTimeTable. In Proceedings of the 1st International Seminar on Railway Operations Modelling and Analysis. Edited by I. A. Hansen, F. M. Dekking, R. M. P. Goverde, B. Heidergott, L. E. Meester (eds.), Delft.

Wuest, R. 2006. Dynamic Rescheduling based on Predefined Track Slots. Proceedings of 7th World Congress of Railway Research, Montreal.