Increasing Railway Capacity and Reliability through Integrated Real-Time Rescheduling

Increasing Railway Capacity and Reliability through Integrated Real-Time Rescheduling

Marco Luethi (corresponding author), Andrew Nash, Ulrich Weidmann

Institute for Transport Planning and Systems, ETH Zurich
Wolfgang-Pauli-Strasse 15, 8093 Zurich, Switzerland
E-mail: luethi\at\ivt.baug.ethz.ch, andy\at\andynash.com, weidmann\at\ivt.baug.ethz.ch

Felix Laube, Raimond Wuest

Swiss Federal Railways, SBB AG
Infrastructure Division, 3000 Bern 65, Switzerland
E-mail: felix.laube\at\sbb.ch, raimond.wuest\at\sbb.ch

Abstract: The paper presents an overview of an ongoing research project designed to identify methods for increasing the capacity and stability of heavily used mixed-traffic railway networks without making significant new infrastructure investments. The proposed approach combines rescheduling with network traffic control in an attempt to minimize schedule reserve times without reducing stability and thereby increase network capacity. The paper presents an overview of the approach, discusses the rescheduling process and issues that need to be addressed in order to effectively use new approach, outlines how the network should be divided to best use the new approach, and presents results of a microscopic railroad simulation that shows the approach can improve capacity, reduce delays and increase schedule stability.

Acknowledgements: This paper was produced as part of the PULS 90 research project and is a result of cooperation with the Swiss Federal Railroad Infrastructure Division. The authors would like to thank their partners at the SBB Infrastructure Division for their ideas, input and funding for this project.

1. Introduction

This paper describes initial results of an on-going research project being completed by the Institute for Transport Planning and Systems at the ETH Zurich and the Swiss Federal Railways (SBB). The project's objective is to identify methods for increasing capacity and stability of railway networks without making significant infrastructure investments.

Research Background: Switzerland's Railway Network

Switzerland has one of the world's most heavily used national rail networks. During the 1990s, the Swiss Federal Railways (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 re-scheduling.

Switzerland's passenger rail service is based on the concept of an integrated clock-face timetable (described in detail by Maxwell 1999 and Nash 2006). This essentially consists of a timed-transfer system for the entire country. The Bahn 2000 program expanded this system to more cities and increased the number of trains operating using this approach.

Initial results, presented by Ullius 2005 and in the SBB's Annual Report 2005, 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 (95.7 % of all passenger trains have an arrival delay of 5 minutes or less). However, it is also clear that the network is operating at the edge of stability (up to 90 trains run through some station areas in a single hour). 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.

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 from two components: safety and schedule reliability. The safety component ensures that trains are separated by enough distance to prevent collisions. 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).

The lowest possible headway is determined absolutely based on safety. This headway is based on the distance it takes a specific train to stop on a specific track segment. There are many strategies for reducing the minimum headway between trains. Many of these are based on communicating "stop" or speed instructions to trains more quickly (e.g. moving block signals, ETCS). These require improvements to signaling systems and on-board equipment.

Once the lowest technically possible headway has been determined, schedule planners add reserve (buffer) time to the schedule to reduce the impact of delays and incidents on network operations, in other words improving schedule stability, but reducing capacity. The relationship between capacity and stability is illustrated in Figure 1.

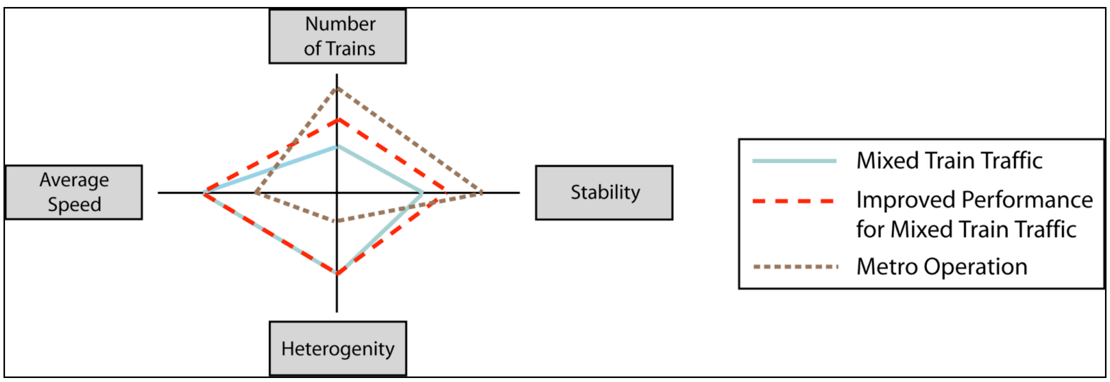

Figure 1: Relationship between capacity and schedule reliability. Source: UIC 406.

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). Eichenberger presented the strategy how ETCS can be used to increase capacity. Revising railway timetables (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 timetable (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 affected 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 affected parties as quickly as possible and these parties must take appropriate actions (which are often different from the planned timetable actions) often immediately.

The process of developing new timetables and communicating them to all affected 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 reduce the need for rescheduling and simplify 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 railroad networks is thus a key factor of success. Several tools were developed to study the impacts of reserve times to capacity and stability. A good example is SIMONE, described by Middelkoop in 2001.

Once the new schedule has been developed and communicated to the affected 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.

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 railroads. For example the COMBINE 2 project (described by Savio 2004). 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.

This paper describes a new framework developed to improve the rescheduling and train control processes thereby increasing network capacity. The framework is called co-production since the infrastructure operator (who determines the schedule) cooperates with the train operator (who controls/runs the trains) in the transportation production process. In addition to helping improve capacity and stability, this new framework can also help address demands for adding slots for freight operations and maximizing the efficiency of rail network operations.

Paper Organization and Definitions

Section 2 of this paper describes the rescheduling process. It includes a description of events that could trigger the rescheduling process and issues related to the implementation of rescheduling on railway networks. Section 3 describes different approaches to classify and manage railroad networks. Section 4 describes the co-production approach. Section 5 presents results of a simulation completed to evaluate the potential benefits of the co-production approach. Finally, Section 6 presents conclusions and an outline of future research.

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).

2. Railway Network Rescheduling Process

Railway network rescheduling is a complex multi-stage process described by Laube and Schaffer in 2006 and is illustrated on a conceptual level in Figure 2.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.

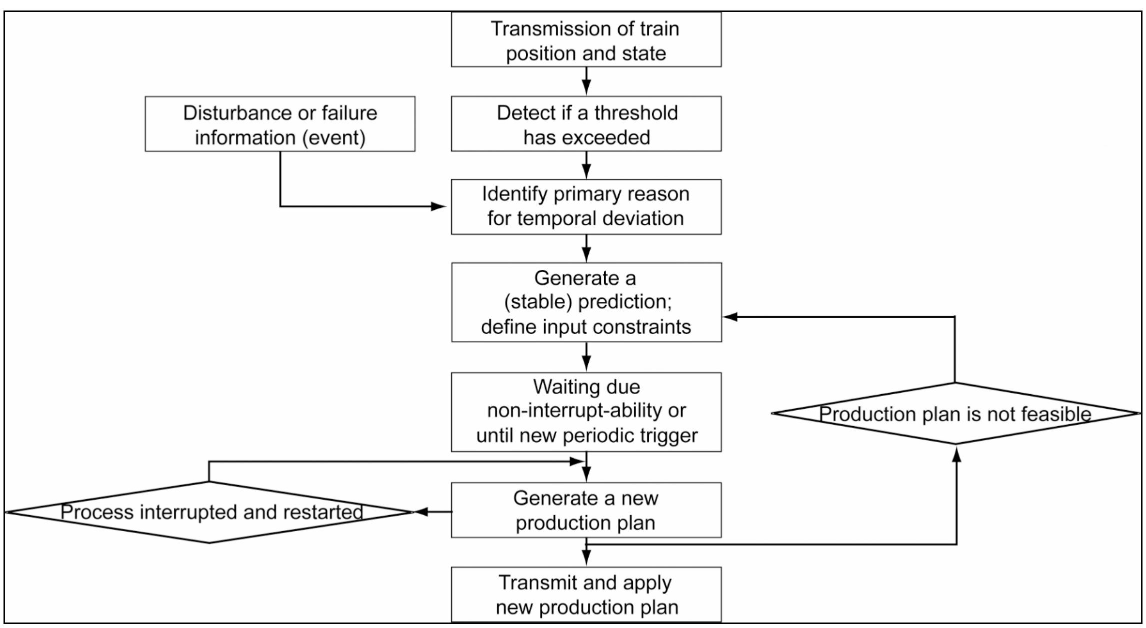

Figure 2: Railway network rescheduling process.

This section describes each of these rescheduling sub-processes in detail and issues related to their implementation. The following sections outline how they relate to the co-production process framework.

2.1 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.

2.1.1 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. Events often lead to schedule deviations. Incorrect inputs or other human errors are also classified as events.
- Service Change – A service change consists of adding or changing trains in the existing schedule (e.g. adding a new freight train). These types of changes may impact other lines 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 event 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.

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.

Type of Problem	Primary Rescheduling Goals	Time Demands
Exceed tolerance bandwidth or event	- maximize flow/maximize productivity - minimize total network delay - ensure connections with connecting trains	high
Reduced availability of vehicle or infrastructure (small disturbance)	- limit delay to a geographical area or to a number of trains - maintain circulation plan	medium
Interruption of infrastructure or vehicle defect (large disturbance)	- ensure the flow of the transport chain - ensure that all stations are served - minimize the number of replacement trains and additional trains needed	low

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

2.1.2 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. This section describes three methods in which a time deviation can be identified and discusses issues related to each of these methods.

The 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 the way 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.

These three methods are explained in detail below. It should be noted that the status of infrastructure elements is also transmitted to the network operator; this is generally done via fixed infrastructure but participants can directly call-in information as well.

Infrastructure-based Train Location Detection

Today, almost the entire railroad system is equipped with various train detection devices (e.g. axle counters, track circuits, balises) designed to operate safety equipment and inform train control decisions. These devices identify trains (actually the heads of trains) at discrete, irregularly arranged points on the infrastructure network. After a short delay, the railroad control centre, the dispatcher and the rescheduling system are provided with the information in combination with additional data when the train passes a detection (reference) point (train number identification system).

This information can be used to determine if a train has exceeded a threshold by comparing it to a list that contains the time windows (in particular, the earliest and latest specified passage times) for each train at the particular reference point. Figure 3 illustrates time windows for a train trajectory at two reference points. A train arriving early will trigger a data event in the control centre. On the other hand, if the train is late (i.e. does not pass the reference point before the latest defined passage time), either the infrastructure operator or the rescheduling system will initiate an event, which activates the rescheduling process.

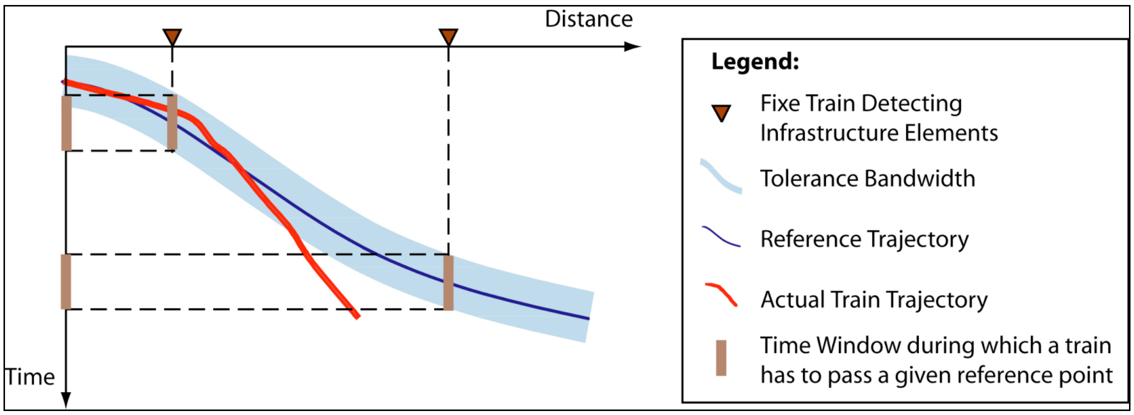

Figure 3: Infrastructure-based train detection

There are several problems with fixed infrastructure transmission. One problem is that in the case of a delay no information about the train location is available (i.e. is the train one-second late or ten-minutes late?). A second problem is that, especially in the case of long distances between infrastructure detection elements, there may be a significant loss of time until a deviation is detected. This problem could be addressed by installing additional passive detection balises to create a denser network of registration points fairly easily. These problems not withstanding, infrastructure transmission is an inexpensive approach, which could be installed and operating within a short time.

Periodic Train Location Detection

In the period train location detection method the train automatically radios location and status information to the control center on a regular basis (e.g. every 30-seconds). Figure 4 illustrates position windows (i.e. the trains should be somewhere between these points at the specified time) for trains using this system. The accuracy of this system can be improved with the aid of infrastructure installations (e.g. balises) to reduce distance measurement errors.

In this system threshold exceedences can be detected and evaluated directly on the train or by comparing to a list similar to that used in the infrastructure detection. The list used in this case is easier to handle than in the infrastructure detection method since the data being evaluated are always the same and a late train is detected in exactly the same way as an early train. A significant advantage of this system over the infrastructure detection method is that it is better able to predict future behavior for delayed trains, in contrast to the infrastructure detection method.

The use of telecommunications technologies can provide operators with a comprehensive picture of all system elements including trains, infrastructure and other elements. Using this approach, it would be possible to monitor the entire production plan and identify potential problems in advance (since each element of the system must be in a pre-defined state at a specific time). Deviations could therefore be detected much earlier and could be accounted for in the development of new production plans (train timetables).

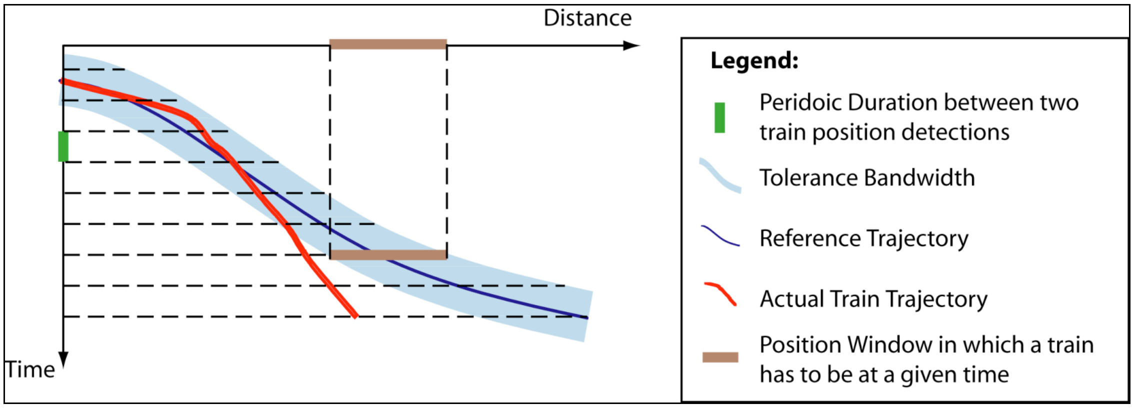

Figure 4: Periodic train location detection

Participant Train Location Detection

A third possibility for detecting and transmitting threshold exceedence information is to use process participants (e.g. locomotive drivers). In this case there would be direct communication between an actor and a person in charge of rescheduling or between and actor and the rescheduling process. In addition to oral communication, it would be possible for participants to directly transmit coded messages to the rescheduling system to accelerate the process. The main problem with this method is the time factor: at least 30 seconds are needed, usually even more until the data and information are processed and realized. So this method could only be recommended for applications that are not time-critical (e.g. during the preparation process or at intermediate stops).

Additionally, oral transmissions would also be valuable to receive precise information when a threshold exceeding already could have been identified otherwise. The more detailed explanations could be useful for more precise predictions over the future behavior and the conditions for generating new timetables could be adapted in a better way.

2.2 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.

2.2.1 Process Outline

Figure 5 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.

At the top of Figure 5, 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 will depend 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 5, 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 5, 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.

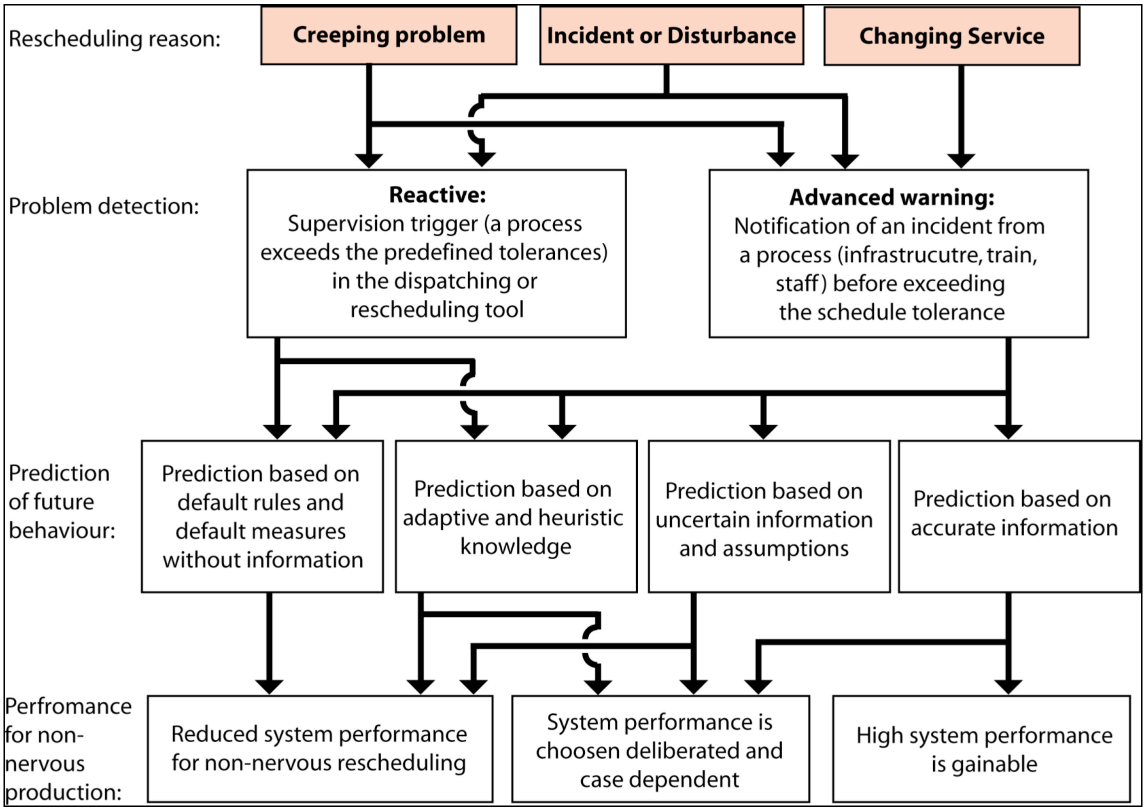

Figure 5: Rescheduling process and performance summary.

When determining whether or not to prepare a new schedule (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.

2.2.2 Prediction of Future System Status

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 predicting a sample of new time windows (slots) for trains at specified reference points (time windows are illustrated in Figure 3 above). These predicted time windows for all trains are the input for the rescheduling system.

It is very important that the rescheduling process be carried out within the shortest possible amount of time. Therefore, a chronological multistage method is a good strategy for rescheduling. This approach consists of developing a 'good' new production plan quickly, although it would be based on limited information. For example, the new production plan could be developed with pre-defined conditions based on the particular event type. As soon as actual information is available, a more precise and powerful production plan could be made on the basis of detailed predictions. This is illustrated in the "prediction of future behavior" row in Figure 5.

2.2.3 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 Burkolter et al. 2005 or Wuest 2006 and by other researchers (e.g. D'Ariano 2005, Wegele et al. 2006 or Weston et al. 2006). 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; 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.

2.2.4 Problem Management Process

All railroads have sets of procedures that actors use to address problems. These problem management processes, which operate in parallel with 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.

2.3 Developing and Implementing New Production Plans

As outlined above, this research project has two components: developing re-scheduling algorithms, and analyzing how these new algorithms can be most effectively implemented using the co-production framework. There are many questions that still must be answered regarding how, specifically, the re-scheduling algorithms should be implemented, these issues are outlined in this section.

Once appropriate re-scheduling 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 re-scheduling process.

In both approaches to generating new production plans (event-driven and periodic), a further question is whether 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 than an optimal production plan which is very unstable.

2.4 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 in 2005, show that operators using this type of display cross reference points very accurately (+/- 15 seconds).

3. Network Classification

The co-production rescheduling process can be most effectively used it the railway network is divided into areas with excess capacity and areas with no excess capacity. This section outlines how using this network helps optimize the rescheduling process.

3.1 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 railroads.
- 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 and Schaffer in 2003, is used in the planning of timetables and schedules, but is generally not formally defined. The co-production framework formally defines and uses this network division to optimize the rescheduling process (outlined below).

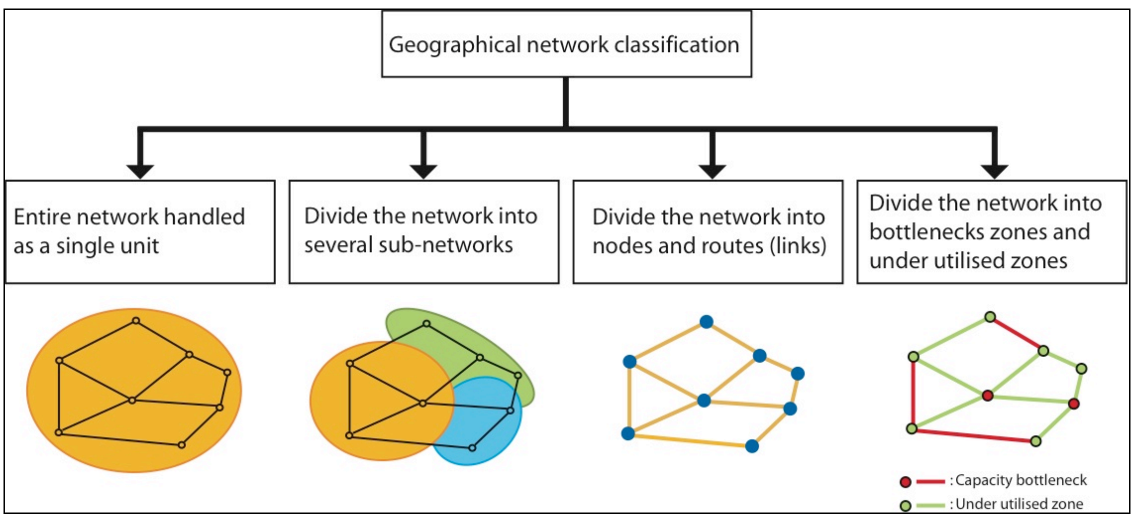

Figure 6: Network classification schemes.

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 is a particularly good way of dividing the network to take advantage of the co-production framework.

3.2 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 co-production framework is designed to provide this type of time and speed information to all affected parties in the network.

The co-production framework resembles air traffic control in the sense that it is based on spacing the arrival of trains at gateways (air traffic: airplanes at airports) by providing operators with detailed time-space trajectory information. However, the railway environment is slightly more complicated since trains have fewer degrees of freedom (i.e. they can only operate on the tracks) and their performance characteristics are more limited (i.e. they can not accelerate or decelerate as quickly as airplanes).

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 co-production 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 co-production 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.

4. Implementing the Co-production 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. This approach is called co-production (since it is based on cooperation of infrastructure operators and train operating companies). The co-production approach can be described as a superposition of two control loops as illustrated in Figure 7.

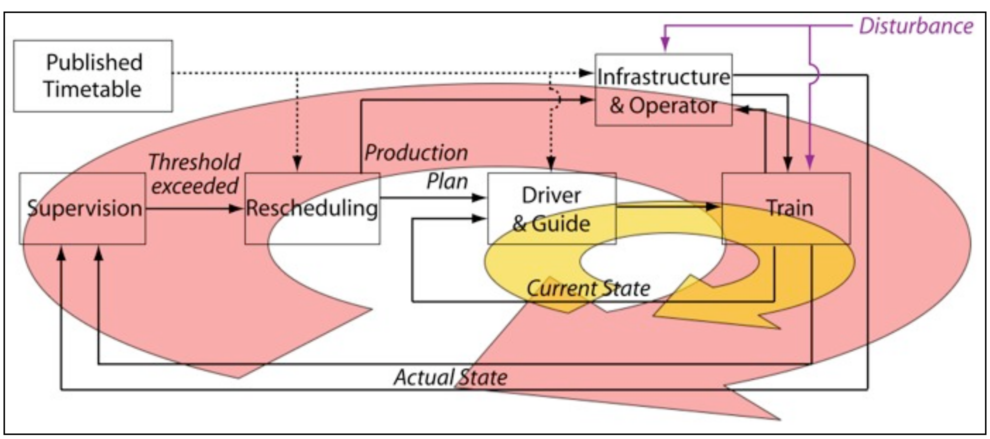

Figure 7: Co-production framework

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.

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 co-production 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 consuming) in the condensation areas. Since most of the additional capacity and schedule stability gained through co-production can be obtained using the existing infrastructure elements and available technology, the approach is extremely cost effective.

5. Simulation Analysis Results

As part of the research project simulations of the co-production 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.

In the simulation OpenTrack allowed uses to selectively control trains (for example exact speed limits can be set for pre-defined sections), which made it possible to directly model the inner production loop (described above). The optimization process, which corresponds to the external rescheduling loop, was modeled manually by performing repeated simulations. This process can be compared with the task of defining a highly dense schedule for a default scenario offline. This approach, described in Luethi 2005, can be regarded as a closed control loop.

The simulation was completed for a network section extending about 25 kilometers around the city of Lucerne and based on the 2005 timetable. This timetable is illustrated graphically in Figure 8 (the illustration is based on the Viriato visual timetable display from SMA 2005, note that in this representation lines indicate train movements not tracks). Lucerne's dead-end station has 10 platforms but just 2 tracks connecting it to the rest of the network.

Results of the simulation showed that the co-production approach can significantly increase capacity and schedule stability. The simulation showed that schedule reserve (buffer) times in the Lucerne station bottleneck area could optimally be reduced by up to 30 seconds per train, although a more realistic scenario (assuming a greater variation in train trajectories) reserve times could be reduced by about 15-20 seconds per train. The headway time in Lucerne is between 90 and 110 seconds. Consequently, capacity could be increased up to 20% without significant stability problems using the new framework with the existing signaling system.

Additionally, the new framework allows system operators to selective when and how they will implement the integrated rescheduling process. Simulations showed, that total delay can be reduced by about 50 – 80% with the assumption of a very fast rescheduling process in combination with a high degree of train control.

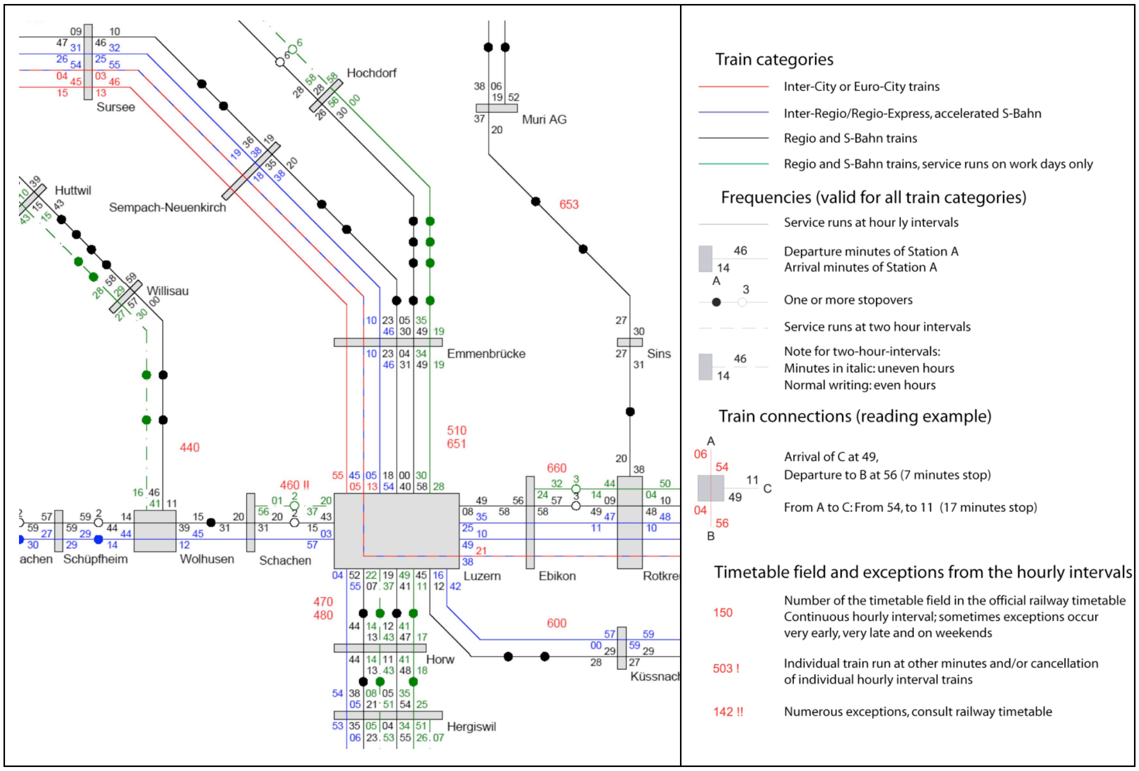

Figure 8: Clock face timetable for Lucerne station, Source: SMA und Partner AG Zurich, Timetable 2005

The simulations were also used to evaluate the impacts of certain constraints on the co-production 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 co-production approach. An especially important finding was the relationship between rescheduling process duration and total system delay; the results (illustrated in Figure 9) 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.

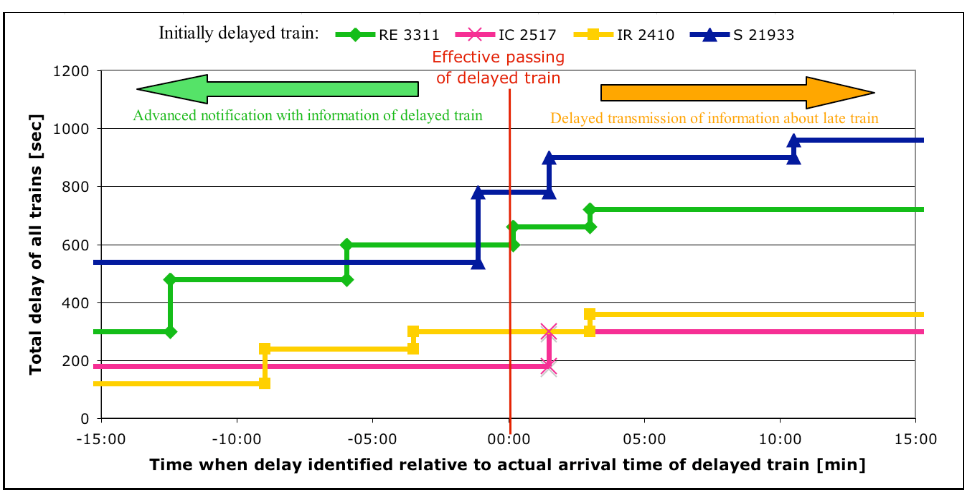

Figure 9: Influence of rescheduling duration on total delay.

6. Conclusions and Next Steps

This paper has shown that rescheduling, in combination with train control, represents a promising low cost approach for increasing capacity and stability of heavily used mixed traffic railroad networks. 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 can the full benefit of rescheduling be achieved.

The next steps in this research project will focus on developing techniques that minimize the time necessary to complete all time relevant sub-processes and real-time algorithms such that the whole rescheduling process is executable within shortest possible time. As part of this effort, the research will investigate the dependencies between topology, rescheduling duration, timetable density, rescheduling reason and the overall performance (total delay) in more detail. Finally, new ideas for transmitting new production information (timetables) to affected actors and interfaces that help them implement these new plans will be developed and evaluated.

The research will continue to use the Lucerne station area as a pilot project area for analysis and evaluation of the co-production approach. This will help show 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.

References

Burkolter, D., Herrmann, T. and Caimi, G. (2005) 'Generating Dense Railroad Schedules', *Advanced OR and AI Methods in Transportation*, Publishing House of Poznan University of Technology, pp. 290-297.

D'Ariano, A. and Pranzo, M. (2005) 'Conflict Resolution and Train Speed Co-ordination for Solving Timetable Perturbations'. Proceedings of the 1st International Seminar on Railroad Operations Modelling and Analysis, Delft, 2005

Eichenberger P. (2007) 'Kapazitätssteigerung durch ETCS', Signal und Draht 03/2007, pp. 6-14.

Fenix, J., Graffagnino, T., Sagot, J.-C. and Valot, C. (2005) 'User centred design applied to increase timetable stability'. Proceedings of the 36th Tagung Moderne Schienenfahrzeuge, Graz, 2005

Laube F., Schaffer H. (2003) 'Specific working results about strategic network division for the PULS 90 project', SBB Bern, 2003.

Laube F., Schaffer H. (2006) 'Specific working results about the rescheduling process and phases for the PULS 90 project', SBB Bern, 2006.

Luethi, M., Huerlimann, D. and Nash, A. (2005) 'Understanding the Timetabling Planning Process as a Closed Control Loop'. Proceedings of the 1st International Seminar on Railroad Operations Modelling and Analysis, Delft, 2005

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

Middelkoop D. and Bouwman M. (2001) 'SIMONE: Large scale train network simulations'. Proceedings of the 2001 Winter Simulation Conference, 2001, pp. 1042-1047

Nash, A. and Huerlimann, D. (2004) 'Railroad simulation using OpenTrack', in Allan, J., Hill, R.J., Brebbia, C.A., Sciutto, G., Sone, S. (eds.), *Computers in Railroads IX*, pp. 45-54, WIT Press, Southampton.

Nash, A. and Weidmann, U. (2006) 'Introducing New Commuter Rail Service on Busy Routes – Case Study: Stadtbahn Zug', Paper presented at 85th Transportation Research Board Annual Meeting, Washington DC, January 2006

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

Savio S. and Masics A. (2004) COMBINE 2, Final Report, EC Contract No: IST-2001-34705, 2004.

SBB (2005), Swiss Federal Railways, *Annual Report 2005*, Bern.

SMA und Partner AG (2004) 'Fahrplan Bahn 2000 1. Etappe', Fahrplan 2005, Zurich

UIC (2004) Capacity, *UIC-Codex 406*, International Union of Railroads, Paris, June 2004.

Ullius, M. (2005) 'Delay Analysis of Rail 2005 1st Phase Using OpenTimeTable'. Proceedings of the 1st International Seminar on Railroad Operations Modelling and Analysis, Delft, 2005.

Wegele, S., Slovak, R. and Schnieder, E. (2006) 'Real-Time Optimisation for Dispatching in Rail Transport'. Proceedings of the 14th International Symposium 'Towards the Competitive Rail system in Europe' Eurnex, Zel 2006, Zilina / Slovakia, pp. 80-89

Weston, P., Goodman, C., Takagi, R., Bouch, C., Armstrong, J. and Preston, J. (2006) 'Minimising Train Delay in a Mixed-Traffic Railroad Network with Consideration of Passenger Delay'. Proceedings of 7th World Congress on Railroad Research, Montreal, 2006.

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