Structure and simulation evaluation of an integrated real-time rescheduling system for railway networks

Conference Paper

Author(s):
Luethi, Marco; Medeossi, Giorgio; Nash, Andrew

Publication date:
2009

Permanent link:
https://doi.org/10.3929/ethz-b-000156910

Originally published in:
Networks and Spatial Economics 9(1), https://doi.org/10.1007/s11067-008-9085-4

Netw Spat Econ (2009) 9:103–121
DOI 10.1007/s11067-008-9085-4

Structure and Simulation Evaluation of an Integrated Real-Time Rescheduling System for Railway Networks

Marco Luethi · Giorgio Medeossi · Andrew Nash

Abstract A critical problem faced by railways is how to increase capacity without investing heavily in infrastructure and impacting on schedule reliability. One way of increasing capacity is to reduce the buffer time added to timetables. Buffer time is used to reduce the impact of train delays on overall network reliability. While reducing buffer times can increase capacity, it also means that small delays to a single train can propagate quickly through the system causing knock-on delays to trains impacted by the delayed train. The Swiss Federal Railways (SBB) and Swiss Federal Institute of Technology (ETH) are researching a new approach for real-time train rescheduling that could enable buffer times to be reduced without impacting schedule reliability. This approach is based on the idea that if trains can be efficiently rescheduled to address delays, then less buffer time is needed to maintain the same level of system schedule reliability. The proposed approach combines a rescheduling algorithm with very accurate train operations (using a driver-machine interface). This paper describes the proposed approach, some system characteristics that improve its efficiency, and results of a microscopic simulation completed to help show the effectiveness of this new approach. The results demonstrate that the proposed integrated real-time rescheduling system enables capacity to be increased and may reduce knock-on delays. The results also clearly showed the importance of accurate train operations on the rescheduling system's effectiveness.

Keywords Real-time rescheduling · Microscopic railway simulation · Rail traffic operation

---

M. Luethi (✉)
Institute for Transport Planning and Systems, Swiss Federal Institute of Technology, Zurich, Switzerland
e-mail: Luethi@ivt.baug.ethz.ch

G. Medeossi
Department of Civil Engineering, University of Trieste, Trieste, Italy

A. Nash
Vienna Transport Strategies, Vienna, Austria

1 Introduction

Railways must become more efficient if they are to be successful in today's highly competitive transport market. One way of becoming more efficient is to increase service frequency, but many railways are already operating at or near capacity and so adding trains would increase unreliability—thus making the railway less efficient. This paper describes an approach for rescheduling trains in real-time that will increase capacity without reducing reliability.

1.1 Switzerland's Bahn-2000 plan

In the mid-1980s Swiss cantons rejected plans for a new high-speed route on the main east–west axis across the country. After defeat of this plan, the Swiss Federal Railways (SBB) adopted a new rail strategy, called Bahn 2000, based on connecting the entire country with an integrated clock-face timetable. The Bahn 2000 infrastructure plan was based on providing the minimal level of investment that would allow the timetable to be operated.

The Bahn 2000 plan was gradually implemented and in December 2004 a major part was put into service. Many routes are now operated on a 30-minute frequency pattern throughout the day. The integrated clock-face timetable provides an optimal timed transfer system for almost the entire country and results in high accessibility and generally shorter travel times for passengers. The service has been extremely successful at attracting more passengers to rail service and demand is expected to increase as additional elements of the Bahn 2000 are completed.

The main problem with Bahn 2000 is that the integrated clock-face timetable means that many trains arrive at and depart from main stations in a short time interval. This means that capacity in these critical locations is at a premium. The increased service has also created capacity constraints at other locations on the Swiss railway network. These capacity problems are compounded by the need for trains to arrive at stations in time for passengers to transfer to connecting trains.

The research project was designed to evaluate the ability of a new method of real-time rescheduling to reduce the impacts of delays on system-wide operations. The SBB is especially interested in this research, given the degree to which it relies on close connections between trains, but increasing reliability is important for all railways, not least because it enables railways to increase service while maintaining reliability.

1.2 Increasing railway network capacity

There are three main ways how railways can increase capacity. They are:

- Infrastructure—build new infrastructure (tracks, junctions, flyovers, etc.);
- Signalling—reduce train headways by reducing block length or introducing more advanced signal systems (e.g. higher levels of European Train Control System, ETCS) (Eichenberger 2007); and
- Operations—reduce train headways by reducing buffer times introduced in the schedule to maintain reliability or by harmonising travelling speed of trains.

The infrastructure and signalling options are expensive and complicated; therefore the SBB is looking for ways to increase service by reducing buffer times.

There are three types of buffer times added to train operating schedules: headway buffers between two consecutive trains, dwell time buffers and running time supplements. Headway buffer times are used to stabilise the system after an interruption, and give the dispatchers time to react. This enables the dispatchers to develop and implement strategies to reduce knock-on delays. Running time supplements are used to reduce the impact of running time variations caused by changing weather conditions or varying train dynamics; it also helps reduce knock-on delays. Reducing buffer times and running time supplement results in a denser level of rail traffic.

The disadvantage of reducing buffer times to add additional trains is that it increases the number of interdependencies between train routes and ultimately increases the number of potential conflicts. Consequently, a single small disturbance can have large impacts on the whole network. Furthermore, the increased number of trains makes it more difficult for human dispatchers to identify optimal strategies for reducing knock-on delays quickly, thus making it difficult to prevent delays from propagating throughout the network. The problem is even worse for integrated clock-face timetables, since train connections at stations are broken to stabilise the system, resulting in passengers missing their connections.

Given their reliance on the integrated clock-face timetable, the SBB is especially interested in examining new methods, ideas and technology for improving both capacity and service quality together. One possible solution is an integrated real-time rescheduling system combining new railway operational strategies with technology. This paper describes initial research on this approach.

The key element of this integrated real-time rescheduling system is that the new timetables and their execution (i.e. driving the trains) must be more accurate than they are today (Stalder et al. 2003). Under this system, when a delay occurs, the rescheduling system generates an extremely accurate timetable that is designed to minimise knock-on delays throughout the system, and that the train can be operated within those parameters.

This paper describes such a system and its benefits. The research was designed to answer the following three questions:

- What are the potential benefits for capacity and stability of using the integrated real-time rescheduling system (for a specific area)?
- What factors have a significant impact on the rescheduling process's overall performance?
- What level of timetable and driving accuracy is needed to most effectively use the new approach?

The next section of this paper describes real-time rescheduling and the proposed rescheduling system's structure; this is followed by a case study of system application (simulation of Lucerne station), and finally conclusions including a discussion of results regarding the research questions.

2 Real-time rescheduling

Rescheduling a railway timetable is extremely complex. The problem is ideal for computers, but it has been too complicated so far to solve the problem in real-time. Today rescheduling is done by human dispatchers. They are generally limited to taking action only after a delay has taken place and, in the short term, have only limited possible corrective actions available including reordering and rerouting trains, or modifying the timetable (assigning new departure and dwell times at stations). Given these constraints, rescheduling decisions made by human dispatchers are often suboptimal and time-consuming.

Consequently, an automatic rescheduling system (calculating an optimal timetable in real-time based on predefined criterions as minimal weighted total delay or minimal amount of missed connections) is needed to support the dispatchers. These new timetables must be based on the actual system state and accurate traffic prediction (i.e. what will the system state be when the new timetable can be implemented?).

This section outlines the main factors influencing the effectiveness of real-time rescheduling and the proposed integrated real-time rescheduling system approach.

2.1 Factors influencing rescheduling effectiveness

There are two main factors influencing the effectiveness of rescheduling processes, first, the effectiveness of the conflict resolution algorithms used to develop the new timetables; and, second, the process for applying these algorithms.

Research on rescheduling algorithms has been underway for many years (Burkolter et al. 2005; D'Ariano et al. 2007; Fay 1999; Jacobs 2004; Wegele and Schnieder 2004). In order to successfully use these rescheduling algorithms in dense railway networks with heterogeneous rail traffic, it is necessary to analyse the whole rail operation process to determine how new schedules can be most efficiently implemented. There are three main problems to address in applying the algorithms.

The first problem is system observe-ability. This refers to the time it takes before a delay can be identified. This problem is minimised in a continuous train detection system, but in non-continuous systems, some time is lost before a delay can be detected. In the worst case, position detection with an intermittent fixed block signalling system, a broken-down or heavily delayed train could mean that no information about the train is provided. This not only delays the rescheduling process, it also makes it difficult to identify the primary delay cause and thus to define an appropriate plan of action.

The second problem is the cumbersome and time-consuming communication of dispatching instructions. Dispatching instructions must be communicated to many different people at many different locations including infrastructure operators, train drivers and train guards. Generally infrastructure operators are called by phone, while in most cases neither the train drivers nor the train guard are informed directly.

Since train drivers have no direct information, they do not know if or how they should adjust their driving behaviour (speed) to prevent conflicts or optimise traffic flow. This means they may inadvertently cause additional delays (e.g. if they drive a bit slower than normal they might not need to stop at a red signal caused by a delayed train and then not need additional time to accelerate from a stop).

Figure 1 illustrates schematically the topology for a train running towards a capacity critical bottleneck area with two time-reference points. The first time-reference point is located in front of the last distant signal outside the capacity critical area, and the second time-reference point lies within the capacity critical area. Figure 2 shows the consequences of the moment a train approaches the capacity critical area in a time-distance diagram for five different arrival times. In the cases A, B and C, the train passes the distant signal when it is closed thus forcing it to slow down. In case A the train fully stops in front of the closed main signal, whereas in the cases B and C, trains do not fully stop. In cases D and E trains pass the distant signal indicating a movement authority allowing them to proceed with their maximal allowed track speed.

Figure 3 shows the nonlinear influence on delay (ordinate: moment train passes the second time-reference point TRP2) depending on the passing time at a reference point in front of the capacity critical section (TRP1, assigned to the abscissa). The final delay (which is similar to the passing time at the time reference point TRP2 within the critical section) depends on when the train passes the first reference point. Of course this delay is also influenced by many other factors including operating rules, the signal aspect update (or the view distance of the main signal), the train dynamics, the desired track speed and the position of the distant and the main signal. Depending on these factors, the time lost because of stopping and reaccelerating could be up to several minutes. However, it must also be stated that, depending on the specific influence factors, it is possible that the earliest passing time of the second reference point can be achieved either for the case when the train passes the distant signal for a set route (case D) or when the train follows an appropriate driving strategy containing a deceleration and acceleration phase when passing the distant signal with closed aspect (case B) (Albrecht 2007).

The third problem in applying rescheduling to complex train networks is the large variation in train running time on route and departure times at stations. This variation creates uncertainty in the prediction of train running times and results in sub-optimal dispatching decisions that unnecessarily delays following trains. In practice, timetables are developed that include extra buffer time to minimise the influence of these variations.

In summary, achieving the full effectiveness of an automatic real-time rescheduling system means not only developing fast and optimal algorithms, but also requires development of very exact schedules (i.e. with an accuracy of 1–5 s), precise train

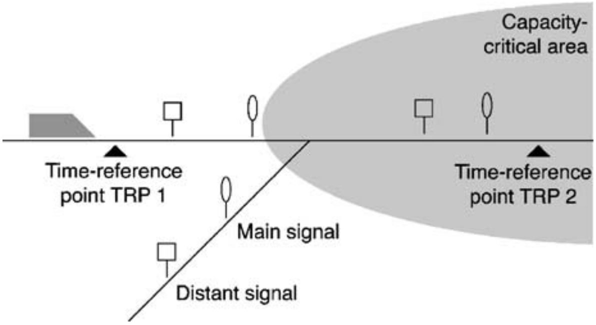


Fig. 1 Schematic topology to analyse the delay effects for trains approaching a possible conflict situation

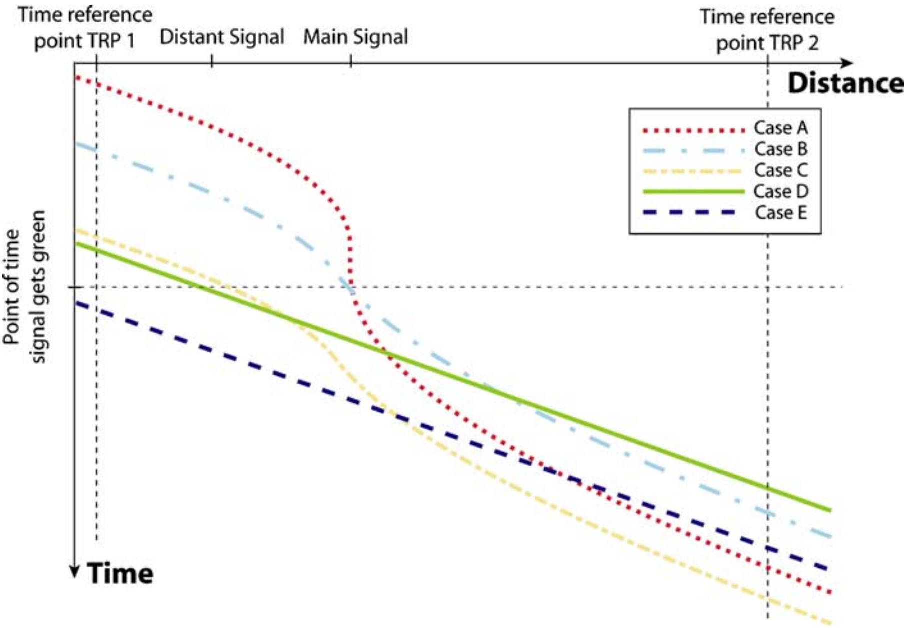


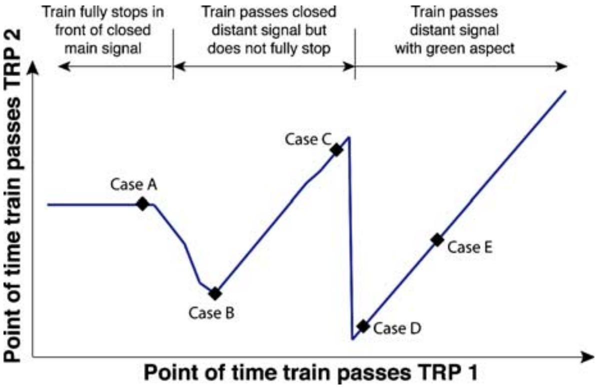
Fig. 2 Time-distance diagram for five different passing times a train approaches a conflict situation

location information, and accurate train operation (i.e. driving the train at exact speeds and crossing time points at exact times based on dynamically changing time-space trajectories). Regarding these requirements, a new integrated real-time rescheduling system is under development. Thereon, the large train heterogeneity, safety requirements and the existing conventional fixed block signalling and interlocking systems have to be respected.

2.2 Integrated real-time rescheduling model

This section describes a new integrated real-time rescheduling model developed as part of the research project. The new model combines real-time rescheduling with

Fig. 3 Example delay effect for trains approaching a possible conflict situation



highly accurate train operation. This combination results in an integrated real-time rescheduling system. In order to better explain this new process, the terms rescheduling and integrated real-time rescheduling are defined below.

Rescheduling is the process of updating an existing production plan in response to disruptions or other changes (Vieira et al. 2003). In the railway industry, the production plan is the schedule or timetable; essentially the railway is producing train movements. The main elements of this production plan (or timetable) in the proposed rescheduling model are:

- Reference times (timetable) for all trains for defined points in the network (stations and on open track);
- Train routings (globally and locally);
- Resources (staff, rolling stock) assigned to the production process;
- Implementation rules or instructions for accurate production (e.g. reference speed or door closing times); and,
- Tolerance bandwidth within which a train is regarded as on time.

Integrated real-time rescheduling is the process whereby new production plans are developed in real-time following a delay, are automatically communicated to all necessary actors, and are accurately executed with the help of supporting tools (e.g. driver–machine-interface). In terms of railways, the affected actors are infrastructure operators, train drivers, train guards and passengers.

Mazzarello and Ottaviano (2007) have visualised the real-time rescheduling and driving process as a single closed control loop under ETCS Level 3. However, in cases where ETCS Level 3 and continuous position detection are not available, the integrated real-time rescheduling and driving process can be visualised as the superposition of two feedback control loops, schematically illustrated in Fig. 4. The outer loop is the rescheduling system and the inner loop is the driving process. The outer rescheduling loop consists of delay recognition, calculation of a new production plan and information update. The inner control loop assures that trains are operated within their bandwidth limits.

This two-loop rescheduling system has two distinctions from other rescheduling system approaches. First, the rescheduling system output is a production plan; this

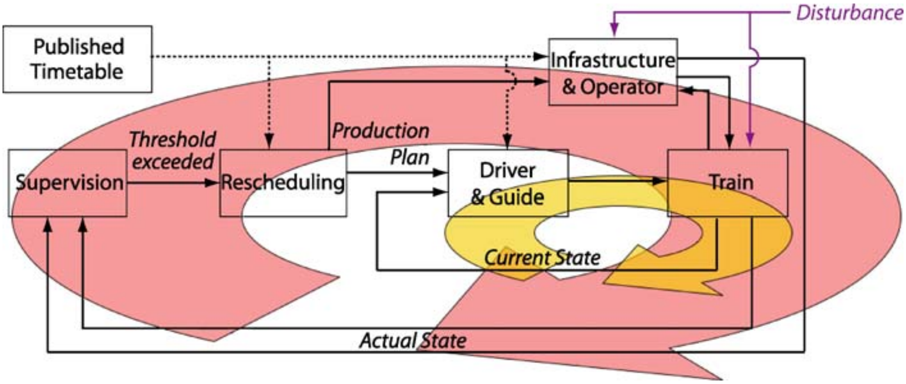


Fig. 4 General model of the integrated real-time rescheduling framework

means it contains much more information than a normal timetable and that the new production plan (schedule) must be conflict-free in all cases. The conflict-free constraint means that no unintended stopping or slowing down of trains should occur when they drive within their given tolerance bandwidth limits. Of course, planned slowing down or stopping of trains as part of the production plan remains possible as long as it is intended.

The second difference is in the dispatching process. The normal dispatching process is: (1) predict the future train movements; (2) detect possible conflicts based on the prediction; and (3) solve the conflicts. In contrast to this, the integrated real-time rescheduling system uses a train-specific tolerance bandwidth (that can vary depending on the train's position and attributes) to detect delays. After the detection of a threshold exceedance, the rescheduling algorithm generates a new production plan (schedule) based on the actual train and infrastructure state. Next, the feasibility of this new production plan is checked before the new schedules are transmitted to the actors. The feasibility check consists of comparing the actual system state to the initially assumed predicted state; if the actual state is within the newly calculated bandwidth the schedule is considered feasible. Figure 5 illustrates the relevant process flow and the actions that take place during the rescheduling process (Laube and Schaffer 2006).

2.3 Improving two-loop rescheduling process effectiveness

The performance of this two-loop approach is optimised if the rail network is strategically divided into bottleneck areas (i.e. areas operating at or near their capacity limit) and non-bottleneck areas (Laube et al. 2007). The SBB uses the term condensation zone for capacity critical areas and compensation zone for non-bottleneck areas.

To optimise system performance, trains should be operated at their maximum allowed speed and with very small buffer times in condensation zones, while in

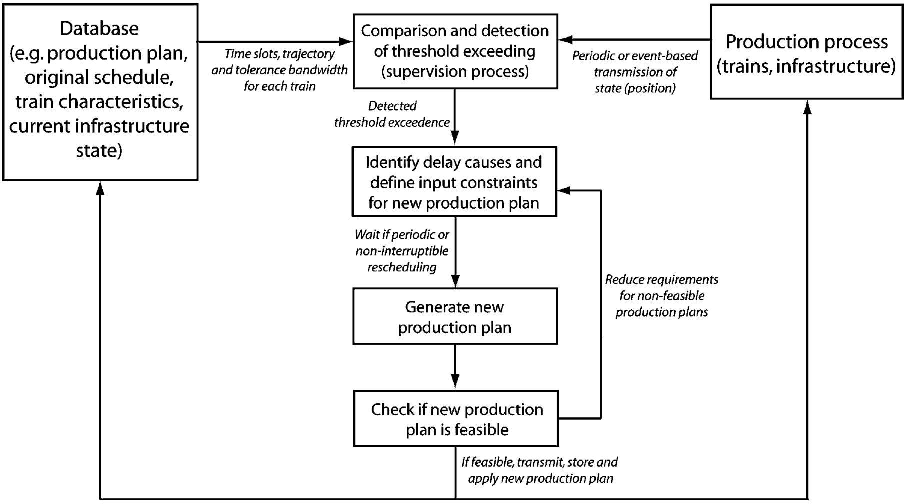


Fig. 5 Time-relevant aspects (actions) of the railway rescheduling process

compensation zones trains should be carefully controlled (slowed down or speeded up) so they arrive at a reference point (i.e. the boundary between a compensation and a condensation zone) at an exact time and with a precisely defined speed. This reduces non-optimal (unintended) signal braking and stops, thereby making it possible to reduce the time and capacity lost due to train acceleration and deceleration in condensation areas. This also makes it possible to avoid the non-linear delay effects due to conflicts (explained in Section 2.1). This approach increases the stability of train operations prediction and reduces the number of exceeded threshold limits thus reducing the rescheduling system's nervousness.

Dividing the network into bottleneck and non-bottleneck zones also helps handle the scheduling problem by creating independent zones (Caimi et al. 2007). Since trains can be operated independently within the zones, it is possible to unite local schedules (within zone) with a global timetable by simply coordinating specific constraints at the zone boundaries. This enables the optimisation to focus on traffic flows within the condensation areas whereas trains operating in the compensation areas can be flexibly controlled with the use of predefined running time supplements (e.g. speeding up or slowing down).

2.4 Improving train operation efficiency

The integrated real-time rescheduling system will require that trains be operated within a tolerance bandwidth of 15–30 s for both running and departing trains. Therefore tools are needed that can provide real-time information about the most current production plan (timetable). Thereby, a larger tolerance bandwidth would unnecessary decrease to possible capacity gain whereas a tolerance bandwidth of less than 15 s appears non-applicable.

A new Driver–Machine-Interface (Fenix et al. 2005; Albrecht et al. 2007) will be developed to enable train drivers to achieve this objective for running trains. The interface will enable drivers to remain in full control of the train; they are simply responsible for insuring that any variation in train trajectory remains within predefined tolerance bandwidth limits. Other types of tools and processes will be necessary to optimise the station departure process so that trains can leave more precisely than they do today. Examples are new dynamic passenger information systems or handhelds for train guards indicating the precise planned departing time.

Figure 5 shows that defining the input constraints assuming the future behaviour of trains and infrastructure is crucial, because a conflict exists between high productivity and rescheduling frequency. All other things being equal, a rescheduling process is more likely to be initiated in the case of schedules with small buffer times and high track occupation. This leads to nervous production behaviour (i.e. frequent development of new timetables) because thresholds are continuously being exceeded which leads to frequent development of new production plans. This should always be avoided. On the other hand, the effectiveness is lowered unnecessarily for large buffer times. This, too, should be avoided. The conflict between nervousness and productivity is therefore a central aspect to consider in the rescheduling process.

A final issue of the rescheduling process is how the inherent inertia and temporal gap between the data measurement, the prediction and constraint definition, the production plan calculation and communication is handled while the dynamic process is going on. The question is: Should a production plan including the tolerance limits be transmitted or not if during the feasibility check a train has exceeded the tolerance bandwidth? The problem is that both, frequent rescheduling and having no valid production plan, should be avoided. Further research, addressing this problem, is ongoing.

3 Simulation studies evaluating the integrated real-time rescheduling system

The real-time rescheduling system approach described above was tested using a simulation study to help estimate its effectiveness and identify possible improvements. These studies were completed using the microscopic rail simulation program OpenTrack (Nash and Huerlimann 2004).

3.1 Simulation of Lucerne station area

The simulation was carried out on the SBB network approximately 15–25 km around Lucerne, located in the centre of Switzerland (see Fig. 6). This area was selected by the SBB for research on rescheduling methods and adjusting railway operation processes since it is a critical bottleneck in the Swiss rail network. The bottleneck area extends over about 4 km, and infrastructure extensions are not

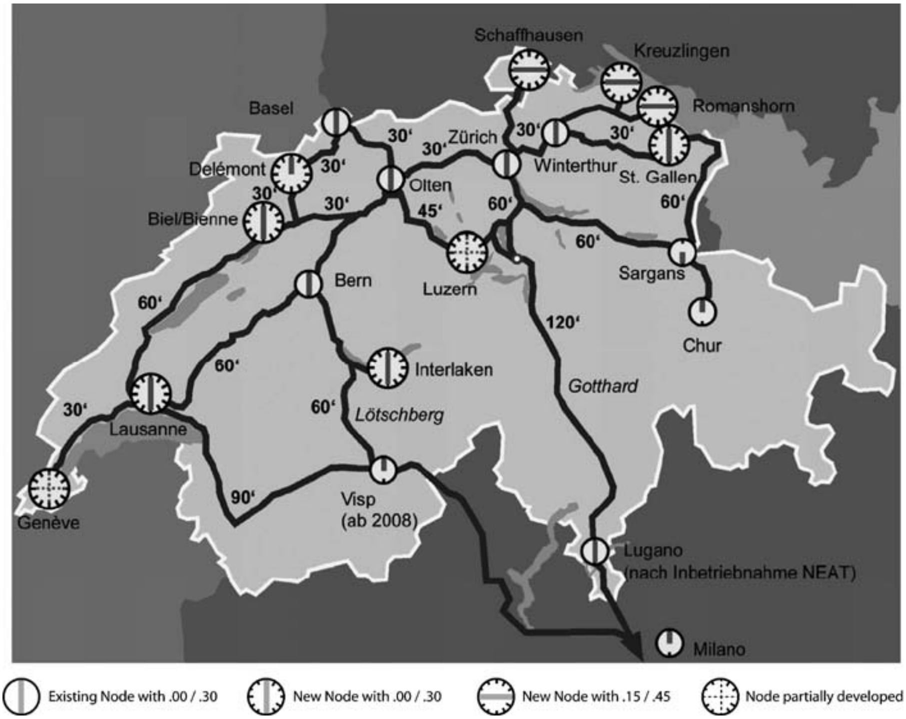


Fig. 6 Future main line network and system nodes of Switzerland (source: Swiss Federal Gazette—Bundesblatt 4/2007, p. 7716)

possible within this area. However, there is an increasing demand in the surrounding area of Lucerne resulting in the request of additional services. Therefore, Lucerne is an excellent area to test the possibilities and limits of the new systems.

Lucerne station serves 30 standard gauge passenger trains per hour, in a terminal (dead-end station) configuration. It has ten station tracks for standard gauge passenger trains, which are connected to the network by only two tracks (see Figs. 7 and 8 for the track topology). Lucerne is not an integrated clock-face station, instead trains arrive and depart distributed over the whole hour. The train intervals vary between 30 and 60 min depending on the lines. Table 1 presents an overview of the timetable of the busiest half hour. The simulation analysis included shunting movements, but neglected the narrow gauge trains serving Lucerne since they do not significantly impact the network of standard gauge tracks.

A conventional fixed block track signalling system is used in the Lucerne station area. The train headway is between 90 and 130 s (depending on train category and direction).

The first step in the study was to identify the most common delay scenarios and typical delay times in the Lucerne station area. This was done using detailed operational data provided by the SBB.

The research evaluated two aspects of the proposed real-time rescheduling approach:

- The first case focused on the outer rescheduling loop; specifically, it investigated the impact of the point in time when the rescheduling is initiated and the rescheduling process duration.
- The second case focused on the inner rescheduling loop; specifically, it investigated the impact of production accuracy (i.e. how accurately the train operator could control the train).

In both cases small original delays were evaluated since the research focus was to analyse the flow and capacity optimisation *within* a bottleneck (condensation) area. This meant that coordination impacts with neighbouring condensation areas could be neglected. The following sections outline the research results.

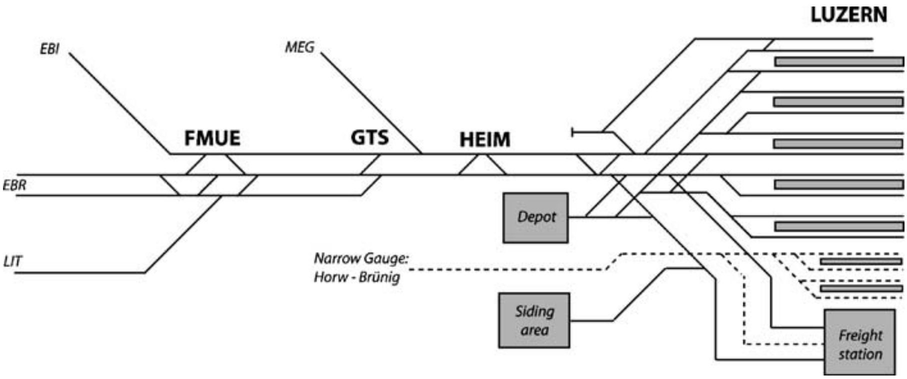


Fig. 7 Topology of the condensation area around Lucerne

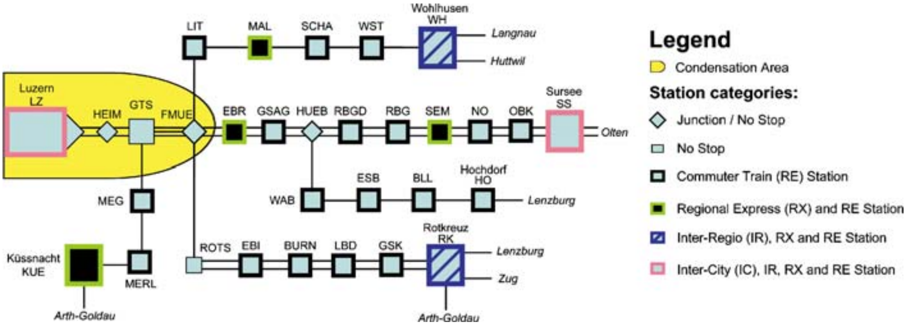


Fig. 8 Aggregated topology of the rail network around Lucerne

3.2 Impact of rescheduling time point and rescheduling process duration on delay

The first set of simulations tested the impacts of the point in time when the threshold exceedence was detected and the rescheduling process duration on the real-time rescheduling effectiveness. The evaluation was done by comparing the total knock-on delays under three cases:

– with no rescheduling,
– with rescheduling but without the ability to speed up trains,
– and with rescheduling including the ability to speed up trains.

Table 1 Timetable for the studies (based on the 2005 timetable; additional trains added to the schedule to study the density effects are indicated with a Z as the train ID)

Train ID	Type of train	Departing condensation area	Arriving (next) condensation area	Platform used in Lucerne	Planned arrival or departure time in Lucerne
2518	Interregional train	Lucerne	Sursee	6	9:55
3320	Interregional train	Lucerne	Wolhusen	5	9:57
21933	Suburban train	Hochdorf	Lucerne	11	9:58
21938	Suburban train	Lucerne	Hochdorf	10	10:00
3311	Interregional train	Wolhusen	Lucerne	5	10:02
Z 1	Suburban train	Lucerne	Rotkreuz	8	10:03
2517	Interregional train	Sursee	Lucerne	6	10:04
3568	Regional train	Lucerne	Sursee	3	10:05
21135	Suburban train	Rotkreuz	Lucerne	8	10:07
Z 2	Suburban train	Lucerne	Wolhusen	4	10:07
2328	Intercity train	Lucerne	Rotkreuz	7	10:10
21339	Suburban train	Lucerne	Kuessnacht	11	10:12
111	Intercity train	Sursee	Lucerne	3	10:12
2410	Interregional train	Kuessnacht	Lucerne	5	10:14
21638	Suburban train	Lucerne	Wolhusen	9	10:15
Z 3	Suburban train	Sursee	Lucerne	4	10:18
21838	Suburban train	Lucerne	Sursee	2	10:18
111	Intercity train	Lucerne	Rotkreuz	3	10:21
90814	Freight train	Lucerne	Rotkreuz	1	10:23
2321	Intercity train	Rotkreuz	Lucerne	11	10:24

For this purpose, eight different scenarios were run; in each scenario a single train was assigned a small original delay between 2 and 4 min.

This range of small original delays was found to be very common in the analysis of SBB operational data. They occur both when trains are delayed entering the bottleneck area or as they are leaving the station. These delays are critical because delays occurring close to the bottleneck area or when trains should be leaving the station have to be detected and solved within a very short time. They directly impact the decisions dispatchers make regarding whether connecting trains should wait or if connections should be broken. The operational analysis showed that the rescheduling measures for such delays vary significantly between dispatchers. Consequently, there is a large potential for optimisation through automated rescheduling systems. In contrast, delays of more than 5 min are normally identified earlier in the process and therefore give the dispatchers more time to react.

In this simulation, three actions could be taken to address a delay:

- Trains could be retimed (this means that trains could operate faster or slower in compensation (non-bottleneck) zones);
- Train departure times from stations could be delayed; and
- Trains could be reordered and rerouted within the condensation (bottleneck) area, but station platforms could not be changed.

In the simulation the amount that trains could be speeded up varied depending on the stopping patterns and the running time supplements (which for the SBB are about 10%, although they vary for all lines). This meant that interregional and intercity trains could be speeded up by a maximum of 1 to 2 min. Increasing train speed to minimise total delay is only possible up to a certain moment. Thereafter, only rerouting, reordering and delaying trains are possible.

The analysis included several constraints:

- First, all actions taken to address delay were required to generate conflict free train paths; this enabled the researchers to neglect non-linear delay effects that would be caused by conflicts. This was done using a time-space discretisation model (Roos 2006; Wuest 2006) with fixed time slots based on predesigned train paths.
- Second, it was assumed that the trains run within their given tolerance bandwidth of 15 s in the case of speeding up or slowing down.
- Third, the delayed train could not be speeded up.

Table 2 presents the results of the first simulation study. It shows the initial delay, the total knock-on delay without rescheduling, and the percentage reduction in knock-on delay with application of rescheduling (rerouting, reordering and retiming), first without speeding up trains and second with speeding up trains. The left portion of the table presents results for the actual schedule and the right portion for the schedule with three additional trains (dense timetable). Without having any delays, these three additional trains would not influence the regular trains.

The following insights can be drawn from the simulation results shown in Table 2:

- The total knock-on delay increases stepwise with respect to initial delay; this means that significant increases in total delay are possible with only a small increase in original train delay.

Table 2 Reduction of knock-on delays with rescheduling (rerouting, reordering and speeding up trains) for regular and dense timetable for a single initially delayed train

Initially delayed train [train-number]	Original delay [s]	Regular timetable			Dense timetable		
		Total knock-on delays without rescheduling [s]	Reduction knock-on delays with rescheduling but no speed up of trains [%]	Reduction knock-on delays with Rescheduling including speed up of trains [%]	Total knock-on delays without rescheduling [s]	Reduction knock-on delays with Rescheduling but no speed up of trains [%]	Reduction knock-on delays with Rescheduling including speed up of trains [%]
2518	120	660	55	82	1620	81	93
3311	120	600	20	70	1860	55	68
3311	240	1500	72	72	2400	33	33
2517	120	180	67	67	900	60	60
2517	240	960	94	94	2820	94	94
111	120	240	25	100	1020	59	100
3320	120	540	67	89	1860	84	97
21933	180	780	23	54	2040	53	53

- The moment when a threshold exceedence is detected and the duration of the rescheduling process (time until a new production plan is applied) has an enormous impact on the total knock-on delay. The earlier a delay is detected and the faster the rescheduling process, the more options for rescheduling are possible and thus the total delays become less.
- Speeding up trains can be a very effective measure for reducing total knock-on delay although not always. The strategy of providing an intelligent and non-linear distribution of running time supplements along a train run can be helpful to provide more time for allowing trains to be speeded up.
- In order to use the measure of speeding up trains effectively, information is needed well before the delayed train enters the condensation area. However, this also increases the possibility that another incident or delay will occur in the meantime and further rescheduling will be needed.
- The effectiveness of the integrated real-time rescheduling system to reduce total knock-on delay is highly dependant on the specific circumstances (timetable, delayed train, train routes, topology of the station and tracks before and within the bottleneck area).
- The possible benefit of the integrated real-time rescheduling system is significantly reduced in cases where the delayed train has only a few interdependencies and conflicts with other trains.
- The specific topology of Lucerne where the condensation area is linked with three short singletrack line sections causes intentional slowing down or even stopping of some trains in the condensation area (in the Lucerne condensation area between FMUE and GTS where three tracks exist). In contrast to the original strategy of preventing trains from slowing down or stopping within condensation areas, in the specific case of Lucerne this increases the overall flow and results in a lower total delay.

- The size of the original delay, even small differences of 2 min, can have a large impact on the difference of the resulting total knock-on delay both with and without rescheduling.
- A denser timetable results in a larger number of interdependencies between the trains and in fewer free slots. Thus, in a denser timetable, the original delay propagates to more trains and it takes a longer time to recover from the delays. Consequently, the total knock-on delays without rescheduling are up to 5 times larger for the denser timetable compared to the regular timetable.
- The absolute delay reduction with rescheduling for dense timetables is normally larger than for less dense timetables. In contrast, no general statement is possible for the proportional delay reduction comparing dense or less dense timetables.
- The dense timetable results in less freedom for rerouting and retiming. On the other hand, more trains offer the possibility of more reroutings, but at a much higher level of complexity.
- The rescheduling system must handle freight trains in a comparable way as passenger trains. However, it is very difficult to achieve the accurate prediction and control of freight trains; therefore, freight trains are generally assigned larger tolerance bandwidths. In the future, adaptive identification methods and supporting tools will be needed to help operate and plan freight trains nearly as exactly as passenger trains.

3.3 Impact of production accuracy on delay

The second set of simulations tested the influence of production accuracy on delay. Production accuracy refers to how close to the exact schedule a train is operated. Small time deviations from the schedule occur frequently in normal operations. Two common situations are the variation in running speed on track sections and in dwell times in stations.

A fundamental question in designing a rescheduling system is determining what tolerance bandwidth should be used to compensate for variation in production accuracy. The tolerance bandwidth has a significant impact on both capacity and stability; too large bandwidth results in lost capacity, whereas too small bandwidth results in unreliability. If a railway can be operated very precisely, then the tolerance bandwidth can be small; on the other hand if it is operated imprecisely, the tolerance bandwidth must be large to achieve the same level of reliability.

The first source of inaccurate production is variation in running times. The analysis of Lucerne area operational data showed that train running times are subject to large variations, results that are consistent with other research (Luethi et al. 2005; Yuan et al. 2004). There are three main causes of running time variation: track conditions, train performance and driving behaviour. The first two are outside of the control of the driver while the third is directly under the driver's control. A driver–machine-interface would make it possible to reduce this running time variation by enabling the driver to precisely follow schedules, and this would work even when schedules are changed in the real-time rescheduling process. In a test where drivers were provided with accurate information about the schedule, it was shown that it is possible to control a train such that it can pass a given reference point with a desired speed within ±15 s (Fenix et al. 2005).

The second source of inaccurate production are delays during the station departure process. Research has shown that even after all departure conditions are satisfied (signal is green, departure time has passed, main boarding and alighting process is finished, train and staff is ready), trains still use a significant amount of time until they actually depart. The observed amount of time spent before leaving was on average 27-seconds for trains without a guard (standard deviation 17-seconds) and 35-seconds for trains with a guard (standard deviation 25-seconds) (Johner and Luethi 2007). The main reasons for these delays are runners (late arriving passengers), blocked doors and train staff that does not react promptly (i.e. the driver does not start the train moving exactly when the signal turns green).

In order to analyse the impacts of production accuracy on delay, the Lucerne case study trains were simulated with an adjusted timetable where the headway buffer times were set to 15 s. In the simulation runs, all trains were then given an initial delay based on three uniform distributions with a width of 30, 60 and 90 s respectively. In this simulation the rescheduling method was not used since the objective was to analyse the effects of the production accuracy on the knock-on delays and system stability.

Figure 9 shows the average knock-on delay of the trains, ordered by the planned arrival or departure time at the platform in the Lucerne station. For small variations (maximum 30-seconds), the knock-on delays are negligible and the timetable remains stable. This is because the non-linear delay effects are very low for three reasons: (1) the low permitted speed in the condensation area of Lucerne where the conflicts occur, (2) the good train dynamics and (3) the signalling system.

For larger variations (assigning trains with a stochastic delay of maximal 60 s), a limited increase of the knock-on delay is observed. Thereby, the knock-on delay is 22 s in average. For even larger variations (maximal 90 s), the knock-on delays increase significantly especially for most departures. Though, average consecutive delays are 55 s.

There are two problems caused by inaccurate production. First, inaccurate production may result in a consistent growth of knock-on delays. Second, accurate production is a fundamental precondition for an effective real-time rescheduling system. As shown in Section 3.1, the rescheduling measures have a significant impact on the overall delay. However, if these measures are not implemented precisely, they can result in suboptimal rescheduling. Furthermore, conflict-free operations (avoiding the non-linear delay effects causing knock-on delays) can only be achieved with accurate production (i.e. when the trains running on open tracks and leaving stations are operated within the given tolerance bandwidth). Specific techniques for minimising the variation in daily operations using new methods and technology are underway in other research projects at the SBB.

4 Conclusions and future research

One way of increasing capacity and quality without making significant infrastructure investments is to use advanced technology to improve system efficiency. This

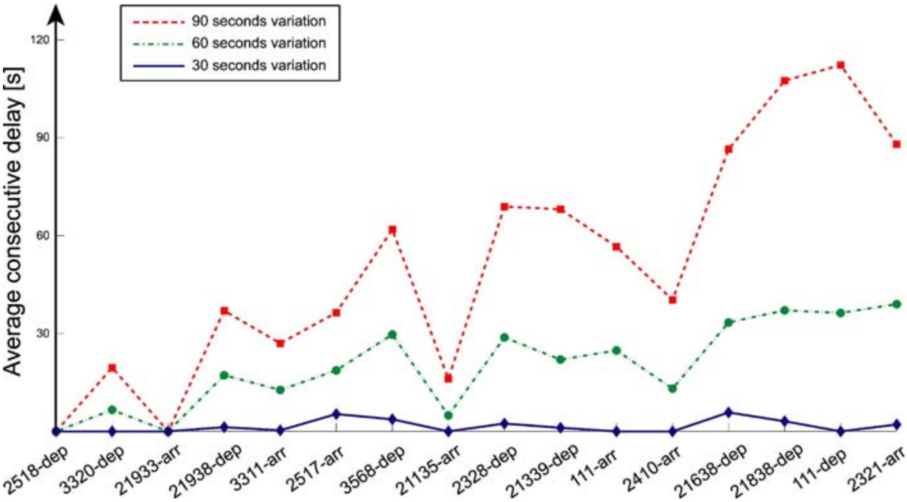


Fig. 9 Average consecutive delay for different temporal variations of delay

research presents an approach that combines real-time train rescheduling with accurate driving of trains to reduce delays and improve efficiency. Simulation experiments show that a real-time rescheduling system could significantly reduce the total knock-on delays even as additional trains are added to the timetable. The earlier a delay can be identified, the more effectively the proposed approach may reduce the total knock-on delays.

The simulations show that the system reliability can be improved if trains were very precisely planned and controlled, for example, arriving at specified times with a certain speed at a specific point (the entry to a bottleneck area) and/or departing from a station very exactly. At present, there is a significant variation in both the segment running time and the station departure time that should be reduced to improve the punctuality. Planning and operating the train movements very precisely would enable a real-time rescheduling system to be most effective.

Finally, the simulations show that the rescheduling approach effectiveness depends upon the specific track topology and schedule. The combination of real-time rescheduling with accurate driving of trains can also improve the railway service in other ways. For example, the continuous information flow to all actors (needed to implement such a system) provides an excellent opportunity for improving the railway's information management system. Similarly, an integrated real-time rescheduling system offers the possibility of introducing additional optimisation goals (e.g. operating the trains to minimise total energy consumption by reducing unintended stops and encouraging smoother driving behaviour) that can enhance railway operations. However, new tools as for automatic communication between all actors or providing on-line information to train drivers require serious investments in hard- and software and implicates also changes from existing operation methods.

Further research should focus on speeding up the rescheduling process by developing intelligent rescheduling algorithms and tools to automatically identify delays and communicate the revised schedule to all involved actors. A challenge is to reduce further the inaccuracy of running trains and improve the efficiency of the station departure process such that the tolerance bandwidths used in scheduling trains are minimised (thus increasing system capacity).

Another key area of research is to determine the size and spatial distribution of optimal running time supplements in a segmented rail network (condensation and compensation areas) in order to provide as much time as possible for taking the best rescheduling measures (i.e. trains can be speeded up to reach the border of a condensation zone at preferred time). This would make the overall system more flexible and improve its ability to react to disruptions without increasing the overall travel times.

Acknowledgements The authors would like to thank the Swiss Federal Railway Infrastructure Division, namely Dr. Felix Laube, Dr. Raimond Wuest, Samuel Roos, Oskar Stalder and Thomas Graffagnino, for funding this research and for their fruitful discussions, and also Prof. Ulrich Weidmann, Prof. Giovanni Longo and Adrian Johner for their cooperation and helpful support. The authors also thank the reviewers for constructive suggestions that helped improving substantially the quality of this paper.

References

Albrecht T (2007) The influence of anticipating train driving on the dispatching process in railway conflict situations. In: Hansen IA, Radke A, Pachl JP, Wendler E (eds) Proceedings of the 2nd International Seminar on Railway Operations Modelling and Analysis, Hannover

Albrecht T, van Luipen J, Hansen IA, Weeda A (2007) Bessere Echtzeitinformationen für Triebfahrzeugführer und Fahrdienstleiter. Eisenbahningenieur 58(6):73–79

Burkolter D, Herrmann T, Caimi G (2005) Generating Dense Railway Schedules, In Publishing House of Poznan University of Technology, Advanced OR and AI Methods in Transportation, pp 290–297

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

D'Ariano A, Pranzo M, Hansen IA (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 99(3):6–14

Fay A (1999) Wissensbasierte Entscheidungsunterstützung für die Disposition im Schienenverkehr. Technical University Braunschweig, VDI Verlag

Fenix J, Graffagnino T, Sagot J-C, Valot C (2005) User centred design applied to increase timetable stability. In Proceedings of the 36. Tagung moderne Schienenfahrzeuge, Graz

Jacobs J (2004) Reducing delays by means of computer-aided 'on-the-spot' rescheduling. In: Allan J, Hill RJ, Brebbia CA, Sciutto G, Sone S (eds) Computers in Railways IX. WIT, Southampton, pp 603–612

Johner A, Luethi M (2007) The train departure process and its impact on the rail network performance. In Proceedings of the 7th Swiss Transport Research Conference, Ascona

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

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

Luethi M, Huerlimann D, Nash A (2005) Understanding the Timetable Planning Process as a Closed Control Loop. In: Hansen IA, Dekking FM, Goverde RMP, Heidergott B, Meester LE (eds) Proceedings of the 1st International Seminar on Railway Operations Modelling and Analysis, Delft

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

Nash A, Huerlimann D (2004) Railroad simulation using OpenTrack. In: Allan J, Hill RJ, Brebbia CA, Sciutto G, Sone S (eds) Computers in Railways IX. WIT, Southampton, pp 45–54

Roos S (2006) Bewertung von Knotenmanagement-Methoden für Eisenbahnen. Master thesis, ETH Zurich

Stalder O, Laube F, Graffagnino T (2003) Increasing Performance of the Rail Network in the Heart of Europe: A Program for the Swiss Federal Railways. In Proceedings for the STECH'03: International Symposium on Speed-up and Service Technology for Railways and Maglev Systems, Tokyo

Vieira GE, Herrmann JW, Lin E (2003) Rescheduling Manufacturing Systems: A Framework of Strategies, Policies, and Methods. Journal of Scheduling 6:39–62

Wegele S, Schnieder E (2004) Automated dispatching of train operations using genetic algorithms. In: Allan J, Hill RJ, Brebbia CA, Sciutto G, Sone S (eds) Computers in Railways IX. WIT, Southampton, pp 775–784

Wuest R (2006) Dynamic rescheduling based on predefined track slots. In Proceedings of 7th World Congress on Railway Research, Montreal

Yuan J, Goverde RMP, Hansen IA (2004) Evaluating stochastic train process time distribution models on the basis of empirical detection data. In: Allan J, Hill RJ, Brebbia CA, Sciutto G, Sone S (eds) Computers in Railways IX. WIT, Southampton, pp 631–640