Doctoral Thesis

Improving the efficiency of heavily used railway networks through integrated real-time rescheduling

Author(s):
Lüthi, Marco

Publication Date:
2009

Permanent Link:
https://doi.org/10.3929/ethz-a-005911895

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
Improving the Efficiency of Heavily Used Railway Networks through Integrated Real-Time Rescheduling

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
MARCO LÜTHI
Master of Science ETH in Electrical Engineering and Information Technology
born 20th September 1977
citizen of Lauperswil (BE)

accepted on the recommendation of
Prof. Dr. Ulrich Weidmann, examiner
Prof. Dr.-Ing. Ingo A. Hansen, co-examiner

2009
Acknowledgements

Thanks to Prof. Dr. Ulrich Weidmann, I had the opportunity to work in the Institute for Transport Planning and Systems and participated in motivating and challenging projects in the transportation sector. I am very thankful for his supervision and helpful support during my work.

I am enormously thankful for the expertise and constructive reviews provided by Prof. Dr.-Ing. Ingo Hansen and his willingness to referee my thesis.

I wish to thank Oskar Stalder and Dr. Felix Laube, who worked both for the Swiss Federal Railways, offering me the possibility to collaborate on this specific topic and to get a deep insight into the world of railway operations. Together with Dr. Raimond Wüst, Thomas Graffagnino, Samuel Roos, Heinz Egli, Ash Smith, Muriel Perron and many more peoples at SBB, I had the possibility for many fruitful discussions.

Many thanks also to Prof. Dr. Hans-Jakob Lüthi, Gabrio Caimi, Martin Fuchsberger, Dr. Dan Burkolter, Dr. Thomas Herrmann and Dr. Marco Laumanns from the Institute for Operations Research for numerous, insightful and constructive discussions. The effort made by Martin and Gabrio to extend the algorithms used to calculate timetable rescheduling scenarios for the area of Bern has to be pointed out especially.

Special thanks go to Dr. Daniel Hürlimann for the collaboration during his time at IVT and afterwards with his spin-off company OpenTrack Railway Technology. His invaluable inputs and discussions undoubtedly improved this thesis.

I also want to thank all my colleagues at the IVT, in particular Dr. Sonja-Lara Beperling, Stefan Bollinger, Bernd Bopp, Stefan Buchmüller, Niklaus Fries, Markus Rieder, Hannes Schneebeli, Jost Wichser and René Zeller. They made the time at IVT so enjoyable and unforgettable for me. I would also like to thank Adrian Johner, Giorgio Medeossi, and other students who have contributed through their work on this research.

During my work at IVT, I had the pleasure to get in touch with Dr. Thomas Albrecht, Dr. Andrea D’Ariano, Dr. Markus Montigel, Andrew Nash, Prof. Dr. Arnd Stephan, Werner Stohler, Dr. Markus Ullius and many more railway experts with whom I had valuable discussions on this work and to whom I want to express my gratefulness.

I would further like to thank Martin Slater for having proof-read my thesis and for having given me valuable advice on linguistic matters. Thanks go as well to all my good friends with whom I have enjoyed some valuable discussions on this work, apart from many other great moments.

Schlussendlich möchte ich mich bei meiner Familie für die Unterstützung und Liebe in all den Jahren bedanken.
Abstract

Demand for rail travel has grown strongly within the last few years and growth is projected to continue in coming years. While central parts of the European rail network are operated punctually at its capacity limits, building new tracks is very cost intensive and thus only possible in exceptional cases. Adding new services to satisfy increasing demand is often achieved by reducing buffer times between consecutive trains. Buffer times are used to reduce the impact of train delays on overall reliability. While reducing buffer times increases capacity, it also means that small delays to a single train may propagate quickly through the network causing knock-on delays to trains impacted by the delayed train. To handle these challenges, new railway operation principles are required.

The thesis introduces new railway operation principles implying a new framework that combines rail traffic management by rescheduling in real-time with advice tools supporting drivers and guards in order to provide information allowing them to follow precisely the dynamically changeable schedules. Designed as a superimposition of two feedback control loops, the framework aims to increase capacity and stability in heavily used mixed rail traffic networks.

The rescheduling part thereby assures that deviations or events causing conflicts are detected and solved automatically within real-time based on predefined optimisation criterions. These updated schedules are transmitted and visualised to drivers and guards who adjust their behaviour accordingly. Recommendations ensure that specified trajectories are followed accurately. The components to improve or develop as part of the new integrated real-time rescheduling framework are presented in this thesis. Possible designs and their consequences are pointed out. Specifications and requirements for components of sub-processes are also described in order to achieve an efficient and effective framework. With this new framework, rail traffic flow is improved and thus unnecessary time and energy consuming signal stops can be avoided. Consequently, delays and their network-wide propagation are significantly reduced.

The thesis points out that enhancements of conventional railway operations must be done in three areas. Firstly, an uplink transmitting the train’s state, position, delay and event information regularly and more frequently than today must be set up. This is used to identify a deviation or event quickly as well as to predict the future behaviour accurately and is the basis for rescheduling. Secondly, generating feasible schedules within real-time for larger networks and high traffic
density is another challenge. Although promising results have already been obtained, developing algorithms to generate new schedules in real-time will still need research. Thirdly, advice tools visualising the new schedules and the connection from the traffic management system to the train providing this information in real-time is required in order to operate trains accurately accordingly to new recommendations by drivers.

A new network decomposition strategy where buffer times are minimised in capacity bottleneck areas and running time supplements are used in compensation areas in-between is proposed. The combination of an integrated real-time rescheduling framework with a new network decomposition strategy allows scheduled buffer times to be reduced in capacity bottleneck areas. As a result, infrastructure utilisation can be improved without having more delays than with conventional railway operation principles.

A new tolerance band principle is also described. Using this, rescheduling as well as planning are more formalised. A train exceeding the given tolerance directly initiates new rescheduling, resulting in fast identification and a clear process for all actors. For scheduling, tolerance bands are added to minimal headways instead of allocating general, imprecisely planned buffer times. Also, open questions regarding inherent conflicts between rescheduling performance, nervousness, duration, and feasibility are exemplified.

Finally, the additional benefits and limits of the new framework are derived and illustrated by simulations and case scenarios in the areas of Lucerne, Bern and Winterthur. It is shown that the knock-on delay of a single train applying the new framework may be reduced by up to several minutes if an unnecessary signal stop is avoided. For bottleneck areas where recorded delay patterns were applied, the thesis also shows that the overall delay may be significantly reduced with the integrated real-time rescheduling framework. However, precise operations following the reference trajectory is required. Test runs have shown that the robustness and reliability of the driver's display tool are necessary prerequisites but difficult to achieve. Test trials with the new display tool proved that following a reference trajectory is possible. Nevertheless, the aspired level of accuracy only seems achievable when the reference speed is displayed. In particular, references with abrupt speed changes, which may occur to avoid conflicts as part of the rescheduling, will not be manageable without speed information.

To summarise, the thesis demonstrates that with the combination of real-time rescheduling and precise train operations, the benefits of new traffic management systems are utilised and also significantly improved. Altogether, the integrated real-time rescheduling framework developed is a promising approach to improve the performance of the railway system both efficiently as well as effectively.
Zusammenfassung


Im Rahmen dieser Arbeit wurde gezeigt, dass mittels Kombination von Berechnung neuer Produktionspläne in Echtzeit mit präziser Zugsteuerung durch das Lokpersonal ein signifikant höherer Verkehrsfluss erreicht werden kann. Eisenbahnnetzwerke können somit bei gleichbleibender Pünktlichkeit wesentlich besser ausgelastet werden. Das neue System ist deshalb ein vielversprechender Ansatz um die Leistungsfähigkeit von Eisenbahnsystemen effektiv und effizient zu steigern.
Contents

1 Introduction 1
   1.1 Challenges for rail transport 1
   1.2 Basic rail operation relations 3
   1.3 The railways in Switzerland 4
   1.4 Aim and scope of the thesis 5
   1.5 Outline of the thesis 7

2 Basic principles for railway systems 9
   2.1 Overview of the railway system 9
      2.1.1 Fundamental domains and demands for railway system 9
      2.1.2 Functional structure of rail operations and planning 10
      2.1.3 Definitions of general railway terms 12
   2.2 Railway safety and signalling principles 13
      2.2.1 Railway safety principle 13
      2.2.2 Railway signalling 14
      2.2.3 ERTMS: The European Railway Traffic Management System 18
   2.3 Timetable principles 19
      2.3.1 Schedule-based rail operations 19
      2.3.2 Planning principles for passenger and freight trains 19
      2.3.3 Basic principles for schedule design 23
      2.3.4 Railway traffic data analysis 29
      2.3.5 Overview of recent research on planning principles for more robust and stable timetables 33
3 Feedback control principles for railway operations 35

3.1 Introduction to control and railway operations 35

3.2 Feedback control principles 35

3.2.1 Feedforward and feedback control 35

3.2.2 Control and optimisation 37

3.2.3 Definitions of railway operation terms 38

3.3 Railway operation principles 41

3.3.1 General rail operation model 41

3.3.2 Dispatching 42

3.3.3 Categorisation of primary delay reasons 45

3.3.4 Overview of state-of-the-art advanced rail traffic management systems 46

3.4 Challenges and requirements for new railway operation processes to improve the performance 51

3.4.1 General approaches to improve performance 51

3.4.2 Requirements for a new railway operation framework 53

3.5 The concept for an integrated real-time rescheduling system 54

3.5.1 Fundamental structure of superimposed feedback control for railway systems 54

3.5.2 Key concept of the integrated real-time rescheduling 55

3.6 Capacity improvement strategies based on an integrated real-time rescheduling system 57

3.6.1 Geographical network decomposition 57

3.6.2 Integrated real-time rescheduling and the systematic management of bottleneck areas 58

3.6.3 Realisation challenges for integrated real-time rescheduling systems 59

3.7 Conclusions 60

4 The real-time rescheduling loop 61

4.1 The rescheduling loop as part of the integrated real-time rescheduling framework 61

4.2 Relevant processes of the real-time rescheduling loop 62

4.3 Elements of the real-time rescheduling loop 64

4.3.1 Train detection 64

4.3.2 Infrastructure supervision 71
4.3.3 Data processing and preparation ........................................... 72
4.3.4 Generation of new production plans ....................................... 76
4.3.5 Transmission of new production plans ..................................... 90
4.3.6 Rescheduling process duration ................................................. 90
4.4 Influence of temporal aspects on rescheduling performance .......... 95
  4.4.1 Phases for rescheduling ....................................................... 95
  4.4.2 Temporal effects on rescheduling performance .......................... 97
4.5 Conclusions ............................................................................ 98

5 Accurate production control loop .............................................. 101
  5.1 Control loops for precise production in railway systems ................. 101
  5.2 The train driving process as part of the integrated real-time rescheduling framework .................................................. 102
    5.2.1 Closed loop design of train driving .................................... 102
    5.2.2 System architecture of the driver-machine interface ............. 103
    5.2.3 Functional requirements for the driver-machine interface ........ 104
    5.2.4 Reference trajectory as input for the driver-machine interface ... 105
    5.2.5 Possibilities of the driver-machine interface displays .......... 107
    5.2.6 Driver in the loop model .................................................. 113
  5.3 On-time departure ................................................................. 122
    5.3.1 Today's departure process ............................................... 122
    5.3.2 Design for new departure process .................................... 124
  5.4 Time-driven infrastructure operations ...................................... 126
  5.5 Conclusions ............................................................................ 127

6 Integrated real-time rescheduling framework .............................. 129
  6.1 Combining rescheduling and accurate production control loops ......... 129
  6.2 Integrated real-time rescheduling data flow diagram ...................... 129
    6.2.1 Data flow diagram principles ............................................. 129
    6.2.2 Data flow diagram model for railway operations ................. 130
  6.3 Interaction of rescheduling and train operation control loops ........ 133
    6.3.1 Time synchronisation problem ........................................... 133
    6.3.2 Differing loop frequencies problem .................................... 133
6.3.3 Concurrence of rescheduling and train operation loops through the design of the tolerance band ................................................. 134
6.3.4 Unscheduled state problem ......................................................................................................................... 136
6.3.5 Summary of rescheduling and precise operation loop interconnections .................................................. 138
6.4 Optimisation potential with integrated real-time rescheduling ................................................................. 139
6.4.1 Overview of optimisation opportunities ......................................................................................................... 139
6.4.2 Traffic flow optimisation ............................................................................................................................... 139
6.4.3 Capacity improvements with integrated real-time rescheduling and network decomposition ................. 146
6.4.4 Further optimisation opportunities through integrated real-time rescheduling ..................................... 148
6.5 Human players in the integrated real-time rescheduling framework .......................................................... 149
6.5.1 Challenges and new duties for players in the integrated real-time rescheduling framework .................. 149
6.5.2 Players affected by the integrated real-time rescheduling framework ....................................................... 150
6.6 Standardisation challenge for integrated real-time rescheduling systems in Europe ............................... 151
6.7 Conclusions ....................................................................................................................................................... 153

7 Case studies ....................................................................................................................................................... 155
7.1 Introduction ...................................................................................................................................................... 155
7.2 Test areas ........................................................................................................................................................ 157
7.2.1 Lucerne area ................................................................................................................................................ 157
7.2.2 Bern area .................................................................................................................................................... 160
7.2.3 Winterthur area .......................................................................................................................................... 161
7.3 Simulation studies .......................................................................................................................................... 163
7.3.1 Microscopic railway simulation .................................................................................................................. 163
7.3.2 Simulation objectives .................................................................................................................................... 164
7.3.3 Impact of rescheduling point of time ........................................................................................................ 165
7.3.4 Impact of production accuracy .................................................................................................................. 172
7.4 SBB test runs .................................................................................................................................................. 175
7.4.1 SBB’s PULS 90 and FARE projects ........................................................................................................ 175
7.4.2 Overview of SBB’s test runs ...................................................................................................................... 176
7.4.3 The Rotsee test case ................................................................................................................................. 177
7.5 Conclusions ................................................................. 186

8 Conclusions ................................................................. 187
8.1 Summary ........................................................................ 187
8.2 Migration possibilities for an integrated real-time rescheduling system .... 189
8.3 Future research and developments ...................................... 191

A Abbreviations .................................................................. 195

B Evaluation of running times ............................................... 197
B.1 Analysis of train driver behaviour ..................................... 197
B.2 Results of train driving behaviour analysis ....................... 198

C Influence of signals on train delays and capacity usage ............ 207
C.1 Introduction .................................................................... 207
C.2 Notation .......................................................................... 208
C.3 Analysis for a signalling system which allows direct reacceleration ... 209
  C.3.1 Derivation of running time and passing time equations for a signalling sys-
      tem with direct reacceleration ........................................ 209
  C.3.2 Example of a signalling system with direct reacceleration .......... 215
C.4 Analysis for a signalling system with fixed approach speed .......... 217
  C.4.1 Derivation of running time and passing time equations for a signalling sys-
      tem with fixed approach speed and no direct reacceleration ....... 217
  C.4.2 Example of a signalling system with a fixed approach speed and without any direct reacceleration ..................................................... 228
C.5 Other signalling systems .................................................. 230

D Energy savings through integrated real-time rescheduling ........ 231
D.1 Introduction .................................................................... 231
D.2 Notation .......................................................................... 231
D.3 Analysis for energy savings by reducing a train’s influences ....... 232
D.4 Potential energy saving ..................................................... 232
  D.4.1 Potential energy savings for an IC2000 train ..................... 233
  D.4.2 Potential energy savings for a FLIRT train ...................... 234
### Rescheduling simulation scenarios

#### E.1 Overview of Lucerne area simulation scenarios for rescheduling point of time impacts
- **E.1.1 Planned schedule**
  - Page 237
- **E.1.2 Rescheduling measures and characteristics**
  - Page 239

#### E.2 Overview of Bern area simulation scenarios for rescheduling point of time impacts
- **E.2.1 Planned schedule**
  - Page 242
- **E.2.2 Rescheduling measures and characteristics**
  - Page 243

#### E.3 Overview of Winterthur area simulation scenarios for rescheduling point of time impacts
- **E.3.1 Planned schedule**
  - Page 245
- **E.3.2 Rescheduling measures and characteristics**
  - Page 246
# List of Figures

1.1 UIC 406 capacity balance (source: adapted from [UIC04]) ........................................ 4  
1.2 Thesis structure ........................................................................................................... 8  
2.1 Functional structure for railway operations and planning ........................................... 10  
2.2 Hierarchical system structure for railway operations ................................................... 11  
2.3 Elements of the blocking time .................................................................................... 15  
2.4 Blocking time stairways ............................................................................................. 16  
2.5 Principle of the integrated fixed-interval timetable ..................................................... 20  
2.6 Example of a section of the integrated fixed-interval timetable for the area around the main station of Lucerne (source SMA und Partner AG Zurich, Timetable 2005) 21  
2.7 Occupancy in Zurich main station (timetable 2006) ................................................... 22  
2.8 Future main line network and system nodes of Switzerland (source: Bundesblatt - Swiss Federal Gazette) ................................................................. 22  
2.9 Planning phases ......................................................................................................... 24  
2.10 Improvement objectives in timetable planning ......................................................... 24  
2.11 Components required to design a schedule for an integrated fixed-interval timetable with symmetrical time at full hour ............................................................ 25  
2.12 Dwell duration as function of the total number of passengers involved in the boarding and alighting process for Zurich’s S-Bahn DPZ ........................................ 28  
2.13 Comparison of measured and fitted log-logistic density functions for 230 recorded weekday morning rush hour train trips of S-Bahn train number 18526 in 2004 from Bubikon to Wetzikon ........................................ 31  
2.14 Variation of dwell times for late commuter trains on through stations (sorted by their median dwell time) based on data recorded by SBB between March and November 2004 ......................................................... 32  
3.1 Schematic layout of an open-loop control system ....................................................... 36
3.2 Design of a generic feedback control system ................................................. 36

3.3 Block diagram of a generic cascade control system ........................................ 37

3.4 Possible control and optimisation architectures. Left: centralised optimisation in hierarchical system; middle: decentralised optimisation in hierarchical system; right: decentralised optimisation and control in an integrated architecture ........................................ 38

3.5 Schema of today’s rail operation .................................................................... 42

3.6 Dispatching process ....................................................................................... 43

3.7 Online scheduling procedure (source [Jac04]) ................................................. 49

3.8 The ERTMS Level 3 operation control loop (source [Maz07]) ......................... 49

3.9 Visualisation of AdmiRail, the intelligent platform concept for a train traffic control and automation system of the future developed by systransis for the ETCS Level 2 Lötschberg-Basetunnel (source systransis AG) ........................................ 51

3.10 Train headway reduction based on reduced and homogenised block length ... 52

3.11 General model of the integrated real-time rescheduling framework ................. 55

3.12 Train control with and without integrated real-time rescheduling ................. 56

3.13 Network classification schemes ................................................................. 58

4.1 General model of the outer rescheduling loop ............................................... 61

4.2 Relevant processes within the rescheduling loop ............................................ 63

4.3 Principle of the tolerance bandwidth .............................................................. 65

4.4 Infrastructure-based train detection ............................................................... 66

4.5 Resulting temporal deviation due to inaccuracy of positioning system .......... 68

4.6 Periodic train location detection .................................................................. 68

4.7 Data flow for periodic transmission (left) and infrastructure based train detection (right) ................................................................. 70

4.8 Two-level prediction process ...................................................................... 75

4.9 Procedure of predicting future behaviour ...................................................... 76

4.10 Visualisation of exemplary penalty-functions for delayed trains .......... ................................. 78

4.11 Multi-level rescheduling process ............................................................... 86

4.12 Comparison of train diagram (left) and pulsing (Source [Roo06]) ................. 86

4.13 Visualisation of different rescheduling policies ........................................... 89

4.14 Comparison of process and transmission durations for infrastructure based train detection .......................................................................................................................... 91
4.15 Comparison of process and transmission durations for periodic status transmission from train ........................................................................................................ 92
4.16 Impact of time on the rescheduling performance .................................................................................................................................................................. 98

5.1 General model of the inner precise operation control loop ........................................................................................................................................... 102
5.2 Design of the inner train driving process loop ................................................................................................................................................................. 102
5.3 System architecture of the DMI with the transmission of a complete trajectory (left) and only reference points and on-board optimisation (right) .......................................................................................................................... 104
5.4 Possible trajectories between two reference points ...................................................................................................................................................... 105
5.5 Principle of defining new trajectories: shifting with temporal offset (left) or distortion of existing reference trajectory (right) ........................................................................................................................................... 106
5.6 Steps in the temporal deviation (right) caused by a variation of the stopping position ........................................................................................................ 108
5.7 Temporal deviations after departing from stations due to varying stopping positions .................................................................................................. 108
5.8 Adjusted reference trajectories depending on the departing position in the station resulting in offset-free trajectories .................................................................................................................................................. 109
5.9 Possible display with speed information including maximum track speed (yellow), reference track speed profile (red), current (and past) speed profile (blue), actual train position with train length and station information (icons) ........................................................................................................................................ 111
5.10 Symbols displaying recommendations how to adjust the driving behaviour ........................................................................................................ 112
5.11 Example of FARE DMI displaying a slight delay ......................................................................................................................................................... 112
5.12 Bathtub (delay-retiming shortly before a conflict area) speed reference trajectory ......................................................................................................... 115
5.13 Block diagram of a train driver in the loop model ...................................................................................................................................................... 116
5.14 Flow diagram for the train driver decision process ................................................................................................................................................ 118
5.15 Transient effects (of the temporal deviation) after an occurring deviation ........................................................................................................ 121
5.16 Setup of today’s departure process .............................................................................................................................................................. 123
5.17 Design of the departure process with improved precision ........................................................................................................................................ 125

6.1 Detailed process and dataflow model of the integrated real-time rescheduling system ........................................................................................................ 131
6.2 Differing loop frequencies for inner and outer control loop ........................................................................................................................................... 133
6.3 Comparison of continuous tolerance band (left) to tolerance slot at reference points (right) causing possible headway conflicts (red) ........................................................................................................................................... 134
6.4 Visualisation of continuously narrowing tolerance bandwidth (left) and gradually tolerance bandwidth reduction at stations (right) ........................................................................................................................................... 135
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.5 Optimistic and defensive prediction after exceeding of the tolerance bandwidth</td>
<td>136</td>
</tr>
<tr>
<td>6.6 Illustration of the unscheduled time period</td>
<td>137</td>
</tr>
<tr>
<td>6.7 Overview of interconnections in the rescheduling and precise operation control loop</td>
<td>138</td>
</tr>
<tr>
<td>6.8 Schematic topology to analyse the delay effects for trains approaching a possible conflict situation</td>
<td>140</td>
</tr>
<tr>
<td>6.9 Time-distance diagram for five different passing times when a train approaches a conflict situation</td>
<td>141</td>
</tr>
<tr>
<td>6.10 Sample delay effect for trains approaching a possible conflict situation</td>
<td>142</td>
</tr>
<tr>
<td>6.11 Influence of driving strategies and rescheduling point of time on energy consumption based on an Intercity train from Rotkreuz to Lucerne</td>
<td>145</td>
</tr>
<tr>
<td>6.12 Influence of the rescheduling point of time on energy consumption based on an Intercity train from Rotkreuz to Lucerne without coasting</td>
<td>145</td>
</tr>
<tr>
<td>6.13 Sample delay distribution for two consecutive trains in today's operation</td>
<td>146</td>
</tr>
<tr>
<td>6.14 Comparison of capacity usage for conventional (left) and integrated real-time rescheduling operations (right)</td>
<td>148</td>
</tr>
<tr>
<td>6.15 Comparison of planning approaches and operational (delay propagation) effects in the event of a train arriving late in a capacity critical condensation area for conventional and integrated real-time rescheduling operations</td>
<td>148</td>
</tr>
<tr>
<td>6.16 Integrated real-time rescheduling Level 0: no data exchange between train and traffic management centre</td>
<td>152</td>
</tr>
<tr>
<td>6.17 Integrated real-time rescheduling Level 1: down-link data transfer from traffic management to train</td>
<td>152</td>
</tr>
<tr>
<td>6.18 Integrated real-time rescheduling Level 2: two-way data communication between train and traffic management</td>
<td>153</td>
</tr>
<tr>
<td>7.1 Topology of Swiss rail network (based on SBB map dated July 10, 2008)</td>
<td>157</td>
</tr>
<tr>
<td>7.2 Aggregated topology of the rail network around Lucerne</td>
<td>158</td>
</tr>
<tr>
<td>7.3 Time distance diagram of the planned schedule for the Lucerne condensation area</td>
<td>159</td>
</tr>
<tr>
<td>7.4 Aggregated topology of the rail network around Bern</td>
<td>160</td>
</tr>
<tr>
<td>7.5 Aggregated topology of the rail network around Winterthur</td>
<td>162</td>
</tr>
<tr>
<td>7.6 Time distance diagram of the planned schedule for the Winterthur condensation area</td>
<td>163</td>
</tr>
<tr>
<td>7.7 Visualisation of the planned schedule with the pulsing tool</td>
<td>168</td>
</tr>
<tr>
<td>7.8 Total relative delay depending on the rescheduling point of time for all rescheduling scenarios for the Lucerne area</td>
<td>170</td>
</tr>
</tbody>
</table>
7.9 Total relative delay depending on the rescheduling point of time for all rescheduling scenarios for the Bern area ............................... 170
7.10 Total relative delay depending on the rescheduling point of time for all rescheduling scenarios for the Winterthur area .................. 170
7.11 Average consecutive delay in Lucerne area for each train for different temporal delay distribution width variations .................. 173
7.12 Average consecutive delay in Winterthur for each train for different temporal delay distribution width variations .................. 173
7.13 Arrival times at stations in dependence of the passing time at a given reference point for the area around Lucerne (only the passing time of specific trains was varied to illustrate the non-linear delay effects) .................... 175
7.14 Detailed topology of the Rotsee single-line section .................................................. 177
7.15 Running time of Interregio trains towards the single line section at Rotsee in dependence of the relative passing time with opposite train (green: major likelihood of no influence, blue: potential likelihood of an influence, magenta: possible likelihood of an influence; red: major likelihood of an influence) .................. 179
7.16 Critical section of running time and relative passing times for Interregion trains at the Rotsee to identify possible influences .......... 180
7.17 Development of the temporal deviation for trains running with FARE from Rotkreuz to Lucerne with minimum running time as scheduled reference trajectory .................................................. 182
7.18 Development of the temporal deviation for trains running with FARE from Rotkreuz to Lucerne under special conditions or occurring events ................................................................. 183
7.19 Development of the temporal deviation for trains running with FARE from Rotkreuz to Lucerne .................................................. 184
7.20 Running time of Interregio trains towards the single line section at Rotsee in dependence of the relative passing time with opposite train (coloured stars: all trains 2321 without FARE during the test period; circle: FARE test runs with normal conditions; triangle: FARE test runs with minimum running time as reference trajectory; square: FARE test runs with special conditions) .............. 185

B.1 Running times for IR 1935 Rotkreuz - Gisikon .................................................. 199
B.2 Running times for IR 2820 Goldach - Moerschwil .................................................. 199
B.3 Running times for IC 615 Cornaux - Le Landeron .................................................. 200
B.4 Running times for IC 759 Mels - Sargans .................................................. 200
B.5 Running times for IC 510 Lengnau - Biel Matt .................................................. 200
B.6 Running times for IC 345 Dagmersellen - Wauwil .................................................. 201
C.6 Detailed visualisation of the four different cases influenced by the signal for the slow speed part at the main signal for a signalling system without direct reacceleration.

C.7 Running time between two reference points depending on the point of time when the first reference point is passed for a signal system without continuous information and no reacceleration until the main signal is passed.

C.8 Point of time second reference point is passed in function of the point of time when the first reference point is passed for a signal system without any loop information and no reacceleration before the main signal is passed.

C.9 Time distance visualisation for the four specific cases for a signalling system with a fixed approach speed and no direct reacceleration.

D.1 IC2000 Intercity train with a Re460 locomotive. (Source SBB AG).

D.2 FLIRT commuter train. (Source Stadler Rail AG).

D.3 Freight train with a Re482 locomotive. (Source SBB AG).

E.1 Topology of the condensation area around Lucerne’s rail terminal.

E.2 Identification of the latest rescheduling point of time to pass the condensation boarder at an earlier or later slot.

E.3 Overview of rescheduling measures for Lucerne scenario 1.

E.4 Overview of rescheduling measures for Lucerne scenario 2.

E.5 Overview of rescheduling measures for Lucerne scenario 3.

E.6 Overview of rescheduling measures for Lucerne scenario 4.

E.7 Overview of rescheduling measures for Lucerne scenario 5.

E.8 Overview of rescheduling measures for Lucerne scenario 6.

E.9 Overview of rescheduling measures for Lucerne scenario 7.

E.10 Overview of rescheduling measures for Lucerne scenario 8.

E.11 Original track occupation for Bern rescheduling scenario (grey: shunting movements, red lines: station platform).

E.12 Overview of rescheduling measures for Bern scenario 1.

E.13 Overview of rescheduling measures for Bern scenario 2.

E.14 Overview of rescheduling measures for Bern scenario 3.

E.15 Overview of rescheduling measures for Bern scenario 4.

E.16 Overview of rescheduling measures for Bern scenario 5.
<table>
<thead>
<tr>
<th>E.17 Overview of rescheduling measures for Bern scenario 6</th>
<th>244</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.18 Overview of rescheduling measures for Bern scenario 7</td>
<td>245</td>
</tr>
<tr>
<td>E.19 Overview of rescheduling measures for Winterthur scenario 1</td>
<td>246</td>
</tr>
<tr>
<td>E.20 Overview of rescheduling measures for Winterthur scenario 2</td>
<td>246</td>
</tr>
<tr>
<td>E.21 Overview of rescheduling measures for Winterthur scenario 3</td>
<td>246</td>
</tr>
<tr>
<td>E.22 Overview of rescheduling measures for Winterthur scenario 4</td>
<td>246</td>
</tr>
<tr>
<td>E.23 Overview of rescheduling measures for Winterthur scenario 5</td>
<td>247</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Optimisation criteria and time requirements for dispatching depending on delay reason</td>
<td>46</td>
</tr>
<tr>
<td>4.1</td>
<td>Overview of characteristics for train detection systems</td>
<td>71</td>
</tr>
<tr>
<td>4.2</td>
<td>Most of all known event types causing delayed deviations</td>
<td>74</td>
</tr>
<tr>
<td>4.3</td>
<td>Rescheduling phases and possible rescheduling measures</td>
<td>96</td>
</tr>
<tr>
<td>5.1</td>
<td>Examples of occurring deviations for running trains</td>
<td>114</td>
</tr>
<tr>
<td>5.2</td>
<td>Overview of departing delays (time taken after all conditions are satisfied until the train departs)</td>
<td>124</td>
</tr>
<tr>
<td>7.1</td>
<td>Overview of case studies and areas</td>
<td>156</td>
</tr>
<tr>
<td>7.2</td>
<td>Basic timetable of Lucerne used for simulation studies</td>
<td>159</td>
</tr>
<tr>
<td>7.3</td>
<td>Basic timetable of Bern used for simulation studies</td>
<td>161</td>
</tr>
<tr>
<td>7.4</td>
<td>Basic timetable of Winterthur used for simulation studies</td>
<td>162</td>
</tr>
<tr>
<td>7.5</td>
<td>Rescheduling scenarios to evaluate the impact of rescheduling point-of-time for the Lucerne area</td>
<td>166</td>
</tr>
<tr>
<td>7.6</td>
<td>Rescheduling scenarios to evaluate the impact of rescheduling point-of-time for the Bern area</td>
<td>166</td>
</tr>
<tr>
<td>7.7</td>
<td>Rescheduling scenarios to evaluate the impact of rescheduling point-of-time for the Winterthur area</td>
<td>166</td>
</tr>
<tr>
<td>7.8</td>
<td>Reduction of knock-on delays with rescheduling (rerouting, reordering and speeding up trains) for a single initially delayed train for the simulation studies in Lucerne (LZ), Bern (BN) and Winterthur (W) areas</td>
<td>169</td>
</tr>
<tr>
<td>B.1</td>
<td>Trains and track sections used for analysis of train driver behaviour</td>
<td>198</td>
</tr>
<tr>
<td>D.1</td>
<td>Overview of possible energy savings for an IC2000 train</td>
<td>233</td>
</tr>
</tbody>
</table>
D.2 Overview of possible energy savings for a FLIRT train 234
D.3 Overview of possible energy savings for a typical freight train 235
E.1 Planned routes for the original timetable 238
E.2 Planned routes for the original timetable 245
Chapter 1

Introduction

1.1 Challenges for rail transport

Transport plays a key role in economic, environmental and social objectives and is an important location factor for the competitiveness of companies and regions. The infrastructure that is required and installed thus defines the land use for decades or longer. Fast transport and well connected networks for various traffic modes of both passenger and goods enlarge the operating range and provide more choice. Transportation can therefore enhance our quality of life. However, transport has negative effects on environment depending on the means of transportation.

Over the years, rail passenger and freight traffic has become less important whereas the amount of road traffic has rapidly increased [EU006, LIT07]. Combined with rapid worldwide traffic growth, this fact leads to a deteriorating environment and diminishing fossil energy resources. Thereby, not only ecological and resource constraints are consequences of increased road traffic. Limited capacity, especially in dense urban areas causes traffic congestion, frequent and serious accidents or reduces our quality of life due to noise and pollution. There is thus a conflict between the behaviour of the users of transportation and our requirements for sustainable development. Public transportation and, in particular, rail traffic play a key role to ensure sustainable mobility. To strengthen their position, railways must become more competitive, more efficient and deliver safe, fast, reliable, cheap and comfortable services that correspond to customers’ expectations.

The rail sector is faced with two main challenges. On the one hand, costs must be reduced and operation needs to be more efficient. On the other hand, the growing demand for transport must be handled attractively and competitively. This not only includes transport costs but also speed, capacity, availability, frequency of services, punctuality, reliability, and comfort.
Within recent decades, pressure on costs throughout Europe has resulted in the closing of train lines, especially in rural and sparsely populated areas. In contrast, rail services in agglomerations, between cities or on high speed lines were extended. The service increase was thereby followed by an immense increase in demand. For this purpose, new lines were built or the existing infrastructure was enhanced, and more services offered. The infrastructure thus often operates at its limit and there is a major need for more efficiency and new capacities on rail networks with minimal investments, for example, in Switzerland and other countries [Fri98, Sta03, Sta04b].

As the number of trains in operation increases, it is also more difficult to cope with delays. This means that even small delays can have a major impact on the entire network. Furthermore, offering new or additional services on the existing infrastructure with existing operation methods may be limited or not even possible because available capacity is already fully utilised. Upgrading the track network and other investments are expensive, complicated, can take ten or more years and are therefore often impossible. Regional development and rail traffic efficiency enhancements were thus unnecessarily delayed.

The historical development in railways with limited cross-boarder traffic led to national rail networks with their own specialities. Because of the lack of technical and operational standards, international travel nowadays has several obstacles. The lack of standardisation of signalling systems led to an initiative in the European Union to develop a new European rail traffic management system (ERTMS) [EU001, The01, The02]. ERTMS can not only improve cross-border rail traffic, but can also optimise safety and reliability. In addition, track capacity may be enlarged. In parallel to cross-border developments, rapid progress in information technology and computer science are generating a multiplicity of new applications offering the opportunity to improve capacity and quality of rail traffic.

Beside the need to improve capacity and punctuality as well as solve interoperability problems, rail traffic and operation nowadays is subject to many other challenges. Market liberalisation and the division into infrastructure management and train operators has resulted in increased competition and potential improvements but also interaction problems due to the boundaries between new companies. In the range of freight trains, a trend to short-term track slot request and reservation is observable. Therefore, to compete with the flexibility of road haulage, an important element is the need to offer possible slots within shortest time. Furthermore, information management for freight customers and passengers must be improved. In particular, the exactness and timeliness of information as well as possibilities to provide specific user-oriented information must be improved. Finally, new strategies aimed at reducing energy consumption are needed for more sustainable rail transport.

Rail industry and rail operation research is faced with all these challenges. Thus, new methods or processes as well as modern technologies must be adapted, developed, and applied in an efficient and cost-effective way. This will result in a more efficient and successful rail system where infrastructure usage together with punctuality can be improved and finally help shifting modal split towards rail traffic.
1.2 Basic rail operation relations

Rail operations are based on historic principles and technologies as well as on physical circumstances (e.g. long braking distances). The operation rules applied influence the effectiveness and usage of the given infrastructure. The capacity of a rail network, a main performance characteristic which must be improved, depends on many parameters. Moreover, specifying capacity as such does not exist, it is heavily dependent on the way a given infrastructure is utilised. Nevertheless, there are numerous recommendations for possible definitions with different meanings [Kru99, Lan08].

However, the capacity of a railway line or section always depends on the minimum time between two trains. Infrastructure and train characteristics are the basis for determining the minimum time between two consecutive trains. This so-called headway is affected by the following relevant factors:

- track topology (single-line section or multiple tracks on open track; number of station tracks);
- signalling system;
- block distance;
- level of automation (sectional route release) and magnitude of process delays (release time); and
- train characteristics (speed, length, deceleration rates).

Beside infrastructure and train characteristics, which define headways, other items seriously affect capacity. Grouping trains with identical characteristics (e.g. speed), or trains running in the same direction for single-line sections, for example, has a positive effect on the number of possible trains which can be operated in a given infrastructure. Consequently, the timetable also determines the possible capacity. For a specific infrastructure, the International Union of Railways (UIC) identified the following operational parameters influencing the capacity [UIC04]:

- number of trains (number of trains per defined time interval);
- stability (buffer times between two consecutive trains and running time supplements);
- heterogeneity (difference in travel time, maximum speed, length, braking and accelerating behaviour of different types of trains); and
- average speed (braking distance increases in the square of speed).

The different parameters that affect capacity are interconnected. For example, a higher heterogeneity of trains results in a change in average speed and in varying headways causing a lower
number of trains using the infrastructure compared to the case with homogenous (identical) trains. The relationship between these influence factors is illustrated in Figure 1.1. The qualitative model has a parameter on each axis starting at the origin. The value of a parameter is represented by the points connected to each axis. The capacity of an infrastructure is illustrated by the overall length of the initial line connecting the four points. Any improvement in capacity would imply an increase in total line length.

![Figure 1.1: UIC 406 capacity balance (source: adapted from [UIC04])](image)

The capacity balance illustrates that improving the number of operated trains results in less stability having identical conditions on a given infrastructure. Thereby, the reduction in stability is created by reducing the running time supplements or buffer times between trains. Time supplements as well as buffer times between trains are used to reduce and recover delays. Consequently, punctuality, a crucial quality criterion for rail customers, is directly coupled to stability. The challenge consists in improving the stability and the number of trains at the same time. New methods, processes and technologies are therefore needed. Consequently, a multiplicity of additional parameters such as the costs of development and total life cycle, safety, availability, reliability or quality must be considered and improved.

### 1.3 The railways in Switzerland

Public and rail transportation, in particular, has played an important role in Switzerland for a long period of time. Market share for public freight and passenger transport is higher compared to most other countries in the world. Switzerland is characterised by a dense population in the Mittelland between the Alps and the Jura Mountains. Actual demand already implies a high utilisation of the infrastructure and almost no new infrastructure can be developed due to limited spatial extension potentials. Infrastructure extensions to improve the attractiveness and capacity of the rail network must therefore be carefully considered.
In 1983/1984 Swiss cantons rejected plans for a new high-speed route (the so-called Neue Haupttransversalen NHT) on the main west-east axis of the country. Following the rejection 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 fixed-interval timetable [Hür06, Krä04]. The Bahn 2000 infrastructure plan was based on a highly systematic timetable and tailor-made new infrastructures. In the new schedule, connections in most major stations are coordinated to minimise waiting and overall travel times.

The Bahn 2000 plan was gradually implemented after 1997 and the first stage (1. Etappe) was put into service in December 2004. Many routes are now operated on a 30-minute frequency pattern throughout the day. The service has been extremely successful in attracting more passengers to rail service and demand is expected to increase as additional elements of Bahn 2000 are completed. The railway thus became a victim of its own success: some parts of the Swiss rail network are used so densely that even minor delays or events can propagate quickly through the entire network and cause large knock-on delays. Having 93.8 trains per day and route on average, the SBB infrastructure operates the most heavily used mixed traffic rail networks in the world on its approximately 3,000 kilometre network. Nevertheless, 96.2 percent of all passenger trains are punctual, defined by SBB as the percentage of all trains that reach a main reference station with a maximum delay of 4:59 minutes [SBB06, UIC08].

Rail infrastructure projects are almost always well supported by state and local authorities in Switzerland and the aim of policy is to get traffic off the roads onto rail. Nevertheless, Bahn 2000 and in particular the two new base tunnels through the Alps tied up a lot of money. Consequently, other infrastructure enhancements, coordinated and known as ZEB (called Zukünftige Entwicklung der Bahnprojekte), have been reduced to a minimum. The network is thus getting increasingly saturated and will be operating at the limits of its stability. To handle increasing demand and to solve the upcoming and inherent quality problems, several programs and projects have been initiated by the SBB.

PULS 90 was one of these projects initiated by SBB to improve the performance of the rail infrastructure by adjusting given planning and operation principles. Among other things, the present thesis describes the framework of the new principle and was part of the PULS 90 project to improve the efficiency and effectiveness of rail operations. The dense rail traffic on SBB’s network is thus an excellent prerequisite to evaluate the framework and possible improvements.

1.4 Aim and scope of the thesis

A major challenge for railway infrastructure and operation companies is to handle the increased amount of traffic more efficiently. New construction projects are limited and consequently new approaches are required. New operation principles including optimisation and control methods are thus one way of achieving the demanding objective target.
The railway system is a large-scale multi-layer structure with high real-time demands and large geographical distribution. Applying optimisation and control systems needs a detailed formal analytical treatment of the structure, processes and involved methods. This is often missing when controllers and optimisation for large-scale systems are developed [Lun92]. The aim of this thesis is to identify and describe a new integrated real-time rescheduling framework to increase both stability (punctuality) and capacity (number of operated trains) concurrently. The key element of this new integrated real-time rescheduling approach is the combination of a rescheduling system generating conflict free schedules in real-time after a delay or incident with the provision of current information to all the actors involved (train drivers, guards, infrastructure operators) in order to operate trains precisely following the dynamically changing schedule.

Consequently, this thesis consists of the following major research objectives:

1. A new integrated real-time rescheduling framework is introduced, modelled, and designed in order to handle rail traffic more efficiently. The feedback control loop concept is the basis for the newly developed and described framework. The particular features of densely used mixed rail traffic are also considered.

2. Requirements to the framework as well as to single processes and interactions of processes are specified based on a systemwide point of view in order to assure maximum improvements.

3. Added value, potential benefits concerning capacity, stability and energy consumption, as well as the limitations of the new integrated real-time rescheduling framework are identified.

The thesis will offer responses to the following hypotheses:

1. Railway operation processes can be formed and operated as superimposed feedback control loops.

2. All process steps from the detection of a deviation up to the transmission of a new schedule are time-critical elements and have a significant effect on the rescheduling performance.

3. The system architecture of the driver-machine interface as well as the information displayed to the driver affect the achievable accuracy of the driving process.

4. The new integrated real-time rescheduling framework increases traffic flow and yields to greater stability (punctuality) and reduced energy consumption.

5. Combined with a new way of network separation, the new framework allows to minimise buffer time allocations in bottleneck areas and thus maximises the capacity usage of the given infrastructure.
The development of the research results was carried out in close collaboration with the infrastructure division of the SBB and the Institute for Operations Research (IFOR) at ETH Zurich and was part of SBB’s PULS 90 project. The special features of Swiss rail operation including the signalling system, large train heterogeneity, dense traffic and specific network topology thus influenced the research and development. Nevertheless, the new approach described with an integrated real-time rescheduling system is a generic framework and can be applied by all infrastructure owners and railway operators.

1.5 Outline of the thesis

The thesis is structured in four parts, illustrated in Figure 1.2. First, an introduction to the problem is described in this chapter 1. The following chapter 2 gives a general overview of fundamental railway principles including signalling and timetable planning.

The second part focuses on rail operations. The new framework is introduced based on feedback control principles and an analysis of conventional operation principles. The structure and enhancements of the new integrated real-time rescheduling system are thus specified. The approach couples state-of-the-art technology with new rail operational strategies and includes feedback control concepts adapted to rail particularities. The analysis results in a new framework, designed as a superimposition of two feedback control loops. This general integrated real-time rescheduling framework is introduced in chapter 3. The outer rescheduling loop is analysed and illustrated in chapter 4 whereas elements of the inner accurate production control loop are described in chapter 5. All sub-processes are thus identified and explained. Requirements for efficient railway operations are also specified. Chapter 6 provides a detailed explanation of the framework including effects caused by the superimposition of the two control loops. Also the potential benefits achieved by the new integrated framework are derived and pointed out in this chapter.

The third part describes the results of simulation studies and the SBB’s field tests with the new iRTR system for the Lucerne area. The studies are used to demonstrate the applicability, effectiveness and limits of the new approach. The setup and results of these case studies are introduced and discussed in chapter 7.

Finally, conclusions are drawn in the fourth part, chapter 8. Recommendations for further developments, possible migration steps and research areas are also emphasised.
Chapter 1. Introduction

Figure 1.2: Thesis structure

Part 1
1. Introduction
2. Fundamental principles for railway systems

Part 2
3. Integrated real-time rescheduling framework
4. Elements of the real-time rescheduling loop
5. Elements of the accurate production control loop
6. Interaction effects potential benefits

Part 3
7. Case studies

Part 4
8. Conclusion
Chapter 2

Basic principles for railway systems

2.1 Overview of the railway system

2.1.1 Fundamental domains and demands for railway system

The fundamental domains, structures, and principles of railway signalling and timetable planning are introduced below. This basis is used to identify interconnections, weak points and possible improvements for new railway operation principles.

In general, a rail system consists of four domains:

- the infrastructure on which the trains run (including its power supply, telecommunication, safety and traffic control systems);
- the rolling stock;
- the schedule, which defines the agreed offer; and
- the operation rules.

Operation rules thus impact the design and construction of both infrastructure and rolling stock and they decisively affect the rail system’s performance and efficiency. Depending on technological or regulatory developments, the operation rules are adjusted continuously to enhance the effectiveness and competitiveness of rail traffic.

Nowadays, infrastructure, rolling stock, and operation rules are designed in a way that passengers and goods can be transported to their destination as safely, quickly, punctually, cheaply and comfortably as possible. It must therefore not be forgotten that several principles are still based on ideas and solutions from the early beginnings of railway operation. Also infrastructure elements, applied technologies and rolling stock have been in use for many decades. Upgrading adjustments are thus not always possible.
Long life cycles for infrastructure elements and trains reduce or slow down the possibilities to change or adjust the production process. New elements are therefore usually introduced gradually in a limited area and network-wide implementations can take up to several decades. In addition, there is a strong correlation between operation processes and available safety and traffic control systems. Existing operation rules can consequently limit possible benefits by new systems.

### 2.1.2 Functional structure of rail operations and planning

The structure of the railway system and its processes are designed as a pyramid with several layers (Figure 2.1). The functional structure consists of several layers with an integrated control and communication system. This general structure is transferred from other automation industries and adjusted to the specific requirements and particularities of railway systems [Fen03].

![Figure 2.1: Functional structure for railway operations and planning](image)

The bottom operational layer consists of the immediate production and its safety supervision. The transport product - the movement of passengers and goods by trains - is thus generated by commanding and controlling infrastructure and trains. The large geographical distribution of the network and the enormous amount of actions taking place at the same time are the specific characteristics of the operation of trains and infrastructure. A signalling system prevents or minimises accidents and hazardous situations due to human errors or technical failures. A signalman in a local interlocking system or movement inspectors with remote control in centralised traffic control centres control the trains by setting routes. Thereby, also the sequence of trains on tracks can be set based on the specifications on the timetable and the actual situation.

The tactical layer within a railway systems is used to supervise the network and to solve deviations, failures, interruptions or unplanned events. The superior overview of the location of all trains and infrastructure status allows conflicts to be identified and solved. This so-called dispatching process is executed by a few traffic management centres. Thereby, connections to be held and decisions with large impacts on trains are defined. Nowadays, dispatching and operation is functionally strictly separated for railway systems. Communication to give orders and measures is mainly carried out verbally by phone; system states and status reports can
only be monitored but not directly accessed between the operational and the tactical layers. A closer connection or merging of both layers into a single unit is foreseeable and will improve rail operation’s efficiency.

Timetables are the basis of operations. Planning the schedule is a strategic task. However, the planning process can also be divided into several sub-levels, starting with long-term tasks and ending with the daily schedule. Experience based on past operational data analysis is used as an important input to enhance the schedule. Furthermore, the top-layer includes management tasks (for example investment strategies, archiving or accounting). These activities are not directly connected with the daily operational work and the time horizon can cover up to several decades.

The resulting hierarchical technical structure is visualised in Figure 2.2. Communication and information exchange between the layers depends on various requirements, for example safety, reliability, availability, or temporal demands and the actors involved. Use is made of telecommunication systems for oral and data exchange, computer networks with data-bus systems, documents as handouts, bulletins or leaflets, visual displays like signals, and so on.

![Hierarchical system structure for railway operations](image)

Figure 2.2: Hierarchical system structure for railway operations
The system structure is influenced by operating rules and developments in many fields but mainly in electrical engineering, telecommunication, and computer science. These principles result in the fact that rail operation is an extremely safe transportation mode even for high speed and hazardous goods. And additionally, up to a certain level of operated trains, the pyramidal structure also allows fast and flexible intervention whenever an error or disturbance occurs. Safe, efficient and cost-effective operation is thus assured. Of course, the technical structure differs for rail networks with low traffic density, less automation or a different traffic management concept.

2.1.3 Definitions of general railway terms

Several general railway terms are used in this thesis. Unfortunately, there are many different definitions for these terms. The terms are thus defined subsequently for a consistent usage in this thesis.

**Capacity:** the total number of possible paths in nodes, individual lines or part of network, in a defined time window, considering a given (e.g. actual) path mix with a market-oriented quality [UIC04].

**Train service intention:** the relevant timetable information including timeframes, stops, connections and train orders for each service along its entire trip [Mah07].

**Stability:** the ability of a system to compensate for delays and to return to a desired state [Han08].

**Robustness:** the quality of a system or process of being able to withstand model errors, parameter variations or changes in operational conditions [Han08].

**Reliability:** the probability that an item can perform a required function under given conditions for a given time interval [IEC90].

**Availability:** the ability to be in a state to perform a required function under given conditions at a given instant of time or over a given time interval assuming that required external resources are provided [RAM99].

**Flexibility:** the ability to adapt quickly to a new environment or new circumstances.

**Performance:** the basic measurement including capacity, stability, robustness and punctuality to benchmark a railway system.
2.2 Railway safety and signalling principles

2.2.1 Railway safety principle

Based on the basic principles and definitions, introduced in the previous chapter 2.1, the safety principles and signalling system are described below.

Railway signalling and operating principles have a significant impact on the effectiveness and efficiency of railway system performance. Particular characteristics will thus be introduced and explained to understand problems, relations and challenges of the railway system.

Safety of rail systems encompasses several aspects. Among other things, preventing collisions between trains and assuring safe train movements are the most important field of actions. Accidents are prevented by rules for drivers, monitoring and control systems for trains and regular inspections of both rolling stock and infrastructure. Furthermore, active train control systems, for example, hot box detectors or flat wheel detectors are used to identify critical conditions resulting, for example, in a derailment or fire on rolling stock.

Braking distances for trains are longer than for other transport modes due to heavy weight, high speed and low friction between wheel and track. Stopping within sight distance is thus impossible in most cases and braking distances can even be up to several kilometres. A variety of principles are therefore used to prevent collisions. One principle is to divide the network in discrete sections, so-called blocks which can be occupied by a maximum of one train. Entry into such a fixed block is only allowed, when:

- the train has a cleared block section ahead;
- the train has a cleared overlap behind the next signal; and
- other train movements are prevented from entering the block section by stop signals.

A route composed by one or multiple subsequent blocks is set in the signal box either automatically or by the movement inspector. Interlocking ensures that safety conditions are fulfilled and that no other train will get clearance to enter the same block.

The information as to whether a train can enter a block is indicated to the driver by a signal. Thereby, the main signal is located at the beginning of a block. To prevent a train from entering a block section which is not cleared to enter, distant signals (indicating the state of the main signal ahead) are positioned before the main signal. The minimum distance from the distant to the main signal is thereby determined in order to safely stop a train running at the maximum track speed allowed before a closed main signal. Fixed signals installed along the tracks cannot be identified at speeds above 160 - 200 kilometres per hour. High-speed trains thus get their movement authority information directly on a display in their cab.
Nevertheless, human error and technical failures can occur and additional safety is required to prevent a train from passing a closed signal. Depending on the technology, automatic train protection systems (ATP) warn the driver in the event of a hazardous situation and apply the brakes if the driver fails to respond to these warnings. Speed and position are thereby taken into account. In addition, several ATP systems also detect excess speed and intervene with an emergency brake.

In addition to ATP, trains can be equipped with a device for automatic train operations (ATO) such that the train runs automatically. The driver’s role is then limited to supervision and he has to intervene in case of an irregularity. Automatic train control (ATC) systems combine the ATO and ATP functionalities. An ATC equipped train can adjust its speed automatically using the codes from the signalling system transmitted along the track.

### 2.2.2 Railway signalling

#### 2.2.2.1 Blocking time theory for fixed block signalling

The time interval in which a block is allocated to a train and blocked for other trains is called blocking time [Pac02]. The blocking time is the most important design variable to estimate infrastructure capacity and to design schedules [Hal59]. The minimum time between two succeeding trains, or headway as it is known, is determined by the blocking time. The blocking time for a fixed block system with lineside signals consist of several elementary intervals [UIC04] (Figure 2.3):

- time to set the route;
- signal watching time;
- approach time;
- running time between the block signals;
- clearing time; and
- route release time.

Figure 2.3 illustrates that the blocking time is significantly longer than the time used to pass the block distance. The occupation and clearance of a track section is detected by a track clear detection device, which could either be implemented by axle counters or track circuits, or by human staff.

It is possible to detect the position of a train using track clear detection devices. This information is used by dispatchers to detect deviations and possible consequential conflicts. This position and time information is discrete data in general. Between two detection points no information
2.2. Railway safety and signalling principles

Continuous train detection systems are available, but are not widespread due to their high cost.

Automated route setting and releasing based on track clear detection systems helps to minimise the fixed, train-independent system times for route setting and releasing. Having a total route-setting and release time of almost two minutes for mechanical interlocking and manual block systems in the 1970s [Pot70], these times are minimised for automatic systems to a few seconds nowadays depending on the complexity of interlocking area and conflicting routes.

In the case of a cab signalling system, the fixed minimum approach time (depending on the distance between the distant and main signal and the track speed) can be replaced by the duration based on the braking time depending on the actual, specific speed and position (braking curves). Therefore, the use of cab signalling systems can reduce blocking times.

By stringing together the blocking times along the line on which a train is running, the blocking time stairways in the time-distance diagram represent the use of the infrastructure perfectly (Figure 2.4). The minimum headway of two trains between two stations, overtakings or crossing points may accordingly be determined. In addition, conflicts in the schedule or during operations can be identified.

Blocking time stairways show that the number of trains operating on a track section is maximised when the blocking time for each train and the speed differences between the trains are minimum. Another insight is that capacity is improved by reducing the block distance. And finally, blocking times depend on train speeds. Higher speeds, on the one hand, reduce the occupation time for
the block section, but increase approach time because of the braking distance extension on the other hand. This leads to a non-linear dependency between travel speed and minimum headway time. Depending on the safety system (especially the fixed block length) and braking coefficient, the optimum travel speed to minimise the headway (and maximise capacity) is between 60 and 100 kilometres per hour [Lan05].

2.2.2.2 Interlockings and remote control

Signals and switches are controlled from signalboxes as they are known. Points and routes are thus interlocked to ensure safe movements. In most signalboxes, the interlocking apparatus only controls points and signals over a short distance. For modern (relay and electronic) interlocking systems, it is technically possible to remotely control several interlockings together from a single central command center by movement inspectors.

Several additional features can thus help to increase the level of automation:

- a track clear detection systems is needed to detect whether a block section is occupied or available for a next train;

- a train describer system is needed in order to assign a train number to the occupation of a block; and

- an automated route setting system where alternative routes are stored and can be set automatically in the case of a conflict or occupancy.

Consequently, automation and, in particular, the remote control of interlockings is an important precondition for efficient and cost-effective rail operation.
2.2. Railway safety and signalling principles

2.2.2.3 Train protection systems

Train drivers get commands and information from two sources:

- data which is independent of the current operational state. This data is mainly based on topology and train characteristics. The information is provided by written rules and fixed signs along the track.

- data which depends on the current operational state. Consequently, this data is a connection between different trains moving in the network influencing each other. The information is provided by signals (either fixed on the track or visualised in the driver’s cab) or as a communication connection between the driver and other actors involved (e.g. dispatchers). Simple signals just provide information on whether a track section may be entered. More enhanced systems can also be used to visualise train speed and distance to next conflict points.

Safety is assured, if operational rules (including signal aspects and speed restrictions) are followed by the driver. To increase safety, automatic train protection systems (ATP) are used to supervise train drivers. Depending on the system’s technical possibilities, deviating driving behaviour (e.g. over-speeding or passing closed signals) is counteracted by a safety action (e.g. an emergency brake). There are two train protection systems: intermittent and continuous.

Data is transmitted for supervision only at discrete points along a track for intermittent ATP systems whereas data is exchanged all the time for continuous systems. On lines with continuous ATP, signals are not needed if all trains are equipped with cab signalling. ATP systems have different features:

- automatic warning;

- automatic train stop (ATS); and

- braking curve (or generally speed) supervision.

Depending on the kind of ATP system, there are different principles of guiding trains and speed profiles which, finally, influence the capacity of the infrastructure [Cat95, RTR01, The07]. Intermittent ATP systems, where information updates and thus speed adjustments (reaccelerations) are only possible at given track points, only allow reduced capacity. In continuous systems, acceleration is possible immediately after the new state is transmitted. Older ATP systems just use the train’s position and speed but no specific train data to calculate the braking curve. New continuous ATP systems take specific train dynamics into account to dynamically compute speed and braking profiles which increase the capacity usage of the infrastructure.
2.2.3 ERTMS: The European Railway Traffic Management System

Railway signalling implies strong connections and interactions between vehicles and track system. Historical developments in the rail signalling field have mainly been nationally driven, resulting in many different systems. Cross-border railway operations are thus restrained due to these technical barriers. At present, more than 20 different signalling and train control systems are used in Europe [Cat95, Pac00, Vin07]. In the late 1950s, initial attempts and a research project to develop an international standard train control system were undertaken [Hür06]. In 1981/1982, the international project was aborted. Instead of a European standard for train control, only the German railways (DB) and their industry partners remained in the project for future cybernetic railways control, resulting in the continuous train control system LZB (Linienzugbeeinflussung).

In 1990, the European Community launched the development of a new European Rail Traffic Management System (ERTMS) [UIC93]. ERTMS combines two principal components:

- European Train Control System (ETCS): a signalling and train control system designed for interoperable rail operation which provides automatic train protection functionality. ETCS has three levels. On level 1, balises (transponders used as data points for an intermittent ATP system) and loops (cables in track used as antennas for data transmission for continuous ATP) are active and transmit safety related data to trains whereas on level 2 balises are only used as reference points to identify the trains position. Both level 1 and 2 are still based on the fixed block system and train integrity is analysed by track clear detection systems. With level 2, fixed signals also disappear and cab signalling is used. Under level 3, trains themselves check their integrity and thus blocks can be variable. ETCS level 1 and 2 are currently under operation in certain corridors whereas further tests and research of level 3 are still needed.

- Global System for Mobile communication - Railway (GSM-R): GSM-based wireless communication developed specifically for data and voice exchange between track and train. With GSM-R, lineside signals can be replaced and signal information is directly transmitted to the train.

ERTMS is intended to improve interoperability. However, capacity can also be increased with ETCS [Eic07]. Especially with ETCS level 2, very short blocks are possible and train headway decreases significantly. Studies also showed, that level 2 with optimised block sections is almost as efficient as level 3; and that operation with relative braking distances will not significantly improve available capacity in comparison to absolute braking distances [Win07, Wen95]. Unfortunately, ETCS was designed as a pure signalling and control system. Operational aspects, such as providing information to adjust driving behaviour and speed, were not intended.
2.3 Timetable principles

2.3.1 Schedule-based rail operations

Schedules have a significant impact on infrastructure utilisation and punctuality. Decisions taken during operations are often based on the timetable. The fundamental principles are therefore described below.

In principle, two general methods are theoretically possible for railway operations: based on a detailed timetable or without a timetable. Regarding passenger trains, a timetable for customer information and trip planning is needed, in Switzerland even by law [Bun98]. The problem of a fixed timetable is that adjustments are difficult to implement. Accordingly, changes of demand or infrastructure availability are difficult to manage and may only be considered with a delay. In particular, demand for freight transport can change rapidly and requests for a train path are more and more short term. Consequently, rail freight traffic can sometimes be operated without a schedule even if all other trains are operated on the basis of an exact timetable.

The schedule has a major impact on the competitiveness of transport and determines expenditure. Therefore short and reliable travel times are needed to attract customers. Optimum resource allocation can make a difference between profit and loss for a railway transport company. Accurate and customer-oriented scheduling is thus a fundamental principle. However, the schedule also has a significant impact on infrastructure usage and punctuality and is linked to traffic management and railway operation tasks. Fundamental planning and scheduling principles are therefore introduced in the following section.

2.3.2 Planning principles for passenger and freight trains

2.3.2.1 Planning challenges

The planning of rail transport is a highly complex task. Many objectives and people involved, departments and companies, conflicting interests and long-life cycle effects complicate the process. Railway companies therefore divide the planning process into a hierarchical process with several steps (phases) [Bus98]. The complicating factor is that structures, standards and requirements between rail freight operators and public passenger companies differ but nevertheless have to be handled consistently for planning.

2.3.2.2 Strategies for passenger rail transport

2.3.2.2.1 Timetable strategies

The timetable is a crucial factor for the success of passenger railways. Frequent, fast and direct connections are desirable but not always possible. The service is thus optimised through
coordinated connections. Basically, we can differentiate between four different timetable strategies [Lie05]:

- **Individual runs**: train runs have no interconnections and are planned on their own.
- **Periodic timetable**: trains operated on the same line and identical directions have scheduled times with an integer multiple offset of the interval time.
- **Symmetric, periodic timetable**: trains operated in a fixed time interval have a symmetry on the same line and both directions.
- **Integrated fixed-interval timetable**: timetables for multiple lines are shifted having one unique symmetrical time in order to provide possible connections between all lines.

### 2.3.2.2 Principle of integrated fixed-interval timetable

The principle of an integrated, fixed-interval timetable is generally regarded as superior [Hat07, Sch95, Sto97] and applied in the Swiss rail network. This timetable strategy has several advantages for customers, but it has also negative consequences which will be pointed out in the following. The principle of an integrated fixed-interval timetable is that all trains arrive shortly before the symmetrical time in main (hub) stations (Figure 2.5). Subsequently, passengers can change from one train to another with a minimum amount of unnecessary and unpopular waiting time. After completion of the dwell process some minutes past the symmetrical time, trains leave the station in all directions.

Figure 2.5: Principle of the integrated fixed-interval timetable
The integrated fixed-interval timetable provides an optimum transfer system and results in high accessibility and generally shorter travel times for passengers. The periodicity of the timetable also simplifies the planning and analysis of the timetable. Thereby, only one system hour (cycle time) must be regarded and adjusted for peak or off-peak hours. Another advantage of the integrated fixed-interval timetable is improved communication. The schedule maps graphically summarise all the relevant information and thus timetables are easy for customers to memorise [Max99]. A section of the 2005 Swiss integrated fixed-interval timetable for the area around Lucerne is visualised in Figure 2.6.

Figure 2.6: Example of a section of the integrated fixed-interval timetable for the area around the main station of Lucerne (source SMA und Partner AG Zurich, Timetable 2005)

However, several requirements have to be satisfied to apply the concept of the integrated fixed-interval timetable. First of all, the interconnection stations must have a sufficient number of station tracks in order to guarantee that passengers can change between all trains. The tracks in front of the stations must also be designed to enable all trains to arrive and leave within a short time (e.g. by flyovers). This leads to the effect that the infrastructure around main interconnection stations is mainly used around the symmetrical time and almost empty in-between. Figure 2.7 illustrates the scheduled fluctuating usage in Zurich’s main station for a regular hour in 2006. It can thus be seen that around the full and half hour, the station is almost fully occupied by Intercity and long distance Interregio trains making connections whereas commuter trains arrive and depart continuously throughout the entire hour having in most cases dwell times of only 2 to 3 minutes.
A second requirement is that travel times between interconnection stations must be strongly coordinated to make sure that they have also an integer multiplicity of the cycle time. Infrastructure enhancements or changes in rolling stock (e.g. tilting trains) are ways of achieving the required travel times. Figure 2.8 shows the planned line network and system nodes for Switzerland for 2030. The infrastructure enhancements and purchase of new rolling stock was carried out iteratively in several stages in order to implement cost-effective projects first [Bub07].
Another characteristic of integrated fixed-interval timetables is the large number of interconnections and dependencies in the schedule. Two different types of interdependencies are especially typical and common for integrated fixed-interval timetables:

- scheduled connections for passengers in stations to change trains; and
- dense sequences of trains (few buffer times between trains) around station areas.

The number of interdependencies reduces the timetable’s robustness, and more conflicts and unnecessary trains stops are the immediate consequences. Consequently, even small initial delays can result in the fast propagation of knock-on delays over the entire network. To reduce this domino effect, connections have to be cancelled frequently even for small delays. From the customer’s point of view, these decisions are not acceptable, and often unnecessary.

2.3.2.3 Planning principles for freight carriers

Integrated timetables are optimised for passenger trains. However, slots for freight trains are also pre-planned in a yearly schedule to assure adequate capacity for freight trains. One problem is the greater differences in the dynamics of freight trains in comparison to passenger trains. These variations combined with the fixed pattern of passenger trains result in the sub-optimum utilisation of the infrastructure. The large delays and unreliability of freight trains entering the network are further challenges. Intelligent route search algorithms are thus needed to allocate free slots for freight trains and to maximise the use of the infrastructure in the short-term [Luc00].

2.3.3 Basic principles for schedule design

2.3.3.1 Planning sequence

Generating a new timetable is a complex process. Even with the aid of computer programs, calculating a new timetable from scratch for an entire, complex network (for example for Switzerland) within a short time is not possible today. As a result, new timetables are usually modified based on the previous year’s timetable in critical sections by shifting train paths for some minutes, extended by the addition of new services or adjusted to a new infrastructure. Comparisons between several different timetables for an entire network, for example, in the Netherlands [Gov02], helps to estimate timetable stability and effectiveness analytically.
A timetable is developed in several phases starting possibly with a conceptual idea about 10 - 20 years before implementation and ending as input for daily operation (Figure 2.9). The accuracy and level of detail increases the closer a phase gets to operation. Thereby, strict borders can be identified:

- 5 years before schedule is in operation: no more new (major) infrastructure adjustments are possible.
- 3 years before schedule is in operation: purchase of new rolling stock fleet is no more possible.
- 1 year before schedule is in operation: contracts with orderer for regional traffic are signed.

In the final phases, planning is strongly connected with defining schedules for staff and rolling stock. In particular, during daily operations, a new schedule must ensure that, after an incident or delay, the rolling stock and staff are ready to run a consecutive or new service.

![Figure 2.9: Planning phases](image)

The improvement criterion in network and timetable planning also changes throughout the phases (see Figure 2.10). At the strategic long-term planning level, economic aspects dominate and capital gain is improved. The maximum possible modal split for railways is also a target. Subsequently, planners try to maximise capacity or the number of trains running on the infrastructure as part of the tactical procedure. Finally, the timetable has to be reliable during daily operations. Interfaces between different planning levels may be sources of errors, data loss and inefficiency. The exchange is usually not automated and manual input is still needed nowadays.

![Figure 2.10: Improvement objectives in timetable planning](image)
2.3.3.2 Timetable components

The final schedule contains different time components:

- the technical minimum run and dwell times;
- running time and dwell time supplements;
- time supplements (or constraints) for connections as part of the integrated fixed-interval timetable; and
- buffer times between trains.

Physical constraints determine the size of minimum run and dwell times. The amount of supplementary and buffer times added in the schedule is based on experience as well as the intended quality (punctuality) and is finally a strategic decision. The different types of time, resulting finally in a schedule, are visualised in Figure 2.11.

![Figure 2.11: Components required to design a schedule for an integrated fixed-interval timetable with symmetrical time at full hour](image_url)
2.3.3.3 Fundamentals for run and dwell time calculations

2.3.3.3.1 Calculation of technical running time

The technical (or minimum) running time can be determined in two ways. The first way is to solve differential equations analytically using train dynamics and topology information with the help of simulation tools or analytically [Wen03]. Assumptions for the train’s weight (load factor, large variations for freight trains), adhesion (weather condition) or braking behaviour (driver) influence the result. The second possibility is to carry out test runs to determine the minimum running time. These results are also used to calibrate parameters in the analytical model.

2.3.3.3.2 Dwell time calculation

Planners often neglect the boarding and alighting process as the basis for the dwell time. Planners simply use predefined dwell times, based on a station and train classification to calculate the schedule. Experienced-based dwell times are also rarely used to classify a stop. These predefined values then contain both minimum dwell times and supplement times. Nevertheless, there are various models to calculate dwell times more precisely [Wei95, Lee07].

The main influencing factors in these models are passenger demand or the number of people involved in the boarding and alighting process as well as the capacity and layout of rolling stock, boarding conditions (design of entrances) and station (including the positions of access points). Changing demand over the day and year as well as variations in passenger distribution and random statistical spread makes it difficult to determine an appropriate dwell time. This problem especially arises for fixed-interval timetables where demand changes over the day cannot be taken into account. Defining an adequate mean value as input in the model is thus a strategic decision, too. Nevertheless, the use of precise models helps to avoid schematic delays because of inadequate dwell times and their propagation in heavily used rail networks [Nas07].

2.3.3.4 Buffer times and time supplements

2.3.3.4.1 Types of buffer and supplement times

Three types of buffer and supplementary times are applied in train operating schedules:

- running time supplements;
- dwell time supplements; and
- headway buffers between two consecutive trains.
Both running time and dwell time supplements are used to absorb variations in daily operations and are additionally used to recover from delays. Headway buffer times reduce the transfer of delays from one train to another and thus minimise the propagation of delays in the network.

There are no general rules describing the amount and distribution of buffer and supplementary times in a schedule although recommendations are proposed [Deu02, Pot70, Sch74, Sew04]. In the final analysis, the size of buffer and supplementary times is always a strategic decision based on several factors:

- level of planned interconnections;
- demand variations;
- expected or planned maintenance tasks;
- characteristics of the topology (number of single-line sections);
- signalling, safety and dispatching systems; and
- operation rules.

### 2.3.3.4.2 Running time supplements

Running time supplements are used to reduce the impact of running time variations caused by changing weather conditions or varying train dynamics. In general, rail infrastructure managers apply different supplements to the running time.

- Firstly, a given percentage of the minimum running time is added. Usually, this value is rounded-up to form an absolute supplement value (between 1 and 3 minutes, depending on the distance to the next node) and added in front of larger stations. This is used to recover from small delays.

- Secondly, a value of 1-2 minutes is added along a line used for smaller maintenance work throughout the year.

- Thirdly, for larger maintenance or construction tasks, running time supplements are specifically calculated and added to the regular minimum running time.

All in all, running time supplements are usually between 5-10 percent (around 7 percent at SBB, standard minimal supplements of 3 percent at DB [Deu02]) of the technical running time. Taken together, these running time supplements may thus increase travel time significantly. Consequently, timetable constraints especially for integrated fixed-interval timetables result in running time supplements being either smaller than desired or excessively great depending on the calculated running time and the schedule’s cycle time.
2.3.3.4.3 Dwell time supplements

Dwell time supplements are used to reduce the impact on changing demands and variations in the dwell process. As already mentioned, dwell times are either defined on the basis of predefined values or are based on an assumed number of passengers with a given distribution on the platform and in the train involved in the dwell process. In order to cope with variations such as higher number of passengers, supplementary times are added. Based on long-term data recordings, principles and variations are well-known [Buc08]. Figure 2.12 illustrates the connection of dwell duration and number of passengers involved in the boarding and alighting process for Zurich’s S-Bahn. It shows the lower limit, mean values and variations in the dwell time.

Figure 2.12: Dwell duration as function of the total number of passengers involved in the boarding and alighting process for Zurich’s S-Bahn DPZ

The strategic decision for dwell-time planning is which trains may experience delays during the dwell process and what amount is acceptable. In larger stations with connections, the departing time of trains is influenced by both dwell time supplements and connection times. Comparable to running time supplements, dwell time supplements may also be used to reduce delays.

2.3.3.4.4 Buffer times between two consecutive trains

Headway buffer times are used to reduce the propagation of knock-on delays in the network. Reducing the size of buffer times results in denser rail traffic and leads to a disadvantage in that the number of interdependencies between train routes increases as does, ultimately, the
number of potential conflicts. Consequently, a single small disturbance can have greater impact on the whole network. Furthermore, the increased number of trains makes it more difficult for dispatchers to identify optimum strategies for reducing knock-on delays quickly in order to prevent delays from propagating throughout the network.

Nowadays, blocking time theory is still rarely used to generate a schedule. Instead, fixed headways between two trains are assumed. Consequently, it is possible that schedules may not be conflict-free. Nevertheless, fixed headways often already include a buffer time between two consecutive trains. However, the result of this unformalised and non-consistent policy is that decisions are usually based on the experience of the planner and insights from the analysis of recorded train data. Precise and correct declarations of infrastructure utilisation are therefore not possible and also maximum utilisation depends on the planner and may vary significantly.

2.3.3.4.5 Unplanned capacity

Buffer and supplement times are used in cases of a delay or an event to reduce the delays, their impacts and its propagation. Another buffer that may be used during dispatching is unplanned capacity. Unplanned capacity means train paths and routes not scheduled for rail operations during a given time period. Thus, the infrastructure not used in the schedule also represents a buffer. This buffer time can also be used for maintenance work if possible. In general, the UIC proposes a factor of 15-25 percent of unplanned capacity during peak times [UIC04].

2.3.4 Railway traffic data analysis

2.3.4.1 Necessity of ex-post traffic data analysis

The ex-post analysis of operational data is gaining in importance. Statistical reports and specific contractual performance requirements (including bonus-malus payments based on statistics) are becoming increasingly common. Using and aggregating available data allows us to detect delay developments, conflicts, running and dwell time variations, delay-prone trains, or capacity bottlenecks. This could be done for multiple trains within a given area or corridor over a specified time period and with the help of specific programs as for example OpenTimeTable [Nas04b, Ull05].

The data can also be used to identify fundamental correlation or insights of the production process. For example, the dwell process or the behaviour of train drivers and possible influencing factors can be analysed. Finally, the data can also be used to verify how efficiently new operation processes are applied.
2.3.4.2 Data analysis framework

Accurate train and infrastructure data from the entire network are the basis of the ex-post analysis. Operations can be analysed precisely and events and their following consequences can be identified. What would be ideal would be a fully observable system including continuous data on:

- route and signal states and their changes;
- infrastructure state;
- train dynamics with speed, acceleration rate and position;
- dwell process data including specific door open time, the number of boarding and alighting persons and their distribution in the train and on the platform.

However, technical restrictions make it impossible for all data to be recordable or available. Nowadays, data is recorded by the infrastructure at given places (mainly signal positions), by onboard units in the train or measurements by the staff at stations. The way in which rail operation data is recorded limits the number of available parameters. In the early days, train arrivals or departures were noted manually with an accuracy of minutes in some given stations. Automation nowadays technically allows to obtain precise data at selective points wherever a sensor is located in the network.

The data recorded from the train detection system in the infrastructure consists of the exact time, the recording position and the train number. Data in the SBB network is available every 200 to 5,000 meters. Gaps in the detection system results in the temporal behaviour of trains often not being identified absolutely reliably and in the behaviour between two detection points having to be estimated.

Due to missing or relatively large distances between consecutive detection points, exact arrival or departure times are also often not available and have to be derived or estimated from other data. Consequently, an larger density of recording points improves accuracy, which is finally important in station and junction areas. The missing speed information can thus be obtained from estimates based on the combination of available recorded data and train information. The tool TNV-prepare, developed in the Netherlands [Gov00, Gov05], allows such a combination of track data (passing times and route states) to determine down to seconds the exact point of time at which a train stops or passes a given point within a station area.

Unfortunately, speed and other train characteristic data are not recorded by the infrastructure. Some trains collect specific data such as speed, acceleration factor, aspect of passed signals. However, this data represents a substantial volume and often requires manual export and processing. As a result, this data is typically not used for analysis. Also, the separation between railway undertaking and infrastructure operator makes it more difficult to exchange this data.
2.3.4.3 Findings based on rail operation data analysis

2.3.4.3.1 Variations during rail operation

Extended statistical analysis on recorded train data is available from the Netherlands and Switzerland [Yua06, Yua07, Lüt05]. The results from the two countries are comparable. For example, the analysis of arrival distributions showed that the two related lognormal and log-logistic density functions fit well in most cases. According to the dwell time analysis in Switzerland, the log-logistic density function also fits best for train running times if trains that experienced a conflict are filtered out [Cha07]. A detailed analysis of running times including the illustration of distribution curves is in Appendix E. An example of the log-logistic density function is

$$f(x) = \begin{cases} \frac{\alpha}{\beta} \left(\frac{x - \gamma}{\beta}\right)^{\alpha - 1} & \text{if } x > \gamma \\ 0 & \text{otherwise} \end{cases}$$

with the parameters $\alpha = 2.51$, $\beta = 9.11$ and $\gamma = 244.98$ is visualised in Figure 2.13 for the running time between Bubikon and Wetzikon of S-Bahn commuter train number S18526.

![Figure 2.13: Comparison of measured and fitted log-logistic density functions for 230 recorded weekday morning rush hour train trips of S-Bahn train number 18526 in 2004 from Bubikon to Wetzikon](image)

The dwell and train departure process is a common reason for delays. Beside external reasons such as waiting for late connecting trains or blocked routes, late passengers, blocked doors, high demand, unequal usage of doors (passenger groups), technical problems or missing staff are also reasons that cause late departures. In general, the variation is higher for the dwell process in comparison to the running time. For running trains (without conflicts), 80 percent of all trains are usually between $\pm 3-5$ percent around the median running time (data analysis based on SBB train recordings from March to November 2004). Measurements of the dwell
process (only for commuter trains with a delay of at least one minute) show that the middle 80 percent vary between ± 10-25 percent of the median dwell time (see Figure 2.14). Also the dwell time cannot be described a priori by a unique density curve. Nevertheless, if dwell times are analysed where boarding and alighting determine the overall dwell time, Weibull, normal and also log-logistic density curves fit best.

The variations for both running trains and trains in stations determine the required size of buffer times between trains for stable operations and thus limit capacity. New operation principles thus have to ensure that variations are shortened for all processes.

2.3.4.3.2 Driving behaviour today

A more detailed analysis of the running time (with trains having a conflict not being taken into account) and their possible influence parameters showed that in most cases, the actual delay of a train has no significant impact on running time (see for more details and examples Appendix B). The general assumption that late trains run faster than early or on-time trains can thus be refuted. Exceptions from this behaviour are only measurable punctually for some drivers and exclusively for Intercity or Interregio trains. Thereby, some train drivers slow down intentionally or start coasting early if they are on-time or ahead of the schedule.

A second important observation from the recorded data is that the running time from a first stop station to a subsequent second stop station has no significant correlation with the running time from the second stop station to the third stop station. The expected correlation, e.g. due to bad/excellent adhesion conditions, high/less load, was, as shown in Appendix B, not recorded in general.
2.3.5 Overview of recent research on planning principles for more robust and stable timetables

Dense network utilisation and robust schedules are the targets of the planning tasks. For many decades, scheduling has been done mainly manually based on experience and supporting tools [Hür06]. Due to the increasing number of trains and more complex infrastructure, research on train scheduling also became the focus of intense discussion. In addition, schedules influence online characteristics and therefore require special attention. Finally, online rescheduling tasks and offline timetable planning are comparable tasks. Consequently, research on offline scheduling helps to improve online rescheduling and vice versa.

Research in offline scheduling can be broken down into three main directions:

- tools that help planners to design a new (feasible, network-wide) schedule;
- tools for optimising a given schedule (mainly in a local sub-area of the network); and
- tools that analyse a given schedule as input for optimisation.

Good overviews and examples of developments are given in Carey and Lockwood [Car95], Cordeau et al. [Cor98], Ghoseiri et al. [Gho04], Goverde [Gov05], Hansen and Pachl [Han08], Huisman et al. [Hui05] or Lindner [Lin00].

Recent research and developments focused on different planning topics to obtain a more efficient usage of the infrastructure:

- Transformation of customer demands into scheduling input: an initial problem with scheduling in general addresses the transfer of customer demands into a description usable for scheduling. To solve this problem, the concept of global service intention is under development at the SBB with the aim of modelling and collecting demands [Lau08, Cai09d].

- Usage of sophisticated scheduling tools: supporting systems help planners during the design process to get a feasible schedule. Tools visualising schedules (for example Viriato [SMA06] or or the newly developed NeTs in Switzerland [SBB05]) can thus be distinguished from more enhanced programs with partially automated route search or scheduling functions (like DONS in the Netherlands [Hoo96, Hoo98], or CAPRES [Luc00]). However, a fully automated scheduling process has not yet been achieved due to the complexity of the networks, the enormous number of routing possibilities and train timing possibilities.

- Strategies for buffer and running time supplement distribution: buffers may sometimes be reduced at particular places and increased in compensation at other parts in the network to obtain robust or efficient schedules. Intelligent distribution of buffer times
and running time supplements is thus of major interest. The target is to develop robust and stable schedules with a minimum amount of supplementary and buffer times. Possible strategies, recommendations and results are described by Kaminsky [Kam01], Rudolph [Rud04, Rud06], Kroon, Dekker and Vromans [Kro05, Vro06].

- More accurate planning by applying the blocking time theory: in order to reduce imprecise planning with constant headways (resulting either in conflicts or unnecessary and capacity wasting unused times), accurate planning (seconds) based on blocking time theory [Wen07] helps to develop conflict-free and precise schedules with reduced unused times especially in bottleneck areas.

- Automated routing of trains through complex station areas: larger station areas offer a multiplicity of routing possibilities that are impossible to handle and optimise by humans. Optimum routing can use the infrastructure more efficiently. Research on this topic is described by Zwaneveld et al. [Zwa01], Burkolter, Caimi and Herrmann [Bur05, Cai04, Her05].

- Applying a micro-simulation to improve utilisation or robustness: delay-prone capacity bottleneck areas may be analysed and selectively improved [Hür02, Lüt05, Nas04a, Rad04]. However, this efficient improvement is time consuming, has to be done manually and only for limited areas. In contrast to other deterministic approaches, stochastic distributions and multi-simulations are used to evaluate schedules.

- Pre-analysis of a timetable concerning feasibility, stability and robustness: new schedules have to be executed and analysed before their application to detect feasibility and stability problems. Weak points in the network may then be selectively improved. Example works for this topic are described by Goverde [Gov05] and Vromans [Vro05]. Capacity bottlenecks, delay crucial areas or delay prone trains can be identified before trains are operated. This data may be used for special attention during operation or for improvements. In the same way, empirical data from daily practice may be used to identify weak points in a schedule to improve performance.

Altogether, new planning methods and algorithms can produce high-quality, conflict-free, stable and robust timetables [Han04]. This is the basis for increasing capacity utilisation and performance on a given network. The temporal granularity of the schedule is thereby reduced from minutes to seconds. Simultaneously, intelligent distribution of buffer and supplementary times improves the robustness and stability of a schedule. Nevertheless, significant research is still needed to develop a network-wide schedule by a computer in a reasonable time. Consequently, the methods described are mainly applied iteratively and in combination with manual tasks.
Chapter 3

Feedback control principles for railway operations

3.1 Introduction to control and railway operations

Temporal variations during production are common everywhere. Railway operations are also affected by these variations. As introduced in the previous chapter, stable production is assured by adding buffers between trains and time supplements which, in the same way, reduces the maximum possible capacity. Feedback control methods allow variations to be reduced and enable deviations to be counteracted earlier, resulting in less knock-on effects, too. Consequently, allocated buffers may be reduced to improve performance. Therefore, the goal is to change the existing rail operation process, which is similar to a functional chain, into a system of superimposed feedback control loops.

First, an introduction to general feedback control principles is given in this chapter. Following this, existing rail operation principles and conventional rail traffic management are described and weak points identified. Based on the analysis, the new integrated real-time rescheduling framework is introduced and described. With this, the fundamental feedback control principles are adapted for railway operations in order to achieve the increase in performance required.

3.2 Feedback control principles

3.2.1 Feedforward and feedback control

Developments for improved railway operations are based on models using fundamental control theory. Subsequently, control principles, used later in this thesis, are presented. Control systems are found throughout the technological world and are a common concept. With the help
of controllers, a system (process or plant) is made to conform to a desired value or state (input reference). Two basic system structures are thereby distinguished [Fra95]:

- open loop (or feed forward) control (illustrated in Figure 3.1); and
- closed loop (or feedback) control (visualised in Figure 3.2).

In open loop control systems, no information on the actual system state is used. In particular, disturbances or deviations are not observed and could thus remain in the system.

![Figure 3.1: Schematic layout of an open-loop control system](image)

In contrast to open loop controllers, specific information from the system (plant or process) measured by sensors is made available to the controller. Responses to disturbances or deviations are examined in this case. In general, the control input generated by the controller is supplied to a actor within the plant. It has to be mentioned that possibly not all system states could be measured and that the measured output could be different to the system output due to some measurement noise [Bur99].

![Figure 3.2: Design of a generic feedback control system](image)

A cascade control system is a multiple loop system where the primary variable is controlled by adjusting the set-point of a related, superimposed secondary variable controller. The inner control variable is affected by the outer loop through the process. Cascaded control is often applied to divide and control a difficult process into two parts [Mor89]. The outer control loop thereby is mainly formed to handle major disturbances whereas only minor disturbances have to be controlled by the inner loop. However, the inner loop must respond to disturbances faster than the outer loop. Also for non-linear processes including actuator saturation, cascade control can improve system performance in comparison to a conventional single loop controller. Figure 3.3 illustrates a cascade control system block diagram.
Control systems are widely used in railway systems. In modern locomotives, traction control systems prevent wheelspin and optimise the traction force to accelerate the train. Another common feature for modern trains is cruise-control functionality. Open-loop controllers are also in operations such as the automatic train protection, door closing or air conditioning systems to name some examples. Also interlocking systems may be described as a feed-forward control system. Driver-less metros, where trains are operated fully autonomously, are an example of closed control loops operations.

### 3.2.2 Control and optimisation

For larger systems like standard gauge railway systems, more layers are used (see section 2.1.2). In general, the hierarchical structure can be divided into two types:

- Control: the processes are commanded or manipulated in a way that the output follows a calculated reference trajectory as precisely as possible.

- Optimisation: the desired references (input for the controllers) are computed taking into account the actual situation of all processes and the global optimisation target (or initial schedule).

Optimisation is part of the tactical and strategical layer in the railway functional structure. Improvements are aspired for the network design and during timetabling. Usually, the distinction between control and optimisation is primarily a separation of time scales with fast reactions and triggering frequencies for control systems and slower updates for optimisation functions. In general, the information flow in such hierarchical systems is based on sending reference values from the above or outer part to the lower, inner part and reporting back state (or problem) reports. Such systems are widely used for distributed (non-centralised) systems. Two types of optimisation can be distinguished:

- global (centralised) optimisation; and

- decentralised (local) optimisation.
In the classical, functional railway hierarchy, the scheduling task, offline during the planning phase or online as a dispatching measure, is therefore part of the optimisation. Driving the train, in contrast, is a control task, designed as a feed forward or feedback system depending on the level of automation.

Another concept would be an integrated approach with optimisation and control in one [Gee07]. The optimisation and control actions in an integrated optimisation and control system are perfectly coordinated. Integrated numerical optimisation and control is best suited to complex multi-input systems with coupling and constraints [Gla03]. In particular, model predictive control (MPC) (or also known as receding horizon control) is becoming popular because computation power has reached a level that allows the generation of a solution in real-time for small sampling times. The core concept of MPC is to take the current state at each sampling time and solve the control problem over a finite horizon [Gar89]. Computation is repeated at each time step, regarding the new state, constraints and optimises a given function. However, integrated solutions like these are often not applicable to larger systems for several reasons such as lack of computing power, distribution or complexity of the processes [Sko96]. This results in the fact, that only a few applications in the railway field try to implement predictive control. One of the examples deals with dispatching [Weg08] and is a pure form of optimisation without integrated feedback control. Nevertheless, possible applications of integrated optimisation and control are conceivable. Figure 3.4 illustrates the different possible optimisation and control systems.

![Figure 3.4: Possible control and optimisation architectures. Left: centralised optimisation in hierarchical system; middle: decentralised optimisation in hierarchical system; right: decentralised optimisation and control in an integrated architecture](image)

3.2.3 Definitions of railway operation terms

Control and optimisation structures have been developed and applied to various branches of industry and productions. In contrast to other areas, standard railway systems are characterised by their high degree of cross-linkings (and thus dependencies), their large geographic distribution, the high number of actors, resources and customers involved, high safety and temporal requirements, reduced observability and missed continuous data and a high number of external influences (disturbances). Therefore, solutions, strategies and methods from other fields cannot
be directly transferred and must be newly designed. In addition, many specifically used terms are not uniquely defined and are used with varying meanings.

To ensure consistent usage in this thesis, we have adapted definitions from general control, optimisation and rescheduling [Vie03] especially for railway operations:

**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. The production plan also defines the reference input for all the actors involved.

**Conflict-free production plan:** is if there are no overlappings of the blocking time graphs (blocking time theory).

**Feasible production plan:** if it is capable of being accomplished. In particular, a feasible train trajectory must be able to be carried out with the given train dynamics on the actual infrastructure with specific state. Also dwell times have to be capable of being carried out and constraints of resources (staff and rolling stock) must be taken into consideration. Freedom from conflicts is part of a feasible production plan.

**Resources:** are used to effect output (operation of trains and infrastructure). Resources in general contain personal actors (drivers, guards, infrastructure operators, movement inspectors, dispatchers), equipment (rolling stock, infrastructure as signal box etc.), and the energy to be supplied.

**Rescheduling:** is the process of updating an existing production plan after a deviation, event or disturbance based on the system’s current state and predicted behaviour.

**Integrated Real-Time Rescheduling:** the combined process of updating an existing production plan (schedule) in real-time, and executing the new plan precisely with the assistance of communication and supporting tools. In other words, a new schedule would be developed based on the current system state transmitted and applied by all actors (e.g. drivers, infrastructure operators, conductors) with the help of technical devices (i.e. man-machine interfaces and/or fully automated systems).

Further terms used in this thesis including the characterising attributes of rescheduling are defined below:

**Rescheduling point of time:** describes the point of time, when a new production plan becomes valid.

**Rescheduling period:** is the duration between two consecutive rescheduling points in time. The rescheduling frequency is the inverse of the rescheduling period.
Rescheduling duration: describes the time span between when a predefined threshold is exceeded until a new production plan is transmitted to the actors.

Rescheduling stability: measures the number of changes to the production plan over a given time period. Rescheduling nervousness is the inverse of rescheduling stability.

Rescheduling robustness: measures how disruptions (events) reduce performance.

Rescheduling performance: measures the quality of the schedule and its updated version after a rescheduling. The performance targets are based on the rescheduling (dispatching) objectives (see also optimisation criterions in section 3.3.2.4) used to make decisions in the rescheduling generation process and to minimise additional costs by cointaneous maximisation of the benefits. The primary cost factors used to describe performance after a rescheduling are additional delays, missed connections, or additional resources (staff and rolling stock). These factors are part of the optimisation function in the rescheduling algorithms.

Rescheduling strategy: specifies the way a deviation or event is handled either by applying heuristic rules or generating a new production plan. Two main strategies can be distinguished:

- Dynamic scheduling: instead of generating a new schedule, dispatching rules and other heuristics are used to handle irregularities. The decisions are made based on the current state. This is the way that dispatching in rail control centres is executed nowadays.
- Predictive-reactive rescheduling: an initially given schedule is updated (or newly generated) after a disturbance, deviation or other event to minimise its impact on performance. For predictive-reactive strategies, rescheduling can be initiated in three different ways: periodically, event-driven (after exceeding of a given tolerance) or hybrid (mix of event-driven and periodical).

Rescheduling method: describes the method used to update a schedule (as part of a predictive-reactive rescheduling strategy). Three main types are possible:

- Right-shift: the entire production plan is shifted by a given value (e.g. all trains are scheduled 2 minutes later).
- Partial: only part of the production plan is updated (or replaced by a new schedule). The partition thereby could either be regional for certain trains or limited by the time horizon.
- Regeneration: the entire production plan is newly generated.
3.3 Railway operation principles

3.3.1 General rail operation model

Developing an adapted framework including new optimisation and control tasks requires a detailed analytical treatment of existing processes and the methods involved. Conventional railway operation principles and recent research activities with the focus on traffic management are thus analysed and described in this section as a basis for the new framework introduced below in section 3.5. The analysis of conventional railway operation principles thereby focuses on identifying weak points and introduces structured descriptions for several tasks combined with dispatching. Based on this, requirements for the new framework are also derived.

No measures are required as long as the infrastructure is operated without any failures and trains are running on time. Dispatchers monitor the train positions according to the occupation state of the blocks. The information update in the traffic management center on the train positions is based on automated position reports from the track clear detection devices when a train passes such a detection point. However, the dispatchers only have the information if a track section is occupied, cleared or reserved. Observation is thus also time-discrete, causing time lags to occur until a deviation or disturbance is detected.

In the case of a deviation or failure, dispatchers decide on measures (e.g. setting alternative routes, or reordering the train sequences) which are usually communicated verbally by phone to the staff in the control centres. The measures taken by dispatchers are thereby heuristic decisions based on their experience. Therefore, no updated schedule is generated either. Oral communication is used in addition to provide or ask for additional information or feedback between dispatchers and movement inspectors in control centres. The consequences of a measure can be observed later in the traffic management center when trains either have a new route or a new sequence. Data flow between control center and traffic management system can therefore be described as a functional chain.

Dispatchers usually do not inform drivers and guards when they take measures. Drivers only get information indirectly through the signal commands. Oral communication between train staff and dispatchers rarely occurs and adjusting the driving behaviour is therefore uncommon. Drivers often just use the information of the static schedule irrespective of the state of other trains.

Driving the train can be designed as a separate loop: The data flow or control loop is initiated by trains passing detection points causing a data event trigger. As a consequence, routes and signals are set by the automation and safety system. Later, drivers recognise the signal state and take appropriate actions for safe and if possible punctual operations causing further data triggers at next detection points. In general, single processes are linked as functional chains but not control loops.
The conventional railway operation principle described is illustrated in Figure 3.5. One main characteristic of the two loops and the actual rail operation in general is the variation for the information transfer between the actors. Variations exist for information-update frequency, volume and accuracy of information, transmission lag or transmission medium.

3.3.2 Dispatching

3.3.2.1 Dispatching principles

Timetable perturbations are common for railway systems and can cause conflicts and delays. To reduce the impact and propagation of a delay or an incident within a network, dispatchers coordinate traffic and determine appropriate measures. When doing this, dispatchers may decide on changes to the existing schedule and communicate them, but they do not directly influence or control the railway infrastructure or the trains. Beside supervising traffic and communicating new schedules, dispatchers are also responsible for informing passengers about the new schedule. For bigger problems such as interruptions to the infrastructure or trains breakdowns, dispatchers organise alternative possibilities using new trains on different routes or with substitute bus services.
3.3. Railway operation principles

Usually, a network is divided into several areas. The size of a specific area which is assigned to and handled by a single dispatcher, can vary between several tens up to several hundred kilometres depending on the complexity and density of traffic [Whi03]. The separation requires dispatching measures to be coordinated between the regions.

3.3.2.2 Dispatching process

Dispatching is a sequential process and can take place over several temporal levels. Three different phases of dispatching are known (see Figure 3.6):

- Pre-dispatching: immediate measures are taken to limit the consequences, thereby gaining time to gather information and to determine new schedule possibilities.
- Main Dispatching: elaboration, communication and application of a new schedule.
- Post-dispatching: normal operation including rolling stock and crew circulations is established.

![Figure 3.6: Dispatching process](image)

Depending on the problem or primary delay reason (introduced in section 3.3.3) as well as the specific circumstances, not all three phases are necessarily applied.

For a fast first reaction after a disturbance, predefined checklists or activity lists are used to determine the initial measures. Together with more precise data about the disturbance and predicted future behaviour, these measures are used as input for the main dispatching process.
Conflicts due to a delay lead directly to the main dispatching process. For disturbances where resources (trains or staff) are in the wrong sequence or in the wrong position, a post-dispatching process is needed after troubleshooting. Thereby, original circulations are restored and normal operation achieved. The long duration of the different phases results in new events are overlaid with the existing dispatching process. Thus, conditions are changing continuously.

### 3.3.2.3 Dispatching measures

The challenge for dispatching is the number of consequences a single measure can cause. The interconnections between the trains can result in fast and network-wide delay propagation. In addition, the number of possible dispatching actions is enormous and their impact is almost impossible to predict without the aid of supporting tools.

For small delays, retiming, reordering and rerouting of trains are common dispatching actions. Also braking connections or transferring the crossing points are possible measures after a delay. For larger delays or events with larger geographical dimensions, stops can be added or cancelled, services can be cancelled, removed or substituted by buses.

The dispatching measures applied depend on many factors. Therefore, an efficient train dispatching support system is a fundamental prerequisite for stable operation in a dense and finely meshed rail traffic network. Consequently, most of today’s dispatching systems have visualised time-distance graphs. Some of these systems can automatically detect conflicts. However, solving conflicts or suggesting possible dispatching solutions are not yet available. Heuristics and experience are thus used by the dispatchers. The applied dispatching measures therefore depend strongly on the dispatcher and his experience [Wik04].

### 3.3.2.4 Dispatching objectives

The optimisation goals and temporal requirements for the dispatching process depend on many parameters. Among other things, dispatchers determine some of the most important influence factors which are:

- primary delay reason;
- location, where an event or deviation has occurred in the network;
- traffic density and rerouting possibilities;
- time of the day; and
- number of passengers involved;
3.3. Railway operation principles

Optimisation objectives can therefore vary significantly. Common optimisation goals are:

- minimise total (weighted) delay;
- ensure maximum number of connections;
- maintain circulation plans for all resources (rolling stock, drivers, guards); or
- ensure that all stations are served with minimum number of additional trains or supplementary services.

While considering the optimisation goals, it must be emphasised that passengers are of maximum importance in the rescheduling process, regardless of what additional actions need to be taken. However, it is always essential to determine whether the extraordinary costs and personnel expenses that are incurred by these additional actions are reasonable and within given limits. Objective functions are introduced in section 4.3.4.1 to compare different dispatching measures.

3.3.3 Categorisation of primary delay reasons

As mentioned earlier, the type of or reason for a delay may have a significant impact on the dispatching measure as well as on other processes. Consequently, categorising the primary delay reasons helps to structure process chains and in addition helps to visualise the temporal demands. We have therefore introduced and categorised four basic types of primary delay reasons:

**Disturbance:** A disturbance means that due to the reduced availability of a technical component, or of a actor participating in the production, production cannot be continued as planned. After the disturbance has been identified, the production plan must be adjusted.

**Incident:** An incident interrupts or delays production on a short-term basis. After an incident, all resources are fully available again and production can continue as planned. Incidents often lead to schedule deviations. Incorrect input or other human errors are also classified as incidents.

**Deviation:** The most common type of deviation is a time deviation, specifically exceeding a pre-defined tolerance bandwidth in the schedule (e.g. a train is late or early). Other types of deviations include a train using a different route than planned or operating a train with different characteristics. Deviations can be the result of the inattention or may also gradually accumulate. Otherwise, a deviation may also be the consequence of an incident or a disturbance. A deviation can be identified both when the deviation occurs or when a deviation is predicted.
Service Change: A service change consists of adding or changing trains in the existing schedule (e.g. changes in a service by adding supplementary stops or adding a new freight train). These types of changes may also impact other lines and services requiring a new schedule.

A disturbance can be distinguished from an incident or a deviation by the fact that after a disturbance, new plan conditions (e.g. changing train or infrastructure characteristics) must be determined and defined while in the case of an incident or deviation, in most cases only the temporal shifting of the schedule (new arriving and departing times) is required. Therefore, knowing the primary delay reason helps to speed-up the dispatching process including the decision on appropriate dispatching measures.

The primary delay reason category and their associated optimisation goals as well as temporal demands are summarised in Table 3.1.

Table 3.1: Optimisation criteria and time requirements for dispatching depending on delay reason

<table>
<thead>
<tr>
<th>Category of primary delay reason</th>
<th>Optimisation criterions</th>
<th>Temporal demands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviation or incident</td>
<td>Maximise traffic flow and capacity; and Ensure connections with connecting trains</td>
<td>High</td>
</tr>
<tr>
<td>Minor disturbance (reduced availability of vehicle or infrastructure)</td>
<td>Limit delay propagation to a geographical area or to a number of trains; and Minimise weighted total network delay; and Maintain circulation plan for staff and rolling stock</td>
<td>Medium</td>
</tr>
<tr>
<td>Major disturbance (infrastructure failure or defective vehicle) or service change</td>
<td>Ensure the flow of the transport chain; and Ensure that all stations are served; and Minimise the number of replacement trains and additional trains needed</td>
<td>Low</td>
</tr>
</tbody>
</table>

3.3.4 Overview of state-of-the-art advanced rail traffic management systems

3.3.4.1 Limitations of existing traffic management systems

Rapid improvements in information and communication technologies have made it possible to imagine the development of a real-time railway rescheduling support systems. In fact, back in the 1950s, automation and cybernetic ideas were studied for application to the field of railways and in particular for the rail traffic management [Hür06]. Nowadays, many modern rapid transit railways currently have some kind of semi-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 dispatching system for a mixed traffic railway network is very difficult [Gat04]. The following challenges have to be taken into account together when doing so:

- network complexity (routing possibilities);
- number of external influences and disturbances;
- lack of accurate train detection and control equipment; and
- high number of different train operating companies.

Research thus focuses mainly on two topics:

- developing fast algorithms to detect and solve conflicts in real-time; and
- developing efficient decision support systems for dispatchers to support them when controlling rail operations as a whole.

### 3.3.4.2 Overview of real-time conflict resolution algorithms

Research on optimising automatic conflict resolution has been carried out for several decades. The research on conflict resolution is thereby closely associated with the development of off-line scheduling for railways. At the beginning, mini-computer based information systems with limited online supporting possibilities such as optimum route planning were developed [Sau83]. Following this, research focused on algorithms for small territories with mainly single line sections and simplified running times. Examples are Kraft [Kra87] or Jovanovic and Harker [Jov91]. Algorithms for more complex networks (but still mainly single-track sections and low-frequency traffic) were developed later by Sahin [Sah99] or Adenso-Diaz et al. [AD99].

These models were not applicable to complex rail networks with high traffic densities as in Europe or Japan. Therefore, research focused on new approaches in order to take rerouting of trains into account. First, heuristic or so-called expert systems were developed and described by Komaya [Kom92] and Fay [Fay00].

In a following step, different mathematical methods were used for conflict resolution including rerouting of trains such as brunch and bound or genetic algorithms [Lam02, Rod07, Weg05b] for example. However, most models are based on fixed speed profiles. Combining the optimisation process with precise running time calculations, e.g. by the use of micro-simulation, allows the accuracy of the calculated conflict resolution to be improved [Weg07]. On the other hand, more precise models including precise running time calculations or blocking time theory, cause longer computation times.
Research on real-time conflict solving for complex rail networks is done by universities with the support of rail infrastructure operators mainly in Europe, e.g. in the Netherlands [D’A07], Sweden [Tör07], France [Gel06], England [Wes06] or Germany [Sch05] but also in the US [Des06], China [Chi02] and Japan [Hir06]. A good overview of different rescheduling approaches, algorithms and their models is described by D’Ariano [D’A08].

At the present time, rescheduling algorithms and methods are not available for large and dense mixed-use railway networks since they cannot handle all the following requirements simultaneously:

- sufficient level of model detail;
- mixed rail traffic;
- accurate prediction of future behaviour (especially after disturbances or events);
- complex network topology/layout (complexity and number of possibilities for train rerouting); and
- conflict resolution and rescheduling within short time.

Nevertheless, it is foreseeable that research progress will allow to use automatic conflict resolution for standard railways in future to support dispatchers in taking appropriate rescheduling measures.

### 3.3.4.3 Activities for a traffic management systems for controlling rail operation as a whole

The main focus of most research activities is directed at developing new and fast rescheduling algorithms. However, the challenge of advanced rail traffic management systems is to get a real-time application and to achieve a reduction in delays, to improve punctuality, and to utilise the existing infrastructure better. Thus, new interfaces designed to integrate all the relevant information and simplified communications are needed [San02]. The target will be to control rail operations as a whole. This would lead to a continuous process where plans are adjusted based on the actual state and transmitted to all involved persons.

Consequently, processes and communications must be handled as a closed-loop control system to control rail operations as a whole including individual train speed management. Fay [Fay99] described the entire process as a superimposition of cascaded feedback control loops. Thereby, offline and long-term strategic actions were also designed as control loops although they work more as functional chains. Schwanhaeusser [Sch97] and Jacobs [Jac04] designed the online scheduling and supervision process as a closed control loop on an abstract level (Figure 3.7). One weak point thereby is the missing definition for out-of-schedule running and the lack of
the analysis of the dynamic processes during operation. As part of the COMBINE projects [Gia04, Sav04], the structure for a new rail traffic management system was developed including a closed-loop rescheduling process for the ETCS Level 3 (Figure 3.8) [Maz07]. However, ETCS Level 3 with continuous position detection, braking curve supervision and update of movement authorities will not be available in existing mainline railway networks within the near future.

Figure 3.7: Online scheduling procedure (source [Jac04])

Figure 3.8: The ERTMS Level 3 operation control loop (source [Maz07])
Bottleneck areas are especially prone to delays. The main goal is that rail traffic flow should not unnecessarily be affected by stopping trains in these bottleneck areas. Recently, a dynamic traffic management approach was presented for the Schiphol tunnel area in the Netherlands [Sch07, Tvd05] for this purpose. The key feature is that even small deviations resulting in conflicts should be solved online. Train orders, in particular, are changed as long as they are not important for the subsequent routes. The concept disregards train connections in bottlenecks and enables the flexible use of either track at platform islands. Furthermore, the performance of the conflict resolution system is still not known. In addition, for integrated fixed-interval timetables with many connections in bottleneck areas as in Switzerland, it remains essential to plan and guarantee connections between different trains at the nodes.

3.3.4.4 Developments of driver assistance systems

The analysis of the traffic management system showed that interaction between trains and dispatching systems are essential to improve quality and performance. Developing driver assistance systems is thus of high importance. Train driver assistance systems have been developed earlier, but they focus on energy-optimum driving [Alb05, Fra02, Lin04, Net05]. Thereby, the transfer of online schedule information to change the planned trajectory has been neglected. Recently, several projects have been launched, to provide dynamic information on the schedule in the driver’s cab [Alb07, Fen05]. The idea is to use driver assistance systems in combination with conflict resolution algorithms to avoid unnecessary train stops at closed signals.

A comparable solution to the driver-machine interface was developed for the Lötschberg-Base-tunnel in Switzerland. Due to heavy use and a long single-line section in the newly built Lötschberg-Basetunnel (see [Hel04] for more details) new solutions were required for traffic flow optimisation. ETCS Level 2, which is installed in the tunnel, does not offer additional information (e.g. for optimisation functions) for a driver-traffic management and communication system. The lack of possibilities to transmit operational data in the cab by ETCS was also a reason why new onboard information display systems are being developed. In the new Lötschberg-Basetunnel, an automatic function (called AdmiRail, visualised in Figure 3.9) was developed by systransis. In case of a delay or event, new speed curves are computed (preventing possible conflicts) followed by the transmission of a speed advice as a free text message via radio interface to the corresponding train [Mon07]. Drivers may slow-down or speed-up to their given recommended speed. Fine tuning improves the entire conflict resolution process.

Instead of having a driver assistance system another opportunity is to replace drivers by an auto-pilot. For automatic driving, small delays occurring daily would be reduced nearly to zero if available buffer times between trains are exploited and technical failures could be kept at low level [Han01]. Also, in case of new directives, variations between drivers would be drastically reduced by auto-pilots, resulting in better predictability, fast reactions and small deviations. Automated driving is becoming more and more common for metro systems whereas only single, and
3.4 Challenges and requirements for new railway operation processes to improve the performance

3.4.1 General approaches to improve performance

The main objective for railway infrastructure companies is to improve the number of trains running on their network without reducing punctuality. The challenge lies in handling the increased demand for both passenger and freight trains. An integrated fixed-interval timetable with fixed patterns as applied in Switzerland therefore represents a substantial obstacle.
Generally, we can identify four main ways in which railways can increase their performance including capacity, stability and reliability:

- **Infrastructure** - building new infrastructure (tracks, junctions, flyovers, etc.), improving existing infrastructure, reducing the number of technical failures and shortening the time-to-repair;

- **Signalling** - reducing train headways by reducing block length or introducing more advanced signal systems (e.g. higher levels of ETCS [Eic07, Knu02, Win98]);

- **Rolling stock** - increasing traction power to harmonise the travelling speeds of trains (in particular for freight trains) and reducing the probability and duration of rolling stock failures; and

- **Operations** - reducing train headways by reducing buffer times implemented in the timetable to maintain reliability; harmonising travelling speed of trains; grouping of trains; or reducing variations in general during operations.

The infrastructure, signalling and rolling stock options are expensive and time consuming. Nevertheless, such enhancements are unavoidable in some cases. Reducing and homogenising block lengths to increase capacity is an efficient and effective investment. Reducing train headways was one of the central aspects of the Bahn 2000 project of SBB to improve capacity on their lines [Krä04]. Figure 3.10 illustrates how capacity could be gained by homogenising block distances. The headway for non-homogeneous and long block distances is large and thus capacity limited [Lüt07].

![Figure 3.10: Train headway reduction based on reduced and homogenised block length](image)

In addition to (expensive) investments in infrastructure or rolling stock, railway companies such as the SBB are looking for other ways to increase their performance. Most approaches result in reducing buffer times or improved schedule and operational robustness. As mentioned earlier, reducing buffer times by simultaneously increased stability is not possible with constant conditions. Given these problems, research and development on capacity improvement strategies, timetable robustness enhancements and new technologies are taking place.
3.4.2 Requirements for a new railway operation framework

Before the requirements for the new framework can be specified in detail, it is essential to identify the general goals. In particular, two approaches with different goals can be distinguished:

- reduction in energy consumption using cheap and simple tools; and
- improvement in capacity in dense areas.

An energy reduction with low investments means that mainly existing processes and simple tools must be developed and used. In addition, train headway buffers will either slightly or not be shortened in order to maintain the existing stability. Consequently, requirements on controllability, reactivity and achievable accuracy are limited in contrast to the case where capacity improvement is the focus of development. Higher requirements for a heavier use of the infrastructure results, in the end, in significantly needed enhancements and comparably higher costs.

This thesis focuses on an improvement in performance including significant capacity usage with new operation principles. Consequently, a new framework for an improved and more efficient usage of the rail infrastructure is required. Based on the drawbacks identified and the weakness of actual operation and planning principles, we identified the changes needed in various areas:

1. New scheduling proceedings and standards:
   - production plans must be more precisely (accuracy of seconds instead of minutes nowadays);
   - production plans must be conflict-free (using blocking time theory); and
   - the distribution and allocation of buffer times must take place in a different way in order to minimise them in capacity bottleneck areas.

2. More precise train operations:
   - trains must be able to follow dynamically changeable trajectories generated by the traffic management system; and
   - trains must be able to depart on time.

3. Fast rescheduling as part of the improved traffic management to:
   - minimise delays and their propagation through the network;
   - increase the accuracy and frequency of data and information; and
   - make data and information more specific and targeted to the receiver using it.
To improve infrastructure usage substantially, the blocking time of a block section in dense areas must be a multiple of the variations in train operations. Having headways of around 90 seconds in dense areas, a maximum variation of 15 to 30 seconds is required. The accuracy required must be followed permanently by all actors resulting in a high required reactivity. Therefore, the rescheduling must also satisfy high requirements. So, a maximum time lag of 30 seconds is allowed to generate a new schedule, and in parallel a large solution space as well as high precision must be satisfied.

The new approach will require precise and well-defined interfaces between the different actors, in particular between the tactical traffic management for optimisation and train control. Consequently, the train control or driving process, the train departure process as well as the optimisation procedure including rescheduling must be changed to meet requirements. The new system therefore has to combine the application of two related areas: the feedback control of each single train in the distributed network and the superimposed optimisation of new schedules in case of a deviation, disturbance or event.

### 3.5 The concept for an integrated real-time rescheduling system

#### 3.5.1 Fundamental structure of superimposed feedback control for railway systems

The present thesis focuses on railway operation processes. Based on the weak points and requirements identified, we developed a new framework including adjusted processes and methods. Substantial improvements are anticipated with the new concept. Available technologies and research developments seem promising ways of improving performance. The fundamental concept is that applying control theory methods for railway operations helps to reduce variations, allows fast and improved responses and thus results in a more stable production. Therefore, the new concept for the future rail operation process based on the superimposition of two general loops is introduced. The two closed control loops are:

- The inner feedback control loop on the lowest layer (driving the train, setting infrastructure elements) is responsible for precise production.

- The outer feedback control loop supervises the train traffic and infrastructure state, develops a new production plan based on predefined optimisation objectives which is subsequently transmitted in case of deviations and events.

Consequently, the new framework combines the generic cascaded control system with the hierarchical multi-layer control and optimisation division (introduced in section 3.2). Thereby, the train driver and the guards as actors in the inner control loop must ensure with the help of support tools that the given production plan (or train trajectory) is precisely followed. Also infrastruc-
tecture operation (provision of the infrastructure as defined in the production plan including on-time route setting) is part of the inner production feedback control loop and must satisfy temporal and quality demands as defined in the production plan. The inner control loop is also responsible for stabilising the process in all operating points and must attenuate small disturbances.

Based on optimisation objectives, new (or updated) schedules must be generated and transmitted to all the actors affected in real-time after a deviation or an event. What is important is that the entire process of the outer control loop (from starting with the input of detection of a deviation or event until new production plans are transmitted) is designed and handled as feedback control loops.

An aggregated illustration with the main processes and links of the new integrated real-time rescheduling (integrated real-time rescheduling) framework is visualised in Figure 3.11. In the following chapters, the newly integrated real-time rescheduling framework is introduced and specified. A detailed description of the outer rescheduling process is described in chapter 4, the inner production control loop is described in chapter 5, and the entire process and its superimposition effects including potential benefits are described in chapter 6.

3.5.2 Key concept of the integrated real-time rescheduling

A key concept of the integrated real-time rescheduling framework is that every train and the entire infrastructure always have a valid and accurate (seconds) production plan (e.g. including position, time and speed information for the driver as well as route and time information for infrastructure operators) which is (continuously) updated [Lüt09]. This is in contrast to today’s rail operation where train drivers only have the initial schedule with an accuracy of a minute which might not be conflict-free. An updated schedule after a deviation or event, which is trans-
mitted to all actors concerned, is not necessarily valid for the entire day but may be applicable over a given, limited future time horizon. Fast decisions and a new production plan (not only a new schedule) are generated as part of traffic management based on predefined optimisation rules and objectives whereas in today's production dispatchers take heuristic decisions and new schedules are not calculated.

With integrated real-time rescheduling, trains can be operated more precisely. In particular, it will be possible to follow a dynamically changeable schedule. Consequently, trains may be intentionally slowed down or speeded up before a conflict point or capacity bottleneck area is passed. A given reference point (e.g. at the boundary of a capacity bottleneck area) is thus passed at a predefined time slot and at a given speed. Consequently, time and capacity consuming braking and reacceleration in capacity bottleneck areas or at conflict points can be avoided. This effect with the new integrated real-time rescheduling framework is illustrated in Figure 3.12 to visualise exemplarily the advantages compared with conventional train operation. Avoiding unnecessary signal stops implicitly improves train traffic flow. Consequently the amount of knock-on delays and delay propagation can be reduced significantly resulting in enhanced punctuality.

![Figure 3.12: Train control with and without integrated real-time rescheduling](image-url)
3.6 Capacity improvement strategies based on an integrated real-time rescheduling system

3.6.1 Geographical network decomposition

Capacity bottlenecks require special attention for both planning and operation. Therefore, an appropriate network separation helps to manage such areas. In addition, decomposition is often unavoidable due to the large geographical distribution of the railway networks.

In general, railway networks can be separated and classified in several different ways depending on the purpose of the ultimate objective (e.g. planning, traffic management, or maintenance). Nowadays, network divisions are based on historic developments in many cases, which can no longer be an optimum for the particular purposes. Figure 3.13 illustrates four ways of dividing a railway network. These are as follows:

- The entire network is planned and operated as a single unit. This is only possible for small networks and is mainly applied to urban rapid transit systems.
- The network is divided into connected sub-networks. Each sub-network is responsible for itself and there is defined coordination between the different sub-networks. This is the classic method for planning and operating railways.
- The network is divided into nodes and routes. This type of decomposition is sometimes used during the planning process.
- The network is divided into capacity bottleneck areas (condensation zones) and areas with excess capacity (compensation zones). This approach, introduced by Laube and Schaffer [Lau03], is a new approach to divide the network and may be used for planning and operations. The division allows us to focus on critical bottleneck areas and to improve usage of these areas.

In terms of traffic management, the advantage of a large network is that complicated, multistage processes including coordination between different areas to generate a new production plan is not required. The disadvantage is that since it is a large network, developing schedules is a complex and long process. The area supervised by a single dispatcher is also limited. Consequently, larger networks always have to be separated to handle the traffic.

Developing a timetable for a divided network is easier in the sense that the problem is smaller, but it adds a need for coordination between the different areas. Therefore, a second working step occurs where train paths and slots must be coordinated with other sub-networks. Dividing the network into condensation and compensation zones is a special example of dividing the network into nodes and links.
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 are bottleneck areas and have no excess capacity (condensation zones). Therefore, the division into condensation and compensation areas helps to plan and operate bottleneck sections more systematically and with special care.

### 3.6.2 Integrated real-time rescheduling and the systematic management of bottleneck areas

The integrated real-time rescheduling concept allows bottleneck areas to be managed more efficiently. By dividing the network into condensation and compensation areas the bottleneck areas can be more systematically saturated (and thus to the maximum possible degree). However, in condensation zones it is crucial that trains are operated extremely precisely. Otherwise, initial delays may propagate to many trains within short time and finally throughout the entire network. In compensation zones, excess capacity provides trains with operational flexibility (i.e. speed control). 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 it is of utmost importance that both correct speed and time of the train passing the gateway is necessary to maximise capacity. Providing an exact departure time for a train from a station platform is another example.

The division into condensation and compensation zones facilitates operating capacity bottlenecks to the best possible degree and therefore guarantees that a network’s current weak spots are always at the focus of planning. Finally, reducing the number of unnecessary and time-consuming signal stops in bottleneck areas by applying the integrated real-time rescheduling
framework allows scheduled buffer times in these areas to be minimised. The combination of
the two concepts integrated real-time rescheduling and smart network division in compensation
and condensation areas, allows stability and capacity to be improved coincidentally.

Decentralising or dividing a network as for example into compensation and condensation areas
results in further advantages:

- computation is easier because less information is needed and manipulated;
- parallel computation is possible without the need for complete synchronisation; and
- improved robustness is achievable because failures and problems are easier to separate
  and segregate.

Nevertheless, smart decomposition designed to achieve the maximum autonomy of sub-problems
is difficult to attain. Avoiding strong coordination is therefore not always possible and is an issue
that must be solved.

3.6.3 Realisation challenges for integrated real-time rescheduling systems

Based on the concept for the integrated real-time rescheduling framework and the network de-
composition strategy, three main challenges for realisation were identified, which have to be
developed and implemented for the successful realisation of the integrated real-time reschedul-
ing framework:

- fast control loops including information transfer, data preparation, rescheduling and its
  application;
- high precision for data, production plan and its realisation; and
- clever system (network) decomposition and coordination.

Beside technical challenges, management, acceptance and migration problems also occur. The
present thesis focuses on changes in the production process including the actors and technol-
ogy and its effects on operations and their efficiency impacts.
3.7 Conclusions

Existing rail operation principles have been described in this chapter, highlighting the weak points of conventional railway operation principles. To improve performance, an integrated real-time rescheduling system for railways must satisfy two main tasks:

- deviations from the schedule or events must be identified, solved and transmitted back as a new schedule in real-time to all actors involved; and
- production (running trains) is executed precisely following a dynamically changeable schedule within a defined tolerance with the help of supporting tools.

To meet requirements, feedback control principles must be adapted for railway operations. This chapter has shown the fundamental basics as well as the adjustments and enhancements needed to form given railway operations as superimposed feedback control loops in an integrated real-time rescheduling framework. This concept of superimposed feedback control loops for railway operations reduces variations, responses within short time on deviations and allows a higher train traffic flow in comparison to conventional railway operation principles. The benefits and implementations are demonstrated in chapters 6 and 7. Improved train traffic flow causes delays and their propagation to be significantly reduced. Also, mention is made of missing elements to be developed for the integrated real-time rescheduling framework and connected main challenges. Finally, this chapter explains how smart network decomposition in combination with the integrated real-time rescheduling framework can be used to minimise the amount of buffer times required and consequently, to maximise capacity in selected bottleneck areas.

In the following chapters, the superimposed feedback control loops as well as their sub-processes, elements and impacts are modelled and described in detail. Requirements for an efficient interaction of the control loops are specified in order to achieve maximum performance for the entire system. Finally, the potential benefits of the integrated real-time rescheduling framework will be derived and illustrated with simulations and case studies.
Chapter 4

The real-time rescheduling loop

4.1 The rescheduling loop as part of the integrated real-time rescheduling framework

The focus of this chapter is on the outer loop of the integrated real-time rescheduling framework with the rescheduling procedure and all sub-processes involved. All elements of this loop will be described in detail. The impacts of the elements on rescheduling duration and performance will be determined. In addition, a reasonable boundary for the entire rescheduling duration, derived from the requirements in chapter 3.4.2, is also determined.

The rescheduling loop, generically visualised in Figure 4.1 and investigated in this chapter, includes the following actors and main procedures:

• traffic management with the supervision of all train movements as well as the generation and transmission of new production plans;

• an infrastructure that is prepared and supplied to trains on scheduled time ensuring the safety of all train movements; and

• drivers, guards and passengers using or operating the train on the infrastructure.

Figure 4.1: General model of the outer rescheduling loop
The outer rescheduling loop is the tactical layer within the integrated real-time rescheduling rail operation framework. Traffic management, as the controller responsible in this outer feedback loop, includes the following tasks:

- the detection of deviations and conflicts;
- the solution of conflicts by generating new production plans based on predefined optimisation objectives; and
- the transmission of new production plans to all actors involved.

For a high performance of the system, process and transmission lags must be minimised. Similarly, new production plans must be developed and finally transmitted to the actors within the shortest possible time. However, technical and operational constraints prolong the duration of the rescheduling process.

Based on the shortcomings of the actual dispatching process, development and improvements must focus on four scopes:

- Fast transmission of position and state reports from the train to the traffic management system.
- Automated detection of a deviation or event, including the identification of the primary delay reason and a prediction of future behaviour.
- Generation of new production plans in real-time. A network-wide optimisation based on the specified objectives of the rail traffic is the goal.
- Transmission of new production plans directly to the actors (e.g. providing driving recommendations).

The elements and processes of the outer integrated real-time rescheduling feedback loop addressing the main scopes introduced are described and evaluated in detail below.

### 4.2 Relevant processes of the real-time rescheduling loop

Rescheduling as part of the integrated real-time rescheduling framework is a complex multi-stage process. Figure 4.2 illustrates the entire process including subprocesses on a conceptual level. These processes are introduced in brief and described below.

The rescheduling process is based on information from all actors regarding the infrastructure and train conditions. Various communication channels (such as phone, fax, or electronic data) are used to transfer the information. The data could be automatically generated (as
changes of switch position or route occupation), or manually triggered (coded or oral information). Nowadays, the automated transmission of continuous position information from a train to the rescheduling system is only implemented for ETCS Level 2. Also, status reports are not automatically transmitted (except with limitations in ETCS). Only radio or phone for verbal exchange between driver, guard, infrastructure operator and dispatcher are available.
Information from actors, trains and infrastructure is collected, prepared and processed by the rescheduling system. The information is used to:

- detect deviations;
- identify the primary delay reason; and
- provide a database for the precise and well-founded prediction of the future behaviour.

A decomposition of the network (e.g. into several areas) may be carried out to handle the immense amount of data and also to simplify and speed-up the next working steps.

The data collected is merged and used for prediction and as constraints, which are part of the input of the algorithms generating a new production plan. The process of generating a new production plan is itself a multi-stage process. Depending on the decomposition of the network and possible knock-on consequences, an exchange between the decomposed network areas may be required.

Finally, the feasibility of the new production plan must be checked before it is transmitted to all affected actors. These recommendations are used as input into their supporting system for an accurate production.

### 4.3 Elements of the real-time rescheduling loop

#### 4.3.1 Train detection

##### 4.3.1.1 The tolerance band principle

One important requirement for the rescheduling system is the early and accurate detection of a deviation or event. To initiate the rescheduling loop, a specific, consistent rule must be defined instead of heuristic, random reactions. Therefore, we define a tolerance band for each train and rescheduling is initiated if the bandwidth is exceeded by the train. The basis for determining if the threshold has been exceeded is thus the precondition that each train (each actor) receives a timetable (production plan) and a bandwidth within which it must operate. The principle of the bandwidth around the scheduled train path is visualised in Figure 4.3.

Using the tolerance band, deviations can be detected before a conflict occurs. This allows dispatchers to react at an early stage. In addition, the tolerance band gives drivers the possibility to react to small deviations, unusual conditions (e.g. adhesion, train weight, traction variations) and provides them with sufficient scope to drive the train. For sections with a planned series of trains, the tolerance band may be used to limit the freedom of drivers in order to minimise knock-on delays.
The decisive variable for the performance of the infrastructure on the one hand and of the stability on the other hand is the size of the tolerance bandwidth. The tolerance bandwidth, assigned for each train and used as the minimum buffer between two trains, reduces capacity and should therefore be as small as possible. On the other hand, a small tolerance band often causes the threshold to be exceeded and frequent rescheduling, which should be avoided. Consequently, the tolerance bandwidth should be defined in a way that trains remain within the limits for regular operations in most cases. However, the tolerance bandwidth may be constant or variable, depending on route, train type, time of day and function. The target size of the tolerance bandwidth is based on capacity and stability requirements and should be around 15 seconds, as requirement introduced in chapter 3.4.2.

4.3.1.2 Train detection methods

4.3.1.2.1 Overview of train detection methods

The most common reason for triggering the rescheduling process is a time deviation, in other words a train that is either late or early exceeding the tolerance band. To identify a deviation, the train’s position and reference production plan must be known by the rescheduling system. The detection of train locations and, in particular, their frequency is of utmost importance for the rescheduling system. In the event of a deviation from the schedule, the detection density defines the time lag until a deviation is detected.

In addition, a high recording density may also help to quickly determine the primary delay reason and may also be used for the more precise and more stable prediction of future behaviour. Three methods are described below of how a time deviation can be identified.

- Infrastructure Train Location Detection - permanently (but irregularly distributed within a network) installed infrastructure elements transmit the train number when a train passes a detection point through the electronic network to the operator and dispatcher.
• Periodic Train Location Transmission - trains automatically transmit their location (and possibly additional information) on a regular, frequent basis (with respect to time) to the dispatcher or rescheduling system (e.g. using GSM-R or radio).

• Participant Transmission - participants (drivers, guards, infrastructure operators) inform the dispatcher or network operator directly using oral communication.

These three methods are explained in detail below.

4.3.1.2.2 Infrastructure based train detection

Today, almost the entire railway network is equipped with various train detection devices (e.g. axle counters, track circuits, balises) designed to operate safety equipment and transmit movement authorities. These devices identify trains (actually the heads or ends of trains) at discrete, irregularly arranged points on the infrastructure network (on SBB’s network, the distance between two detection points is in most cases between 300 and 5,000 metres). After a short delay for data processing, the railway control centre, the dispatcher and the rescheduling system are provided with the information by the 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 a time window (derived from the tolerance bandwidth) for each train at the particular reference point. Figure 4.4 illustrates the time windows for a train trajectory at two reference points.

A train arriving early will trigger a data event in the control centre. In contrast to this, if the train is late (i.e. does not pass the reference point before the latest defined passage time), either the infrastructure operator (manually) or the rescheduling system (automatically) will initiate an event which activates the rescheduling process.

The time lag until an excess of the threshold is detected depends on the train’s position and speed as well as the distance to the next detection point. The maximum lag occurs, when a
train exceeds the threshold directly after passing a train detection point. Thus, the maximum time lag $t_{\text{lag, max}}$ is determined by the distance to the next detection point $s_{\text{next, detect}}$ and the planned track speed $v_{\text{planned}}$:

$$t_{\text{lag, max}} = \frac{s_{\text{next, detect}}}{v_{\text{planned}}}$$

(4.1)

For a sample speed of 120 km/h and a block distance of 1,500m, the time lag until the exceeding of a threshold is detected, is at most 45 seconds.

There are several disadvantages with fixed infrastructure transmission:

- One disadvantage is that in the event of a delay no information about the exact location and state of the train is available: Is the train only one second late because of driver inattention or will the train be late for ten-minutes because of a temporary breakdown? Only information about the occupied blocks is available.

- A second disadvantage is that there may be a significant loss of time until a deviation is detected especially in the case of long distances between infrastructure detection elements. This problem could be addressed by installing additional passive detection balises to create a denser network of registration points.

- And finally, infrastructure based detection allows us to measure the exact point in time when a train passes the given detection point, but no additional information (e.g. train speed).

Nevertheless, infrastructure transmission is an inexpensive approach, because it is already in operation. For denser detection, it may be upgraded with additional detection points (e.g. balises) which do not require more enhancements or adjustments in the system. Another advantage of the system is that the time delay when a train passes a detection point until the information is available is very small and train position data is very precise.

### 4.3.1.2.3 Periodic train position transmission

In the periodic train location detection method, the train automatically transmits (e.g. by GSM-R) location and status information (e.g. actual speed) to the control and traffic management center on a regular basis (e.g. every 30-seconds). The highest frequency is thereby determined by the maximum possible amount of data transmitted between trains and the traffic management system as well as the number of trains in a traffic management center.

The basis for a periodic data transfer system like this is the precise determination of the position. For that purpose, each train needs either a satellite-based locating tool (GPS or Galileo) or onboard odometer equipment or, best of all, a combination of both. The accuracy of both systems can be improved with the aid of infrastructure installations (e.g. balises) to reduce distance.
measurement errors (validation and calibration at reference point). Commercial GPS tools with the standard positioning service have a domain accuracy of 13 meters (36 meters worst case) for horizontal applications. Differential GPS improves precision in a way that accuracy to one meter is possible [Det06, HW08]. Nevertheless, greater inaccuracy of the position determination is possible, which results, depending on the train’s speed, in greater errors in temporal deviation. This effect, illustrated in Figure 4.5 also points out that especially for slower speeds, larger temporal errors are possible. This effect, together with the large number of tunnels within Switzerland enforces the need of to combine odometer and satellite based location systems to get a precise and robust positioning system.

![Figure 4.5: Resulting temporal deviation due to inaccuracy of positioning system](image)

For periodic train position transmission, threshold excesses can be detected and evaluated directly by comparing the actual position to a given distance window for each triggered point of time saved in a list. Consequently, no additional list needs to be set-up to detect late trains. Figure 4.6 illustrates position windows (i.e. the trains should be somewhere between these points at the specified time) for trains using periodic position transmission.

![Figure 4.6: Periodic train location detection](image)

The list (defining lower and upper position bound for each train) is easier to handle than in the infrastructure detection method since the data being evaluated is always the same and a late
train is detected in exactly the same way as an early train. Nevertheless, the temporal bandwidth must be translated into a distance bandwidth for each point of time.

A significant advantage of this system in comparison to the infrastructure detection method is that future behaviour for delayed trains can be predicted more precisely because additional information (e.g. actual speed, door positions during the boarding and alighting process, or coded information for events) may be available from the train. And for high sampling rates (less than 10 seconds), almost continuous observation of behaviour is possible and can be used for a prediction with the highest possible accuracy.

4.3.1.2.4 Participant transmission

A third possibility for detecting and transmitting information when thresholds are exceeded is the communication between the actors and the rescheduling system or dispatcher. Oral communication or the transmission of coded messages are possible.

For oral transmission, the main problem is the time delay: at least 30 seconds are required, usually even more until the data and information are explained and processed. Therefore, oral communication may only be recommended for applications that are not time-critical (e.g. during the preparation process or at intermediate stops) or for gathering additional data on an event. However, direct communication between dispatcher and train driver are still used frequently at the SBB. Coded messages (for example, as part of a driver-machine interface) may be an efficient way, in which event or deviation (delay) reasons can be transmitted within shortest time.

However, the transmission of additional information requires special attention on the part of the actors. If there is a deviation or an event, drivers and guards are already distracted, and their full attention is required to solve the problem or to handle the train (e.g. for bad adhesion conditions). Therefore, one simplification may be that creeping deviations (e.g. due to bad adhesion) need not be transmitted and only disturbances and incident reasons have to be transmitted by drivers and guards.

4.3.1.2.5 Comparison of train detection methods

Participant transmission is not appropriate for reporting delays in general and should be applied only in specific circumstances. Therefore, only the two methods of infrastructure-based and periodic transmission are compared. A fundamental difference between the two methods concerns train supervision. A time-window is used for supervision using the infrastructure-based train detection method. In contrast to this, the train-based periodic transmission uses a position window which is supervised by the rescheduling system. To point out the differences, the data flow for infrastructure-based train detection and periodic-position transmission is visualised in Figure 4.7.
An overview of the two train detection methods is provided in Table 4.1. It shows that the density of detection points and the sample frequency are the two decisive parameters for performance. The main advantage of periodic transmission is the possibility of frequent data transmission even in the event of a breakdown and the option of transmitting additional information allowing the faster identification of a delay and the more accurate prediction of future behaviour. Thereby, the combination of available odometer positioning (including a framework for calibration when an infrastructure detection point is passed) with satellite-based positioning may make the system almost as accurate as exclusive infrastructure-based train detection. However, a data-uplink from the train to the rescheduling system leads to a more complex and consequently more expensive tool and system architecture.

The periodic transmission method may be extended by sending event messages from a train to the rescheduling system between two regular samples. This would result in a hybrid-state transmission. In particular, in the event of exceeding a threshold, a message could be sent by the train automatically. This would require an onboard tool, analysing the temporal deviation of the train. Also, the tolerance bandwidth has to be known by the onboard tool for the entire route. Using a hybrid method, temporal lags may be minimised in the train detection process.
### Table 4.1: Overview of characteristics for train detection systems

<table>
<thead>
<tr>
<th></th>
<th>Infrastructure based train detection</th>
<th>Periodic train position transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precision</strong></td>
<td>Highest accuracy</td>
<td>Accuracy of 15 meters in general</td>
</tr>
<tr>
<td><strong>Regularity</strong></td>
<td>Sporadic data trigger when train passes detection point</td>
<td>Periodical data trigger</td>
</tr>
<tr>
<td><strong>Recording density</strong></td>
<td>Depends on the distance between two detection points and the track speed</td>
<td>Defined by the sampling rate</td>
</tr>
<tr>
<td><strong>Time delay to detect exceeded threshold</strong></td>
<td>Time taken before the next detection point is passed (early trains) or has to be passed at the latest (delayed train): between 15 seconds and 3 minutes on average for a network like SBB</td>
<td>At next data trigger: sampling rates of around 15 seconds are the target</td>
</tr>
<tr>
<td><strong>Possibility to transmit additional data</strong></td>
<td>Not possible</td>
<td>Automated communication offers the opportunity to send additional information (e.g. speed, door state or coded event information)</td>
</tr>
<tr>
<td><strong>Developments and new systems</strong></td>
<td>Passive balises for more dense train detection</td>
<td>Up-link to transmit train position (and state) by GSM-R or other communication channels</td>
</tr>
</tbody>
</table>

#### 4.3.2 Infrastructure supervision

Similar to the supervision of the train position and state, events and deviating states of the infrastructure (switch positions, signal states, barrier position etc.) must also be monitored and collected. The occupation state of a route is nowadays already available in the traffic management centre. Incorrect infrastructure states or reduced infrastructure availability are usually communicated verbally. Infrastructure information is a required input for the rescheduling system. Manual input of data should therefore be avoided due to excessive time requirements and the likelihood of errors.

Deviating and abnormal infrastructure states should, whenever possible, be automatically (electronically) transferred to the rescheduling system. If this is not possible, operators in the control centre can send coded messages to speed-up the rescheduling process. The information should contain the kind and place of deviation as well as a predicted duration until full availability is re-obtained. In this way, time-consuming verbal communication can be limited to clarify imprecise, vague or missing information. Especially in cases of vague knowledge after an event, updated information must be transmitted to the rescheduling system regularly in order to assure that the behaviour is as predicted or if changes are required.
4.3.3 Data processing and preparation

4.3.3.1 Primary delay identification

4.3.3.1.1 Advanced information for quick primary delay identification

The immense volume of incoming data is used by the rescheduling system to identify the primary delay reason. This is of the utmost importance, because the identification of the primary delay reason is the fundamental basis for predicting the future behaviour of the trains and infrastructure. Thereby, two requirements may be contradictory: on the one hand the primary delay reason should be identified as fast as possible, and on the other hand also faultlessly and precisely. For better prediction, often more time (and data) is needed for the unique identification of the delay reason. A multi-level rescheduling approach, introduced in chapter 3.3.2.2, thus seems helpful and most promising. During an initial pre-dispatching phase, measures based on vague data are used to generate new schedules. In parallel, time is used to gather more information on the primary delay reason, which is used subsequently for more precise prediction and rescheduling.

Incidents and disturbances often occur before they cause a temporal deviation. Information sent to the rescheduling system at an early stage can therefore be used to adjust the production plan before the consequences are observable. Examples are infrastructure failures observed by movement inspectors, maintenance staff or others before a train passes the section. In addition, late connections (not observed by the rescheduling system, for example, bus services or international trains) or failures to trains observed before departure may typically be reported before a consecutive delay occurs.

4.3.3.1.2 Primary delay reasons

Identifying the primary delay reason is more difficult if no advanced warning or information is available and a train exceeds the tolerance band. The identification of a delay is related to accident or failure investigations where a multiplicity of formal methods exist. Examples are the why-because analysis, fault tree analysis, the root cause analysis or multilinear event sequencing [Ben75,Fah00,Mon99,NM96]. What these approaches have in common is that they are designed and carried out manually and a long time after the event. In contrast to accidents, deviations from the tolerance band may be assigned in the majority of cases to one single delay reason. In addition, the number of delay reasons is also limited.

For trains ahead of schedule, the reason may be ascribed in most cases to wrong driving behaviour or an erroneous production plan whereas a late train can have several possible delay reasons. However, the problems are that different events may have similar effects and that one failure may have varying consequences. Especially for trains not departing from a station, an
identification of the reason is impossible without additional information. Of course, if the primary delay reason is not identified within reasonable times, assumptions based on earlier data (statistics) may be useful. These assumptions can change (which causes a restart of the entire rescheduling process) when additional data is made available and causes differences in the prediction of future behaviour.

The main target during the identification phase is to determine if the event is either a disturbance (which ultimately results in changed conditions) or an incident. For an incident, the behaviour until fully availability is regained and the consequential delay must be predicted. If the deviation is caused by a disturbance, the new parameters for the infrastructure, train or staff must be determined and used for prediction.

For the simplified handling of event data during the rescheduling process, we structured the event types. This helps to speed up the prediction and timetable generation process. Table 4.2 gives an overview of the majority of all known event types.

4.3.3.1.3 Challenges for identification of primary delays

The identification of the primary delay reason is of special difficulty for infrastructure-based train detection. A missing train at a reference point means that the position and state of the train is totally unknown for the rescheduling system and in addition, it is also possible that data transfer from the infrastructure to the rescheduling system has been interrupted. Therefore, the rescheduling system must wait until the driver notifies the dispatcher or the train runs over a next reference point. But even after passing the next reference point, identification of the delay reason is almost impossible (e.g. did the train slow down and reaccelerate due to an incident or has the train got reduced availability and is thus running more slowly?). Consequently, data from multiple reference points (which may take several minutes to pass) are required. In contrast to infrastructure-based train detection, periodic data transmission has two main advantages. Firstly, it ensures that interruptions between train and traffic management centre are identified and, secondly, the train’s behaviour and position are available for the rescheduling system.

If an incident causes a delay, the reason often remains unknown for the rescheduling system if no additional information is provided. However, it is crucial to identify disturbances causing consequences such as reduced train dynamics, bad adhesion conditions or reduced availability of the infrastructure. This could be achieved if the data of the train delayed is compared with its earlier run and running times of other trains on the same track section. Bad adhesion conditions are likely if other trains have also lost time on the same track section. Earlier delays in the train run indicate that reduced train dynamics are most likely. Similarly, overlong dwell times can be explained by unscheduled high demand and may also be avoided if longer boarding and alighting durations are scheduled for the next few stations. Of course, the superimposition of different events (for example bad adhesion and reduced train dynamics) is also possible and even more difficult to identify.
### Table 4.2: Most of all known event types causing delayed deviations

<table>
<thead>
<tr>
<th>Event cluster</th>
<th>Specific delay reason</th>
<th>Driving mode</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Train disturbance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train failure</td>
<td>Running and departing train</td>
<td></td>
</tr>
<tr>
<td>Changed (reduced) train dynamics</td>
<td>Running and departing train</td>
<td></td>
</tr>
<tr>
<td>Missing train</td>
<td>Departing train</td>
<td></td>
</tr>
<tr>
<td><strong>Train incident</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short time train malfunction</td>
<td>Running and departing train</td>
<td></td>
</tr>
<tr>
<td>Distractions caused by passengers (emergency brake, blocking doors)</td>
<td>Running and departing train</td>
<td></td>
</tr>
<tr>
<td>Operating error (staff)</td>
<td>Running and departing train</td>
<td></td>
</tr>
<tr>
<td><strong>Infrastructure disturbance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduced or broken infrastructure</td>
<td>Running and departing train</td>
<td></td>
</tr>
<tr>
<td>Changed (reduced) adhesion</td>
<td>Running train</td>
<td></td>
</tr>
<tr>
<td><strong>Infrastructure incident</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wrong route</td>
<td>Running train</td>
<td></td>
</tr>
<tr>
<td>Obstacle or people close to or on track</td>
<td>Running and departing train</td>
<td></td>
</tr>
<tr>
<td>Unscheduled signal influence</td>
<td>Running and departing train</td>
<td></td>
</tr>
<tr>
<td><strong>Staff disturbance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staff missing</td>
<td>Departing train</td>
<td></td>
</tr>
<tr>
<td>Unfeasible schedule</td>
<td>Running and departing train</td>
<td></td>
</tr>
<tr>
<td><strong>Schedule disturbance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scheduled conditions differ from actual conditions (passenger demand, wrong train dynamics)</td>
<td>Running and departing train</td>
<td></td>
</tr>
<tr>
<td>Late connections by unrecorded transport</td>
<td>Departing train</td>
<td></td>
</tr>
<tr>
<td><strong>Creeping deviation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unobservant, careless driving behaviour</td>
<td>Running and departing train</td>
<td></td>
</tr>
<tr>
<td><strong>Other incidents</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delayed custom clearance, cleaning or others</td>
<td>Departing train</td>
<td></td>
</tr>
<tr>
<td><strong>Transmission errors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interrupted, restricted or faulty transmission</td>
<td>Running and departing train</td>
<td></td>
</tr>
<tr>
<td><strong>Data errors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erroneous train (positioning, speed) or infrastructure data</td>
<td>Running and departing train</td>
<td></td>
</tr>
</tbody>
</table>

Handling transmission and data errors requires special attention (if the data is plausible) and are not evaluated in this thesis. Nevertheless, these errors are possible and must be taken into consideration if primary delays are to be identified.

In summary, with today’s distances between the detection points train detection based on infrastructure allows no unique assignment to a primary delay reason within reasonable time. The same is true for periodic train detection with large sampling times (e.g. over one minute). However, even higher sampling rates will not ensure that primary delay reasons can be identified uniquely and within the shortest possible time.
Consequently, the up-link from the actors (trains) to the rescheduling system should be used to send additional messages. According to the different delay reasons in Table 4.2 train drivers may send a coded message (number) whenever a threshold is passed to inform the rescheduling system. It is even conceivable that the rescheduling system in such a case actively demands further information. It must then be ensured that sending a message takes place rapidly and that drivers are not distracted from their regular work. Of course, the message should be checked and compared with recorded data for plausibility. Using coded messages by train drivers, the primary delay reason is determined in the shortest possible time and is also as correct as possible.

4.3.3.2 Process of predicting future behaviour

Based on the primary delay reason and type identified, assumptions for future condition, behaviour and their period of validity may be established. In fact, during the rescheduling proceedings, two states must be predicted (Figure 4.8 illustrates the two prediction parts in a time-distance diagram):

- Firstly, the behaviour of the train during the whole rescheduling process is predicted. This will be checked for feasibility at the end of the rescheduling procedure.
- The second part consists of predicting ‘long-term’ future behaviour based on the delay reason identified. Accurate prediction is possible in the event of detailed information.

Data mining or adaptive principles may also be used for more precise prediction, especially in the event of deviations (e.g. due to bad adhesion, new train dynamics, increased passenger demand resulting in longer boarding and alighting). Based on previous data (e.g. from earlier
days, former train behaviour or other trains), the parameters for the future behaviour can also be adjusted.

More accurate prediction finally leads to a more stable rescheduling system. The entire process of predicting future behaviour should be fully automated and based on the primary delay reasons identified. As mentioned, the accuracy of the prediction depends on the precision of the database and the information. However, even precise knowledge of the problem (e.g. catenary breakage) can cause imprecise predictions (e.g. the duration until a breakdown is repaired). The correlation between the primary delay reason and the way the future may be predicted is illustrated in Figure 4.9.

![Figure 4.9: Procedure of predicting future behaviour](image)

Finally, the prediction is used to determine the earliest (and also latest) possible point of time when the next reference points may be reached by the trains. This information is used as input for the rescheduling system to generate a new production plan.

### 4.3.4 Generation of new production plans

#### 4.3.4.1 Rescheduling performance measures

##### 4.3.4.1.1 Delayed trains rescheduling optimisation objectives

The key optimisation criteria for dispatching depends on the possible consequences as well as the number of actors involved (see also chapter 3.3.2.4). To measure the quality, efficiency and performance of a rescheduling system, an optimisation objective must be defined. Subsequently, several possible optimisation functions are introduced.
4.3. Elements of the real-time rescheduling loop

A first possibility is to sum up the delays \( t_{d_{i,j}} = t_{\text{measured}_{i,j}} - t_{\text{planned}_{i,j}} \) of all trains \( i \) arriving or passing given reference stations \( j \). The objective function \( f_{\text{obj}} \), which will be minimised, would then be:

\[
f_{\text{obj}} = \min \sum_{j=1}^{n} \left( \sum_{i=1}^{m} t_{d_{i,j}} \right)
\]

(4.2)

These simple objective functions (e.g. used by [Gel06, Lam02, Rod08b]), may be modified and extended by weighting factors for each train and station \( a_{i,j} \) or penalty functions depending on a train’s delay \( f_d \). The objective functions can therefore be extended and expressed as:

\[
f_{\text{obj}} = \min \sum_{j=1}^{n} \left( \sum_{i=1}^{m} a_{i,j} f_d(t_{d_{i,j}}) \right)
\]

(4.3)

Possible penalty functions (illustrated in Figure 4.10) can be expressed as:

\[
f_{d1}(t_{d_{i,j}}) = t_{d_{i,j}}
\]

\[
f_{d2}(t_{d_{i,j}}) = \begin{cases} 
0 & \text{for } t_{d_{i,j}} < t_{t1} \\
 a_{k1} & \text{for } t_{t1} \leq t_{d_{i,j}} < t_{t2} \\
 a_{k2} & \text{for } t_{t2} \leq t_{d_{i,j}} 
\end{cases}
\]

\[
f_{d3}(t_{d_{i,j}}) = \begin{cases} 
0 & \text{for } t_{d_{i,j}} < t_{t1} \\
 a_{k1} + a_{k2}\log(t_{d_{i,j}}) & \text{for } t_{t1} \leq t_{d_{i,j}} 
\end{cases}
\]

\[
f_{d4}(t_{d_{i,j}}) = \begin{cases} 
0 & \text{for } t_{d_{i,j}} < t_{t2} \\
 a_{k1}t_{d_{i,j}} & \text{for } t_{t2} \leq t_{d_{i,j}} < t_{t3} \\
 a_{k2} & \text{for } t_{t3} \leq t_{d_{i,j}} 
\end{cases}
\]

with the times \( t_{t1} \) (a minimum acceptable delay regarded as on-time, e.g. two minutes), \( t_{t2} \) (a delay with which planned connections can still be guaranteed, e.g. five minutes) and \( t_{t3} \) (a larger delay requiring that major measures like the operation of an additional train must be taken, e.g. 30 or 60 minutes) and weighting parameters \( a_{k1} \) and \( a_{k2} \).
Chapter 4. The real-time rescheduling loop

Figure 4.10: Visualisation of exemplary penalty-functions for delayed trains

So far, all delays have been taken into account. Another possibility would be to only consider and sum up the knock-on delays (e.g. used by [Mar95]). An optimisation objective, which does not minimise the sum but the maximum knock-on delay, is proposed by [D’A07]. The optimisation function in this case can be expressed as:

\[ f_{obj} = \min \| (t_{d_{i,j}} - t_{p_i}) \|_{\infty} \]  

(4.4)

with the train’s primary delay \( t_{p_i} \).

### 4.3.4.1.2 Connections oriented rescheduling optimisation objectives

In highly integrated networks, as is the case in Switzerland, breaking connections are unpopular measures and may cause passengers to wait for a longer time. Thus, another important objective is to maintain as many (important) connections as possible in the event of a delay or event. Connections can thus be weighted, for example, by the number of passengers transferring between two trains. Taking all the planned connections \( l \) into account, the objective function is:

\[ f_{obj} = \min \sum_{j=1}^{n} \left( \sum_{l=1}^{p} b_{l,j} s_{l,j} \right) \]  

(4.5)

with \( s_{l,j} = \{ \begin{array}{ll} 0 & \text{for connection } l \text{ hold} \\ 1 & \text{for connection } l \text{ broken} \end{array} \) \)

and with the planned connections specific weighting factor \( b_{l,j} \) which could, for example, depend on the number of passengers with a planned connection \( l \) in station \( j \).
4.3.4.1.3 Secondary rescheduling optimisation objectives

Further objectives can be the total number of trains assigned with a knock-on delay, the number of required track or platform changes (applied by [Weg05a]) or the total number of unscheduled influences by a closed signal. From an operational point of view, the last criteria may be of special interest because energy consumption and thus costs are directly associated with the number of trains influenced by a signal. Consequently, the number of unavoidable signal influences $u$ to a train $i$ have to be minimised and are described in the objective function as

$$f_{obj} = \min \sum_{i=1}^{m} c_i u_i$$

with the specific weighting factor $c_i$ with which the train’s specific parameters can be considered.

4.3.4.1.4 Recommended rescheduling optimisation objectives

Depending on the network and the timetable strategy applied, the objectives and parameters are of different importance. Consequently, for rail networks and service principles as in Switzerland, we propose a composed optimisation objective with the focus on

1. maintaining connections (high weighting parameters);
2. minimising knock-on delay; and
3. avoiding signal stops

resulting in the following function:

$$f_{obj} = \min \left[ \sum_{j=1}^{n} \left( \sum_{i=1}^{m} a_{i,j} f_d(t_{d_{i,j}}) + \sum_{l=1}^{p} b_{l,j} s_{l,j} \right) + \sum_{i=1}^{m} c_i u_i \right]$$

(4.7)

4.3.4.1.5 Rescheduling quality criterions

The effects or optimisation targets, which could be compared and used as a performance measure, were introduced in the previous section. However, the rescheduling process as a whole must be considered. Hence, two further criteria need to be used to determine the performance of a rescheduling process.
A initial quality criterion is the rescheduling stability or total number of initiated reschedulings over a given time period. This is an important characteristic and also decisive if it is accepted by the actors (too many reschedulings in a short time are not appreciated by actors).

A second quality criterion is the rescheduling duration: the time period from the point a threshold is exceeded until a new, feasible production plan is generated and transmitted to all the actors involved.

### 4.3.4.2 Rescheduling input

A detailed and accurately updated database is the basis for rescheduling. The inputs required for the task of generating a new production plan are:

- the current infrastructure, train status and actual behaviour;
- the predicted behaviour (two-level process) based on the primary delay identification; and
- the initial production plan (as a basic reference) combined with the optimisation objectives.

Data may be classified as offline and online data. The offline data consists of the initial production plan and the regular, planned infrastructure and rolling stock data. Infrastructure data consists of an exact description of the topology and the associated signalling system (e.g. including route reservation times). Rolling stock characteristics include specific data for the planned conditions. Changes from the originally planned conditions because of a disturbance are online data. A characteristic of online data is that it can be assigned with a limited validity (e.g. reduced speed on a given track section until a certain point of time during the day). The actual production plan at that time is also part of online data.

Working with the enormous amount of data requires attention to be paid to the design of the system architecture:

- Storage of redundant data should be avoided. This requires a stringent separation of online and offline databases.
- Only the data required should be stored. The problem is then that other applications (e.g. a description of maintenance status) will use the same database. There is therefore an immanent conflict of avoiding redundant data and minimising the size of the data base.

An extension of available sensors in the future makes it conceivable that the volume of online data (e.g. the distribution of passengers in trains or waiting on platforms) will increase and will allow us to make even more precise predictions and thus more stable production plans.
4.3.4.3 Rescheduling measures

During the rescheduling process, different kind of measures are possible to generate an optimum and feasible production plan. For deviations and events with limited delays of up to several minutes, three different rescheduling measures are mainly used:

- Retiming: Trains are assigned with changed reference times. This may cause a speeding-up or slowing-down of trains. Also longer or shorter (if possible) scheduled dwell-times and new departure times can be assigned to trains. Finally, the determination of keeping to or breaking connections is also part of the retiming rescheduling measures.

- Reordering: Changing the order of trains in bottleneck areas or the position of a crossing or take-over. Sequence changes are implicitly combined with the retiming of the trains involved.

- Rerouting: Other routes may be assigned to trains locally within bottleneck areas.

Depending on the network layout (topology) and the timetable strategy, the application of these three different measures may vary strongly. The local rerouting of trains, in particular, is of maximum importance for areas with feeding lines merging a long distance before the station with many switch areas and mixed traffic. However, the increased number of routing and thus rescheduling possibilities concurrently improves the solution space and thus the complexity of calculations.

The implementation of the retiming measures may also vary a lot. In a simplest version, retiming is not applied for running trains and affects only changes for departure times and the planned dwell times if possible. More sophisticated rescheduling processes develop speed advisories in a second step after conflict solution by the traffic management. Depending on the network topology and traffic density, speed management may be implemented punctually or network-wide. In the most enhanced rescheduling algorithms, speed profiles are calculated as part of the rescheduling process to ensure feasibility and the best possible performance of the new production plan.
4.3.4.4 Generating new production plans

4.3.4.4.1 Strategies for efficient and rapid rescheduling

Generating a new production plan is one of the core processes of the rescheduling process within the traffic management system. Dynamic capacity utilisation and delay propagation are decided by it. Three different ways for generating new production plans are possible:

- Manual decisions based on simple, predefined rules and heuristic measures based on human experience.
- Semi-automated generation through a combination of manual decisions (initially to restrict the number of possible solutions and, finally, to choose a solution) and automated processing.
- Propositions generated fully automatically based on efficient algorithms.

For all ways of generating new production plans, various strategies and simplifications to optimise the duration and complexity are possible:

- time discretisation;
- limited prediction and calculation horizon;
- reduced routing possibilities;
- fixed speed profiles;
- fixed headways;
- pre-calculated solutions;
- geographical decomposition;
- multi-level approach; and
- combinations of different strategies.

These strategies for an efficient and rapid generation of new production plans are described below.
4.3. Elements of the real-time rescheduling loop

4.3.4.4.2 Time discretisation

The post-analysis of rail traffic data has shown, that trains arrive with specific distribution curves at reference points. To simplify the solution space, only discrete times or time slots can be assigned to the trains. This will reduce the solution space considerably. However, applying time discretisation may result in a loss of capacity depending on the discretisation size. And, it is essential to ensure that trains arrive during their assigned time slots.

4.3.4.4.3 Limited prediction and calculation horizon

The propagation of a single deviation can cover a large part of a network and can have consequences during a longer time. To prove the quality of a new production plan, all consequences must be taken into account. This would cause complex calculations and, combined with the fact that deviations occur again and again, new production plans may be needed before all the consequences of the first deviation are solved. Therefore, a reasonable prediction and calculation horizon (e.g. one hour), depending on network and timetable structure, helps to make the generation of new production plans more efficient, even if not theoretically optimum (compared to a solution that takes the entire day as a time horizon into account).

4.3.4.4.4 Fixed speed profiles

Calculating precise travel and passing times of reference points is complex and time-consuming especially for variable speed profiles and when the influence of signals must be taken into account. Using fixed speed profiles calculated offline therefore simplifies the process of generating new production plans. However, the actual status of trains not taken into account results in the generation of non-feasible production plans.

4.3.4.4.5 Fixed train headways

Fixed train headways reduce the complexity of algorithms comparably to fixed speed profiles. Constant instead of specific, train and infrastructure dependent headways may either reduce capacity or otherwise lead to headway conflicts because the fundamental blocking time theory will be violated.

4.3.4.4.6 Reduction of routing possibilities

Complex station areas have a multitude of switches and thus also many routing possibilities for each train. In the event of a disturbance, delay or maintenance, other routings may therefore be used without resulting in additional delays. However, if all routing possibilities are taken
into account, the solution space is unnecessarily large. A reduction of possible routes is therefore useful to minimise the duration of rescheduling. Nevertheless, possible solutions may be omitted. The challenge is therefore to find and define general rules for convenient routings.

4.3.4.4.7 Pre-calculation of event-specific production plans

Data mining of previous rail traffic data is used to identify common delays. This may be used to adjust the production plan generally or to pre-calculate new production plans offline. If a deviation is identified and assigned to a pre-calculated scenario, the adjusted production plan may be applied directly from the database without new calculations. Larger variations and thus deviations often occur during the boarding and alighting process in peak hours. Therefore, offline solutions with pre-calculated production plans may be conceivable for such cases.

In areas with high traffic density, delays propagate quickly through the network. Fast decisions are essential, even if the reasons for delay and the future behaviour of the train affected are vague or unclear. In particular for such cases, pre-calculated solutions immediately applied as urgent measures are of major importance. Therewith, traffic flow can be assured during the initial phase, which is used in parallel to collect more precise data for prediction and to generate a new production plan as a second stage.

4.3.4.4.8 Geographical decomposition

Decomposition of the network for dispatching is applied in all larger networks, otherwise it would be impossible to handle all traffic. Geographical decomposition therefore simplifies, on the one hand, data handling and, on the other hand, generating new production plans to solve conflicts or deviations are easier and quicker as long as no coordination with neighbouring areas is needed. A clever decomposition (e.g. with the concept of compensation and condensation areas, introduced in chapter 3.6), combined with predefined rules and predefined limited freedom for measures in each area can help to reduce the amount of coordination between separate sub-networks. However, in the event of larger delays, coordination between sub-network must be ensured.

4.3.4.4.9 Multi-level approach

A multi-level approach in which the possible rescheduling measures are extended step by step, is possible. The solution space and calculation complexity is extended gradually in order to reduce the number of trains involved and the area affected making the rescheduling process more efficient without reducing performance. In addition, a clear distinction between automated steps and manual decisions is possible.
Firstly, a solution is sought in which only a limited retiming of the train with the deviation is allowed. The rescheduled train can, in this case, be slowed down or speeded up and also obtain new dwell times on its route. Train sequences, routes and connections are fixed and not changed for all trains as well as all times for all other trains remain identical.

If a feasible solution cannot be found after the first step, rerouting of all trains without changing times will be applied to find a new production plan as a second step. A possible approach, actually used for offline planning is described in [Cai04]. For the first two steps, the result of the constraints is that performance measures will be equal for all cases and the optimisation objective in the second step is to minimise the total number of reroutings.

Retiming of all trains is allowed in a third step. In this, trains will only be shifted slightly in order to maintain all previously scheduled connections. Of course, reordering and rerouting of trains is also allowed. The limited time shifts of the trains also results in running time supplements being used to reduce the delay propagation and generating new production plans can still be done within a limited area without any coordination of neighbouring areas. Preceding offline calculations or post-processing after generating a new schedule may be used to determine the flexibility (or maximum shift) of the slots of specific trains [Cai07]. The first three steps can be implemented fully automatically without any input by human dispatchers. For the next rescheduling steps, decisions with larger consequences, such as breaking a connection, may require manual input.

If no solution can be found without larger temporal shifts, coordination between separated areas is required. Once again a twofold approach seems promising to generate a new production plan: firstly, at a macroscopic level, time slots are defined (coordination between areas) for all trains and secondly, feasible solutions are tried to generate locally at a microscopic level [Cai09a]. Certainly, the procedure is repeated if no local feasible solution is generated, and new slots have to be generated at macroscopic level. The macroscopic level also ensures that high performance is achieved. In particular, for performance functions with high weights on connections, this approach is well suited and appropriate because connections are defined at macroscopic level. Another coordination approach would be a direct exchange of constraints at the interfaces (boundaries) between two areas respectively. This exchange may be rule-based and requires an iterative process because a global coordinating level above is missing.

Finally, drastic measures, such as changing services (e.g. adding or omitting stops, adjusting circulation for rolling stock or changes in crew rostering, substitute trains) have to be taken into account for larger delays or disturbances. Traffic flow improvement is thereby of minor interest and also temporal requirements to the rescheduling system are of less relevance. Of course, these procedures also have to be taken into consideration, but are not of special focus within this present work. The flow-chart for the multi-level rescheduling approach is visualised in Figure 4.11.
4.3.4.4.10 Pulsing method

One approach, which combines several of the proposed strategies, is the pulsing method as it is known [Roo06]. The capacity critical condensation area is the basis for the rescheduling. Firstly, homogeneous speed is assigned for all trains within the condensation area and stops or slowing-downs are avoided whenever possible in the schedule. Secondly, at the boundary of the condensation area (so-called portals), only discrete arrival times are assigned to the trains. The topology and time within the condensation area is also discretised and coordinated. This results in so-called pulses, which are indeed individually used by trains, but designed in a specific way in order to ensure that blocking times of (most) trains remain feasible and within the pulse.

Homogenisation leads to the fact that the capacity may not be used to the maximum extent and in addition restrictions are needed for exceptions (especially in cases where occupation is longer than the pulse duration) [Roo06]. A comparison of the general, regularly used train diagram and a pulse-diagram is illustrated in Figure 4.12. Thereby, the pulse and the topology (tracks, platforms and switch areas) as well as the train routes and schedule can be identified.

However, reduced use of the infrastructure due to the pulsing model may be acceptable in a first proof-of-concept phase. The model allows the efficient identification of dependencies and conflicts in the schedule. And, in addition, new production plans can also be generated manually.
with the help of a visualisation tool within the shortest time and, depending on the complexity of
the topology and experience of the dispatcher, close to the maximum possible performance.
As long as conflict-free schedules including train trajectories have to be executed (partially)
manually, the pulsing model helps dispatchers to generate new schedules for condensation ar-
eas within shortest time. And, additionally, for trains approaching a condensation area, the train
path is automatically re-calculated in order to allow trains to pass at the portals at the given point
of time at the predefined speed. Also new routes and train sequences are directly visualised
for movement inspectors allowing them to take appropriate measures without intensive com-
munications. Of course, for larger delays, coordination between different traffic management
centres (condensation areas) still has to be done verbally by phone or other communication
channels based on manual (heuristic or rule-based) decisions. Nevertheless, the (theoretically)
opimum utilisation of a given infrastructure will not be achieved until time discretisation is small
(1-5 seconds). But this is definitively not manageable manually by dispatchers and will require
an automatic generation of new production plans.

4.3.4.5 Rescheduling policy

4.3.4.5.1 Overview of rescheduling policies

The rescheduling policy implemented has a significant impact on rescheduling duration and the
point of time new schedules are provided to all actors involved. In the event of a predictive-
reactive rescheduling strategy where new production plans are generated after a deviation or
event, three different ways to initiate the rescheduling process, rescheduling policies as they
are known, are possible: periodic, event-driven and hybrid. The three policies and their conse-
quences are described below.

4.3.4.5.2 Periodic rescheduling policy

For a periodic rescheduling policy, the state of the system is checked periodically by traffic man-
agement using a given sampling rate. Messages and information are collected in the meantime.
In the event of a deviation, a new production plan is generated. It is essential to ensure that the
duration needed to generate a new production plan is less than the cycle time. Another disad-
vantage of periodic rescheduling is that valuable time is wasted until the rescheduling procedure
is initiated.

On the other hand, in the case of geographically decomposed networks, coordination between
traffic management areas may be well structured as a two level procedure with global and
local rescheduling tasks. The time needed to generate a new production plan determines the
maximum sampling rate and is thus the most important design parameter. Particularly in cases
of events with drastic consequences, the limited calculation duration due to the sampling rate
may reduce the performance of the rescheduling system. On the other hand, low frequencies also unnecessarily reduce the performance of the rescheduling system in the event of a small deviation or events with restricted impacts on the network.

### 4.3.4.5.3 Event-driven rescheduling policy

As long as all system states are within their tolerance bandwidth, no event initiates the rescheduling procedure for an event-driven rescheduling policy. The rescheduling process will be initiated after exceeding of the threshold or due to other event messages. Thereby, two models can be distinguished: interruptible or non-interruptible event-driven rescheduling.

In the case of an interruptible event-driven rescheduling policy, an ongoing calculation is immediately stopped and directly restarted with the information of the new state. This may lead to the problem that the time required until a new production plan is generated is not predictable and may even be long if a multiplicity of events occur within the shortest time.

Overlong calculation duration due to a multiplicity of events may not occur for a non-interruptible event-driven rescheduling policy. However, an upcoming event before completing the rescheduling process is not taken into account and a non-feasible production plan may be transmitted. Therefore, after the transmission of the new production plan, a new rescheduling run is initiated directly, comparable to periodic rescheduling. Although the transmission of a non-feasible production plan may be avoided, it is finally preferred in comparison to an interruptible event-driven rescheduling policy where actors may have no production plan over a longer time period. Special attention is required if delays affect neighbouring areas, and coordination will be needed. Thereby, waiting times are unavoidable until a new production plan or constraints are generated in all affected areas.

To combine the advantages of interruptible and non-interruptible event-driven rescheduling policies, each event may be assigned a priority. If the sum of one or all the incoming events collected is larger than a given limit, rescheduling can be interrupted. Otherwise, rescheduling may not be interrupted anymore if the events handled (which initiated the rescheduling) are of a total priority that is higher than the limit.

### 4.3.4.5.4 Hybrid rescheduling policy

A hybrid rescheduling policy starts rescheduling periodically and in the case of major events immediately. The advantages of periodic and event-driven rescheduling policies are thus combined resulting in simply manageable coordination as for the periodic rescheduling and for urgent events, instantaneous rescheduling and no time wasting is thus possible. Consequently, it is necessary to prioritise the incoming events or deviation messages. High priority has to be assigned to all events or deviations in or close to capacity bottleneck areas (condensation
areas) and to disturbances, with the result that track sections will be unavailable for a minimum amount of time.

An overview and comparison illustrating the pros and cons of the different rescheduling policies is visualised in Figure 4.13.

![Figure 4.13: Visualisation of different rescheduling policies](image)

**4.3.4.6 Feasibility check of new production plans**

Predictions or assumptions of future behaviour are determined (see chapter 4.3.3.2) before the algorithms to generate a new production plan are initiated. After generating a new production plan, these assumptions must be compared with the actual situation. If the recorded data is within a given tolerance bandwidth compared to the prediction and if the new production plan is capable of being accomplished with the new characteristics, the production plan is feasible and may be transmitted to all actors afterwards.

A short calculation period results in a reduced number of non-feasible production plans. However, especially for non-periodic (infrastructure-based) train detection or longer sample times (in comparison to rescheduling period), it is possible that no updated information for a comparison is available. Consequently, a wrong prediction will result in the rescheduling procedure being initiated shortly afterwards when the next train detection point is or should be passed. This may finally lead to a frequent and thus nervous rescheduling procedure. To achieve an optimum balance between short calculation periods and nervous rescheduling, extended experience and field tests will be required.
4.3.5 Transmission of new production plans

The rescheduling process is finally completed by transmitting the new production plan to both the database for the supervision part in the traffic management system and to all actors affected (drivers, guards, infrastructure operators and movement inspectors). This down-link can also be carried out automatically by GSM-R, electronically or other communication channels. Otherwise, it may be possible to handle the large dataset (entire trajectories for driving recommendations or new routes) in combination with the up-link from the actors to the rescheduling system. A two-way communication system can be set up therefore. The maximum possible amount of transferrable data per time must also be respected.

4.3.6 Rescheduling process duration

4.3.6.1 Impacts of train detection on rescheduling duration

The overall duration of the rescheduling process from the point of time when the threshold is exceeded until a new production plan is transmitted to the actors is a key characteristic of the outer rescheduling loop. The first part of the process depends mainly on the train detection method.

Based on the model developed and given infrastructure characteristics, the time lags for infrastructure-based train detection was identified (see Figure 4.14). The detection density and thus the time taken until a next detection point is passed (for an early train) or has to be passed (for late trains) are the factors with major impacts on the overall duration. Depending on the train’s position and speed, the time lag is between several seconds and up to 5 minutes for a network like the SBB. In average, the time lag is expected to be between 30 and 60 seconds. The time required also depends significantly on the level of automation of the data processing. For a system where threshold excesses are detected and forwarded fully automatically to initiate the subsequent rescheduling process, the analysis showed that an overall delay of less than 2 seconds is expected. A manual approach, on the contrary, is expected to cause an additional time delay of at least 30 to 90 seconds.

In contrast to infrastructure-based train detection, in which data flows differ for early or late trains, periodic state transmission from the train to the rescheduling system always has the same data procedure. The analysis of the time periodic transmission modelled has shown that the time taken from the point of time when the train exceeds a threshold until data is prepared for rescheduling depends almost solely on the time taken for the next sampling. Technical constraints delaying the process beside the sampling rate are expected to be around 2 seconds. In order to speed up the process and to cancel the time taken until a next sampling occurs, a data trigger from the train to the rescheduling system may be set up. The train also requires the actual trajectory and information on the tolerance band as well as an intelligence which can
4.3. Elements of the real-time rescheduling loop

Figure 4.14: Comparison of process and transmission durations for infrastructure based train detection

detect a threshold excess. The sub-processes, transmissions and time lags for time periodic transmission are summarised in Figure 4.15.

Consequently, as long as train update frequencies are smaller than the frequency of passing train detection points, the time lags and delays for infrastructure-based train detection will be significantly larger in average compared to periodic train position transmission. To minimise the overall rescheduling duration, the periodic train position transmission should thus be preferred.
4.3.6.2 Rescheduling principles and their consequences on the overall time taken

Several fundamental decisions that have a major effect on the time required for rescheduling must be taken:

- Which actions must be carried out by human dispatchers?
- Should new production plans be generated if only improper, vague or no information is available?
- Should non-feasible production plans nevertheless be transmitted?
- Which rescheduling policy should be selected?

These questions and their consequences are discussed below. The tasks which must be completed as part of the integrated real-time rescheduling framework by the dispatcher will, of course, influence the time needed and the quality of overall rescheduling. Ideally, dispatchers should only act in abnormal situations. For example, requesting further information from actors in the event of missing data or entering data manually as input for the
rescheduling system. Also, in the event of larger delays where, for example, connections may be broken, dispatchers should select or agree on an automatically generated production plan. If other activities such as calculating new schedules are done manually, significantly larger and also imperfect solutions (compared to automated rescheduling with a given objective function) are expected.

Continuing the rescheduling process based on poor data avoids long delays until more precise data is available. Consequently, assumptions based on general rules or empirical values must be made. However, as soon as updated information is available, previous assumptions can be replaced and the rescheduling process must be re-initiated. Consequently, rescheduling based on improper or vague information leads to a higher nervousness of the rescheduling system. However, the train detection method applied from the first part of the outer rescheduling loop not only affects the time taken in this phase, it has also impacts on the quality and time required of the subsequent process. For infrastructure-based train detection, less data (quantity and quality) is often available in comparison to time periodic transmission. Coded messages sent by actors in the case of a deviation or event as part of the time periodic transmission would thus help enormously to speed up the rescheduling duration. Also, prediction quality would be significantly improved, which results in a reduced number of reschedulings initiated and, finally, in a less nervous system.

The question of whether non-feasible production plans should be transmitted is similar to the problem with improper data. If a non-feasible production plan were transmitted, a new rescheduling process would directly be initiated. This would therefore lead to a nervous rescheduling system. On the other hand, if a non-feasible solution is not transmitted, there is a risk that no feasible solution is found for several rescheduling cycles. This may result in the problem that over a longer time period no production plan exists. As a consequence, trains will be influenced or come to a halt which finally must be absolutely avoided.

As described in section 4.3.4.5, the rescheduling policy also has a significant impact on the time taken for rescheduling and on the rescheduling nervousness. The problems are thus comparable to improper data and non-feasible production plan problems.

Consequently, general principles and recommendations are not yet possible for these fundamental decisions. However, the computation time to generate a new, optimal production plan enlarges the entire rescheduling duration and must be taken into account for the analysis. We suggest carrying out an extended data analysis including a survey with test runs in order to evaluate which strategic decisions should be taken for all of these problems in order to achieve short rescheduling duration as well as a balance between rescheduling performance and rescheduling nervousness.
4.3.6.3 Time lag for the transmission of production plans

The time taken for the last process of the outer rescheduling loop, transmission of a new production plan to the actors involved only depends, in principle, on the level of automation of communication. In the event of oral communication, e.g. instructions by phone, a delay of around 15 seconds must be expected for each actor. Consequently, if several actors need to be informed by the dispatcher, the time required to inform them all will increase significantly. For the automated transmission of new production plans, a technical lag of one second is assumed. Thereby, multiple trains or actors can be informed almost simultaneously. However, also the time lag until a new production plan is applied by actors has to be regarded. For big yards in particular, significant time lags of up to 10 seconds to set new routes are possible.

4.3.6.4 Summary rescheduling duration

Manual actions in the rescheduling loop cause substantial time lags and should therefore be avoided whenever possible. For automated communication and control actions, the major delays are caused by the train detection process and the generation of the new production plan in combination with the prediction of future behaviour. The relevant time lags and times taken for processes are summarised below:

**Train detection (Identification of event or deviation)**

- Infrastructure-based train detection
  - Train is early: 2 seconds + time until next detection point is passed
  - Train is late: 1 second + time until next detection point is passed + tolerance bandwidth

- Time periodic train transmission
  - Purely periodic: 2 seconds + time until next data sampling
  - Hybrid transmission: 2 seconds

**Identification of primary delay reason - prediction of future behaviour - generating new production plan**

**Two-level approach**

- Identification and prediction based on vague data: 2 seconds
- Generating new production plan: Time to generate new production plan without coordination

**Optimal performance**

- Identification and reliable prediction: Time used to get oral information from actor and/or duration for multiple train detection data
- Generating new production plan: Time to generate new production plan including coordination with neighbouring areas

**Transmission of new production plan to all actors involved**

- Automated transmission of new production plan: 2 seconds
4.4 Influence of temporal aspects on rescheduling performance

The overview shows that the train detection method has a significant impact on the time taken for rescheduling. In a worst case, a temporal deviation is detected with a delay of the sampling duration (for time periodic transmission) or until a next detection point is passed (which could take up to several minutes depending on the density of infrastructure-based train detection systems). The technical delays during train detection are expected to be marginally around 1-2 seconds.

The second major time lag includes the tasks of identifying the primary delay reason, the prediction of future behaviour, and the generation of the new production plan. In a twofold approach, vague data is used for prediction and only a local valid solution is generated. Very short time lags thus seem reasonable. In a second step, more reliable data for prediction can be gathered. This includes time-intensive oral information exchange or data from several subsequent train detections. Also, coordination between several areas and thus longer computation times may be required in this case. Altogether, several minutes may be required for the entire process for the second level.

Overall, based on the requirements, a target time of 30 seconds for the entire rescheduling process for local measures within one area should be maintained. The analysis showed that for automated systems and dense train detection, a threshold of around 10 seconds to determine a new production plan exists. The developments on fast algorithms thus seems reasonable for satisfying the requirements. Fast reactions, in particular to improve train traffic flow in dense areas, can therefore be assured. In addition, the twofold approach guarantees that as soon as additional or more precise data is available, improved solutions are determined. Nevertheless, the interactions between short rescheduling duration and low rescheduling nervousness must be investigated under daily operations. This consideration also requires an extended data analysis and field tests before a general conclusion may be derived.

4.4 Influence of temporal aspects on rescheduling performance

4.4.1 Phases for rescheduling

The performance of the rescheduling system is heavily dependent on its temporal and quality characteristics. The relevant impacts and effects of time on performance are used to specify the limits for each task of the entire rescheduling process. An analysis of rescheduling showed that the following identified aspects must be considered:

- point of time when (or position of the train where) the threshold is exceeded (and consequently the time taken until this train enters the next critical bottleneck area where possibly other trains will be affected);

- time taken until a new production plan is generated and applied by all actors involved; and
time period and running time supplements available for other trains to use in the updated production plan to react to the deviation until they pass the critical bottleneck area.

The later the rescheduling process is initiated or the longer the generation of new production plans lasts, the more the possible number of rescheduling measures decreases. Basically, four different rescheduling phases are identified and can be distinguished:

- **Phase 1**: The rescheduling process is completed so early that all rescheduling measures including rerouting, reordering or retiming of all trains are possible. This may include a (partial) catch up of the primary delay of the train affected from the position the deviation (delay) occurred until the critical bottleneck area is entered. It may also allow the speeding-up (earlier reference times than in the original schedule) of other trains in order to change the sequence of trains in the targeted bottleneck area. Unnecessary signal stops or influence of distant signals can be avoided for all trains.

- **Phase 2**: The rescheduling process is terminated ensuring that enough time is left to react in order to prevent possible conflicts. Nevertheless, it is no longer possible to speed up trains heading for the critical bottleneck area. Consequently, only delaying, reordering and rerouting are achievable measures.

- **Phase 3**: Conflicts (unintended slow-down or stops in front of signals) are either unavoidable or have already occurred before the rescheduling process is completed. However, to minimise the further propagation of knock-on delays, rerouting, reordering and delaying trains are possible measures. The later the rescheduling is completed, the more trains are already affected by conflicts. Similarly, the number of rescheduling possibilities and, consequently, performance decreases.

- **Phase 4**: Finally, no rescheduling action is taken. Buffer times as well as running time supplements are used to get all trains back on-time. This (theoretical) case corresponds to the worst performance.

Table 4.3 summarises the four different phases and their possible rescheduling measures.

<table>
<thead>
<tr>
<th></th>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Phase 3</th>
<th>Phase 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conflicts prevented actively</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Speeding up trains</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Delaying trains</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Reordering trains</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Rerouting trains</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>
4.4.2 Temporal effects on rescheduling performance

The resulting performance after rescheduling as well as the boundaries of the different phases depend on many factors and the actual situation. The most relevant aspects include:

- controllability (or achievable accuracy) of trains even for late rescheduling (requiring abrupt changes in driving behaviour, q.v. the bathtub problem described in chapter 5.2.6.1);
- size of the delay and duration until the first conflict occurs if no rescheduling measure is taken;
- traffic density (number of involved trains);
- amount of running time supplements for all trains involved (possibility to speed up trains mainly used in combination with the reordering of trains); and
- number of rerouting possibilities.

Based on running time analysis and field experiences, the boundary between phase 1 and 2 is between 10 - 15 minutes before the train initiating the rescheduling is planned to enter the bottleneck area. The critical boundary between phase 2 and 3, where conflicts are becoming unavoidable, is based on the circumstances and is between 2 - 5 minutes before the train causing the rescheduling enters the bottleneck area.

Each rescheduling measure has a latest point of time by which it must be applied. With each rescheduling measure omitted, rescheduling performance also deteriorates. This characteristic leads to steps in the performance function. The position, size and quantity of the steps are also heavily dependent on the specific circumstances. Temporal discretisation to simplify or speeding-up the rescheduling process also affect the resulting function. The correlation between time and rescheduling performance is illustrated exemplarily for a generic case in Figure 4.16.

The importance of taking rescheduling actions as early as possible and of minimising the time taken for rescheduling to obtain maximum performance are also pointed out in Figure 4.16. In addition, this interrelations shows that in cases with limited time a multi-level approach with a pre-dispatching and a main dispatching phase (as introduced in section 3.3.2.2) may be applied. In the first phase, imminent conflicts should be solved immediately (e.g. by using predefined rules or pre-calculated solutions). During the second phase, the main rescheduling including optimisation can be executed.

Taking pro-active measurements before a conflict occurs (especially rescheduling phase 1) contradicts current principles where dispatching measures are mainly taken to react to conflicts. Consequently, phase 1 can not be implemented with conventional railway operation principles. Phase 2, where conflicts still are actively avoided, can also only be achieved coincidentally when
sufficient buffer times and/or running time supplements exist. The performance of dispatching algorithms not embedded in an integrated framework is thus unnecessarily reduced.

The controllability or achievable accuracy of running trains also impacts the rescheduling performance function. Imprecise train operations result, firstly in a worse performance index because more buffer times are required. Secondly, the boundaries between the different phases are shifted forward. This means, for example, that with the possibility of accurate train driving (including following an abrupt changeable trajectory) a conflict may be avoided whereas with reduced controllability probability increases.

4.5 Conclusions

As part of the integrated real-time rescheduling framework the outer rescheduling loop is responsible in the case of a deviation or event for generating and supplying new production plans to all actors and passengers involved within the shortest time. The temporal aspects of all processes of the rescheduling loop thereby have a significant influence on the resulting performance. In summary, these processes are:

- Detection of threshold excess and data preparation: The train detection method thereby has a major impact. For infrastructure-based train detection, the recording density is decisive. The delay in this case may be up to several minutes with today’s network characteristics. Train-based periodic transmission causes a maximum delay similar to the time taken for sampling. Additional lags of at least 30 seconds occur, when the detection of a deviation and data preparation are carried out manually by dispatchers.

- Prediction of future behaviour and generating a new production plan: A balance between short rescheduling times and nervous rescheduling is required. Depending on the inter-

![Figure 4.16: Impact of time on the rescheduling performance](image-url)
actions of the superimposed control loops and the specific circumstances, it is possible to define the requirements for the time taken for rescheduling.

- Transmission of new production plans to the actors involved: In the event of a fully automated system, the delay is below one second whereas orally communications with a actor involved causes a time lag of at least 30 seconds.

Summarising the rescheduling processes, it can be concluded that a total rescheduling period of 30 seconds or less is possible for local rescheduling measures. For this purpose, the following points must be fulfilled:

- What are required are periodic train-based transmission with a sample duration of 10 - 15 seconds or infrastructure-based train detection with comparable density (which implies a multiplicity of new detection points).

- Multi-level rescheduling where in the event of vague data or imminent conflicts, a new production plan is available within maximum 10 seconds. Improved solutions based on rescheduling algorithms and predefined objective functions are generated subsequently using more precise prediction data and requiring more time. However, an overall time of less than one minute is the goal as well as for reschedulings require the coordination of multiple areas.

- Manual actions within the rescheduling task should focus on collecting detailed missing data for precise prediction and selecting a possible rescheduling solution. The completion of time consuming tasks such as the detection of a deviation, developing or transmission of new production plans manually should be implicitly avoided.

Short rescheduling periods may be achieved based on vague data (awaiting more precise incoming data would increase the time), non-optimum solutions (network-wide coordination and optimisation is done subsequently) or non-feasible production plans. This may lead to a nervous rescheduling system with frequently changing production plans. New production plans distract actors, in particular, train drivers. Consequently, further research and a survey of field tests are suggested to strike a balance between short rescheduling times and an acceptable number of new production plans to be handled by a single actor approaching a bottleneck area.

Finally, the evaluation of the rescheduling on performance has shown that the point of time at which new production plans are transmitted to all the actors involved has a significant impact on performance. Early rescheduling also allows all available rescheduling measures to be used including the speeding up of trains and avoidance of conflicts, which is new in contrast to all existing rail traffic management systems. Consequently, the early detection of a deviation and short rescheduling time are a permanent goal.
Chapter 5

Accurate production control loop

5.1 Control loops for precise production in railway systems

The inner accurate production feedback control loop is described in details in this chapter. Special focus is placed on the train driver and the DMI with an analysis of different system architectures and designs as well as their impacts on achievable accuracy. The decisive elements and parameters of the inner accurate production control loop will be modelled and evaluated. In addition, the train departure process will be analysed, improved, and described in this chapter in order to achieve the precision that is also required for departing trains.

Improved stability and increased use of the infrastructure, which is the goal of integrated real-time rescheduling framework, requires trains to be operated precisely according to a dynamically changeable production plan. Accurate train operations include two parts: driving and departing. The goal for both modes is that a small tolerance band can be defined in which the trains remain under normal conditions. Therefore, the following objects are required:

- A detailed database (of topology, trains and passenger demand) to define feasible production plans by the traffic management system;
- A tool (man-machine interface, handheld) which supports the driver and guards with accurate information on the specified trajectory and the actual state; and
- Direct (and if possible continuous) communication for all actors involved with the rescheduling system.

The general process with the relevant actors of the inner control loop is visualised in Figure 5.1. Firstly, the elements and architecture of the precise operation control loop with focus on the driving process are described and evaluated. Secondly, an enhanced departure process, and finally, time-driven infrastructure operations are introduced.
5.2 The train driving process as part of the integrated real-time rescheduling framework

5.2.1 Closed loop design of train driving

The inner precise train operation control loop is developed as part of the integrated real-time rescheduling framework to improve the achievable accuracy and in particular to follow dynamically changeable trajectories. Nowadays, direct speed adjustment, e.g. verbal instruction by a dispatcher, is uncommon. Sometimes, passing times at specific reference points are suggested or drivers are asked to drive as fast as possible over a given section. Because oral commands are quite rare and undocumented, the accuracy obtained cannot be investigated. However, the idea of advising the driver has been used for a long time. Before radio communication was possible, movement inspectors or station staff showed a table at stations telling the drivers either to slow-down or speed-up for the next few sections [But72].

To satisfy the requirements, the information from the traffic management has to be transmitted automatically to the driver, designed as a closed control loop. Therefore, the inner train driving closed control loop consists of the three main elements: driver, train and driver-machine interface (DMI). The structure of the driving control loop is illustrated in Figure 5.2.

Figure 5.2: Design of the inner train driving process loop
5.2. The train driving process as part of the integrated real-time rescheduling framework

An essential factor for the success of an integrated real-time rescheduling framework is the architecture and design of the DMI. The DMI must be:

- intuitive and understandable;
- the data must be traceable and reliable;
- high availability and robustness must be guaranteed;
- the driver should not be distracted during safety-critical actions;
- complex and time-consuming inputs must be avoided;
- production, operation and maintenance must be cost-efficient.

In particular, in the case of a deviation or event, drivers are under maximum workload and should then be supported with and not disturbed by additional information or tasks. Coded messages sent by drivers to the rescheduling system shortens the duration of the rescheduling process. Consequently, such tasks have to be implemented in a way that the data transfer is completed within one or two working steps (e.g. key-press on the DMI). To minimise driver distraction, head-up displays may be a solution. Initial test runs with a head-up display have been carried out in England [Rod08a].

5.2.2 System architecture of the driver-machine interface

The focus of the train control loop is the DMI. Thereby, the system architecture of the DMI is linked to the design of the rescheduling system and the data which has to be transferred. In principle, two different strategies are possible to define a reference trajectory. The first possibility is to calculate the trajectories precisely in the traffic management centre as part of the rescheduling loop and then transmit them to the DMI. The other possibility is to only specify reference times at particular points (e.g. in front of a condensation area) by the rescheduling system and the intelligence (optimisation algorithms) calculating the trajectory remains on the DMI.

The strategy has a large impact on the system architecture of the DMI, illustrated in both cases in Figure 5.3. It shows that for the strategy with on-board optimisation and trajectory calculation, the tool is more complex, more data needs to be stored (entire topology) and automated data actualisation (online data with restrictions, e.g. for speed) must be handled. On the other hand, the data volume to be transferred is reduced for systems calculating the trajectory itself in the DMI. This may be of the utmost importance when information exchange between traffic management system and trains is technically limited. In addition, centralised optimisation also ensures that trajectories are conflict-free. To summarise, an architecture with transmission of complete trajectories is preferable.
5.2.3 Functional requirements for the driver-machine interface

The DMI has the task of visualising the reference trajectory, the actual state as well as the deviation for the driver. The values indicated are then processed by the driver which results in correcting measures. Depending on the requirements specified and the system architecture selected, the inputs for the DMI are:

- the reference trajectory;
- an exact, synchronised time;
- the actual train state including its position (e.g. GPS and/or odometer data) and speed; and
- data for the calibration of the tool (e.g. passing a reference point).

Depending on design and observability, inputs are either continuous or discrete.

The set-up of the DMI is decisive for the achievable precision. In particular:

- the displayed parameters and values;
- the accuracy and correctness of the values; and
- the display’s or parameters’ update rate.

Besides precision, the reaction of the driver after receiving a new production plan is also important in order to ensure that the changed reference trajectory is followed precisely from the beginning. The train’s high inertia must be taken into account by generating adjusted production plans. In modern trains, drivers are supported by a speed controller which results in less variations during operation. In addition, the number of inputs expected from drivers is reduced allowing them to pay more attention to the signals and the DMI.
5.2.4 Reference trajectory as input for the driver-machine interface

5.2.4.1 Modelling trajectory calculations

The reference trajectory is defined by the outer rescheduling loop and is the input for the inner precise train operation control loop. The shape of the reference trajectory thus has an impact on driving behaviour and achievable accuracy, especially after rescheduling.

The reference trajectory must be feasible and precise so that trains are able to follow it. Basic inputs to calculate the reference trajectory are:

- an accurate description of the topology and track characteristics (e.g. grades, curves, tunnels, track speed, switch positions, insulated sections);
- train dynamics (e.g. tractive effort, resistance, length, load);
- a description of specific operation actions (e.g. testing the braking effect after departure from terminal stations); and
- the schedule for the reference points at the beginning and end of the section including the desired speed at this points (determined at a macroscopic level as part of the rescheduling process).

Theoretically, if the problem is feasible, a multiplicity of reference speed profiles are possible. The enveloping solutions are given by the physical constraints, illustrated in Figure 5.4.

Various ways of calculating the reference trajectory are possible. One possibility is to divide the trajectory in several track sections. Each section then is weighted by a factor. The overall running time supplements are then iteratively distributed along the sections based on weight factors resulting in reduced speed. Speed changes and supplementary time distribution strategies can thus be taken into account [Gra06]. An approach using graph theory for segmented track sections to find conflict-free track paths is described by [Cai09c]. Another possibility is to consider the calculation of the reference trajectory as an optimum control problem [Gee07]. Energy optimal operations can thus be taken into account. One possible analytical method, solving
the optimum control problem by applying the maximum principle, is described in [Liu03]. However, speed changes in the updated reference trajectory should be planned whenever possible at points where the maximum allowed track speed also changes.

### 5.2.4.2 Strategies for calculating trajectories after rescheduling

After exceeding the tolerance bandwidth, a new trajectory is calculated and transmitted to the train. The new trajectory can either be the old trajectory shifted by a given value, or distorted based on the actual position and state of the train (see Figure 5.5 for a comparison of the two strategies).

Shifting a schedule is simple and only a limited amount of data needs to be transmitted. On the other hand, shifting a schedule prevents the tolerance band principle from being applied strictly. It is then possible that the position of the train after rescheduling can be outside the tolerance band. Therefore, a point of time must be defined until the train must be within the tolerance band again. Consequently, deviating behaviours will not be identified until this point of time. A further drawback of shifting a schedule is that overshooting may happen more frequently due to the fact that drivers may not be well supported by the available data.

The actual state (speed and position) of the train is the basis for distorting a reference trajectory. Especially for trains having a deviation, prediction of behaviour may be error-prone and distortion may also cause trains to be outside the tolerance bandwidth. In this case, the rescheduling process will be re-initiated. However, the behaviour of trains and thus deviations may be detected continuously resulting in fast rescheduling and fewer conflicts. The last fact is more important and, consequently, distorting an existing reference trajectory should be preferred.

![Figure 5.5: Principle of defining new trajectories: shifting with temporal offset (left) or distortion of existing reference trajectory (right)](image-url)
5.2.5 Possibilities of the driver-machine interface displays

5.2.5.1 Overview of display possibilities

Supported by the DMI, train drivers should follow the given reference trajectory as accurately as possible and remain within a given tolerance bandwidth. The display should therefore provide information concerning the reference trajectory and the actual state. Basic requirement for displaying a parameter is its observability and, in particular, its update frequency and available data accuracy. Parameters which can be displayed individually or in combination are:

- deviations from the reference trajectory (temporal);
- information on reference points (position and time);
- speed information (recommended, maximum and actual);
- recommended driving mode (hold speed, slow down, speed up, coasting); and
- foresight of data (similar to receding horizon control).

The DMI must be intuitively recognisable, readable and understandable. Consequently, only as much information as required should be displayed. Also to avoid confusing drivers (signalling aspects), colours should be avoided or used only sporadically. The high inertia and reduced controllability due to actuator saturation (limited braking and acceleration forces) of a train must also be considered. Providing a foresight and not only the actual state and deviation may help drivers to anticipate their behaviour. Many display possibilities are conceivable and a promising selection is described below.

5.2.5.2 Displaying temporal deviation

5.2.5.2.1 Determination of the temporal deviation

The deviation between the reference trajectory and the actual state is the most important parameter for a feedback control system. For public mass transit systems, temporal deviations have been used for a long time to inform drivers. The temporal deviation at a given position $\Delta t(x)$ is defined as:

$$\Delta t(x) = t_{ref}(x) - t_{act}(x),$$

where $t_{ref}(x)$ and $t_{act}(x)$ are the reference and actual passing times at the reference point.

Within station areas, the stopping position of trains may vary by up to several dozen meters. Using equation 5.1 strictly as defined results in the temporal deviation having steps with the size of the scheduled dwell time (see Figure 5.6 for an illustration of the effect).
Varying stopping positions are another source of temporal deviations, in particular at the beginning of the acceleration process or at the end before the train comes to a standstill (illustrated in Figure 5.7). Whereas temporal deviations within stations due to varying stopping positions cannot be avoided, these steps should not be displayed. Preferably, drivers should be supported by a display of the exact stopping point and the exact scheduled point of time of the departure instead of the actual deviation. A mode change is therefore suggested between station areas and open track.

**Figure 5.6:** Steps in the temporal deviation (right) caused by a variation of the stopping position

**Figure 5.7:** Temporal deviations after departing from stations due to varying stopping positions

### 5.2.5.2.2 Temporal deviation in station areas

To avoid steps within station areas, the reference trajectory must be adjusted. For arriving trains, the reference stopping position would be of most importance. Displaying time deviations are not helpful for this topic. Consequently, the temporal deviation can be suspended when a train enters a station which will avoid the step function for arriving trains. Instead, guidance to the intended stopping position should be preferably provided.
The difference between reference and actual stopping positions $\Delta s_{\text{stop}}$ due to imprecise operations can be used to adjust the reference trajectory in combination with a target speed after departure $v_{\text{target}}$ (e.g. maximum speed after short acceleration). Hence, the shifted reference trajectory $t_{\text{ref,shift}}(x)$ within the station area is determined as:

$$
 t_{\text{ref,shift}}(x) = t_{\text{ref}}(x + \Delta s_{\text{stop}}) - \frac{\Delta s_{\text{stop}}}{v_{\text{target}}} \quad \text{for} \quad \Delta s_{\text{stop}} > 0
$$

$$
 t_{\text{ref,shift}}(x) = t_{\text{ref}}(x - \Delta s_{\text{stop}}) + \frac{\Delta s_{\text{stop}}}{v_{\text{target}}} \quad \text{for} \quad \Delta s_{\text{stop}} < 0
$$

Figure 5.8 illustrates the offset-free trajectories for adjusted departure references in station areas. Of course, the trajectories are only offset-free if the trains have the same speed at the target point. However, it can be seen that temporal deviations from the original reference trajectory within the station area may not be avoided.

To react to changes and to take appropriate measures, it is not only important to know the actual temporal deviation, but also the temporal development of the deviation. In particular, the high inertia of trains can then be taken into account better. Therefore, rounded displayed deviation values (e.g. 10 seconds) should be avoided. However, this implicitly requires an accurate, realistic reference trajectory, synchronised times for all actors and precise position determination. Interviews with drivers showed that a frequent display update (1-5 seconds) should be the goal. Behaviour can be observed, checked for traceability, and divergences can be identified directly used to take appropriate measures finally.

### 5.2.5.3 Displaying reference points

Verbal contact between dispatcher and driver are rare nowadays. Recommendations on avoiding a conflict often contain information about passing times at specific points. This is also a
possibility for visualisation. Drivers would get a list with exact passing times for specific points as reference information on their display. This is comparable to the information drivers have nowadays, but more precise and adjusted to the actual circumstances to avoid signal influences. However, this visualisation would give drivers greater liberty and, consequently, the tolerance band principle would be more difficult to apply.

5.2.5.4 Displaying a foresight

Reference trajectories must satisfy the physical constraints of the train. Automated systems compare the difference between reference trajectory and actual state and deduce the control variable (actuator). The system characteristics (especially inertia) are taken into account in the control law. Nevertheless, controlling a system with large inertia is difficult for human operators, effects are observable after a time lag. Consequently, a foresight of the reference trajectory helps to plan the next actions. Doing so, appropriate and timely measures taken by train drivers can be executed. Similarly, displaying part of the behaviour from the past helps to identify diverging actions from which correcting measures may be derived.

A foresight helps drivers to use their experience and freedom of action to satisfy the requirements of following a given trajectory within a predefined tolerance bandwidth comparably to the case with the display of reference points. A special example of foresight is the DMI RouteLint [Alb07], tested by the Dutch railways. In this, the sections in advance and their occupation state (including train numbers) are displayed to the driver. Combined with the knowledge of dynamics and schedule of the trains in advance, the drivers can adjust their speed in order to avoid the influence of signals from other trains.

5.2.5.5 Displaying speed recommendations

Train drivers are used to working with speed information. Modern traction units have cruise control systems, so drivers only have to set the target speed and the traction or braking force. Consequently, the actual reference speed or the speed to be achieved after accelerating or braking can be displayed by the DMI. The recommended speed information displayed may be extended from a snap-shot view of the actual reference speed to an overview of the reference speed trajectory around the actual position. A small part of the previous behaviour and a foresight of the future can be displayed. A proposal in which the maximum track speed is also displayed, is visualised in Figure 5.9.

The difficulty with speed information is that they are safety-related. Overspeed due to a lack of missing information or disregarding information can cause accidents. Therefore, the displayed speed must be equal to or less than the maximum allowed speed. Particularly for temporary reduced speed sections (e.g due to short-term work-site), the updating of the database is of the utmost importance. However, ensuring that false information will not be displayed is cost-
intensive for signalling systems without continuously updated information as with ETCS level 2, for example. Nevertheless, train drivers are nowadays used to handling different speed information from varying sources which could even be contradictory.

An indication of the reference speed requires deviations from the original trajectory having to be taken into account and the trajectory having to be re-calculated. Without adjusting the reference input, an offset (delay) would remain if only speed information were displayed.

### 5.2.5.6 Displaying operation recommendation

Trains are either controlled by the driver through direct acceleration and braking commands or by setting a target speed. Consequently, following a reference trajectory, an announcement of the recommended operation (braking, acceleration, hold speed, coasting) is one possibility. Another way would be to use symbols, illustrating the action to take. Figure 5.10 illustrates seven possibilities:

- a: Speed up to the maximum permissible speed in the shortest possible time.
- b: Significant acceleration to higher speed.
- c: Slightly acceleration to higher speed.
- d: Maintain speed and driving behaviour.
- e: Reduce speed slightly.
- f: Reduce speed significantly.
- g: Reduce speed sharply within the shortest possible time.

Using these symbols avoids a direct display of the reference speed. Speed information is thus not directly provided. However, the reference trajectory is also not displayed with the best possible accuracy. This implies that operational accuracy also will be reduced.
5.2.5.7 SBB’s DMI FARE

Based on discussions with the Swiss Federal Office of Transport (BAV), the SBB decided to avoid displaying speed information to avoid safety relevant confusion on their test DMI. As part of the SBB’s FARE project, the tested DMI developed (a schematic illustration of the FARE-DMI for a typically delayed train is visualised in Figure 5.11) provides the driver with three types of information [Wüs08]:

- actual deviation in seconds with regard to the current schedule;
- recommendation for the driving behaviour/mode; and
- a foresight of 8 minutes indicating a $t - \Delta t$-Diagram.

Selection of the information displayed as well as the design of the DMI was carried out in cooperation with train drivers.

Whereas the first two types of information are intuitively understandable, the $t - \Delta t$-diagram needs a more detailed explanation. Two types of information are indicated by the 8-minute ahead preview. Firstly, the possible time compensation (considering technical and safety restrictions) is indicated as white bars for each minute. The possible compensation time thereby
describes the difference between maximum speed and planned reference speed. Secondly, the recommended time compensation per minute (based on a specific algorithm [Röt07]) is indicated as a grey bar. Positive values indicate the amount of seconds which must be made up for (i.e. when the train is late) and negative values indicate that the train has to lose time (i.e. when it is ahead of schedule).

Based on the size of the white bars, the driver can determine how close to the maximum track speed he has to run the train. Changes in the schedule accordingly affect the size in the white bars. An increase of the bar’s height implicitly means that the scheduled speed is lower than in the original schedule and vice versa. The interpretation and application of the bars and foresight involves a good knowledge of the track ahead. Also, being familiar with the original schedule helps to identify changes in the schedule and their consequences on the driving behaviour.

The reasons for choosing a $t - \Delta t$-Diagram by SBB are:

- avoiding wrong information of actual valid maximum speed;
- the bars with an indication of 8 minutes ahead allow drivers to use their freedom of action; and
- time instead of distance on the abscissa allows an appropriate preview to take relevant measures independent of the speed.

Technically, the FARE-DMI receives the current train position from the on-board unit and compares the value with the planned position to the actual time. The original schedule and the minimum running time are transmitted before the train trip starts. The reference schedule (or trajectory) is updated each time after a rescheduling when a change in time or route occurs for the affected train.

### 5.2.6 Driver in the loop model

#### 5.2.6.1 Source and size of temporal deviations in the train running process

##### 5.2.6.1.1 Deviations during daily train operations

Analysing train driver behaviour and the impact of different display possibilities, the following two criteria are of interest:

- maximum deviation in which a train remains under normal conditions; and
- train behaviour (including maximum temporal deviation) after rescheduling.

Initial sources of temporal deviations and its consequences are described below. A general train driver model then is introduced to gain insights into train running behaviour. This enables the impacts of different DMI design parameters to be analysed.
Temporal deviations from the reference production plan can occur continuously. In general, three reasons can be identified:

- actual speed is different from planned speed;
- acceleration or braking rate deviate from the planned rate; and
- initial braking or acceleration point differ from the planned braking or acceleration point.

As shown in Table 5.1, all three cases may cause temporal deviations of more than 10 seconds within a short time. Again, the need for frequent display updates as well as the requirement that deviations should be accurately displayed to also monitor the changing rates of temporal deviation are pointed out.

<table>
<thead>
<tr>
<th>Deviant speed</th>
<th>Planned speed</th>
<th>Deviant speed</th>
<th>Time taken before temporal deviation is greater than 10 s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>30 km/h</td>
<td>20 s</td>
</tr>
<tr>
<td>60 km/h</td>
<td></td>
<td>50 km/h</td>
<td>60 s</td>
</tr>
<tr>
<td>65 km/h</td>
<td></td>
<td>60 km/h</td>
<td>20 s</td>
</tr>
<tr>
<td>120 km/h</td>
<td></td>
<td>115 km/h</td>
<td>240 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>130 km/h</td>
<td>120 s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deviant braking rate</th>
<th>Planned braking rate</th>
<th>Deviant braking rate</th>
<th>Final deviation for a train braking from 140 km/h to 60 km/h with differing braking rates.</th>
</tr>
</thead>
<tbody>
<tr>
<td>−1 m/s²</td>
<td>−0.8 m/s²</td>
<td>4 s</td>
<td></td>
</tr>
<tr>
<td>−1 m/s²</td>
<td>−0.6 m/s²</td>
<td>10 s</td>
<td></td>
</tr>
<tr>
<td>−0.8 m/s²</td>
<td>−0.6 m/s²</td>
<td>6 s</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deviant braking point</th>
<th>Planned braking point</th>
<th>Deviant braking point</th>
<th>Final deviation for a train braking from 120 km/h to 40 km/h with a constant braking rate of −0.6 m/s² but differing initial braking points.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 m</td>
<td>−300 m</td>
<td>11 s</td>
<td></td>
</tr>
<tr>
<td>0 m</td>
<td>200 m</td>
<td>−6 s</td>
<td></td>
</tr>
</tbody>
</table>

### 5.2.6.1.2 The bathtub problem

In the event of a delay-retiming rescheduling measure shortly before a conflict area has to be passed results in a train having to slow down significantly and reaccelerate afterwards (see Figure 5.12). By doing so a conflict-free joining or crossing of trains is obtained. Consequently, it is
5.2. The train driving process as part of the integrated real-time rescheduling framework

absolutely essential to precisely follow the rescheduled trajectory. For this particular rescheduling scenario, also called the ‘bathtub problem’ (due to the shape of the speed reference trajectory), all three deviation cases may have a possible impact on the resulting deviation: too late slowing-down, wrong (inadequate) braking force and divergent reduced speed.

![Bathtub speed reference trajectory](image)

Figure 5.12: Bathtub (delay-retiming shortly before a conflict area) speed reference trajectory

The specific problem of bathtub rescheduling trajectories is that they occur shortly before a train enters a conflict point or conflict area. Thus, delayed or premature train arrivals will result in conflicts with preceding or following trains.

Having no information about an abrupt reduced speed target thus makes it almost impossible for drivers to remain within a reasonable tolerance bandwidth. When the train driver tries to achieve a too high speed target, the train will arrive at the conflict point too early resulting in unnecessary influence from the signal and forced to slow down. On the other hand, if the aspired target speed is too slow, the train will be delayed too much causing possible conflicts with following trains. In general, slowing down a train substantially (like the bathtub reference trajectory) simply to improve traffic flow (without having a closed signal or a special event) is new for train drivers and will require special training.

### 5.2.6.2 General driver in the loop block diagram

Developing a driver model helps to understand a train driver’s behaviour and can be used for the development of a DMI. Of course, factors such as motivation, stress, workload, fatigue, experience or changing conditions have an impact on a driver’s performance and behaviour, but have not been included yet in the model description because of a lack of data.

Whereas models for train drivers have not been developed so far or only used to analyse driver performance [McL05], the human pilots in the loop models have been of interest for control system engineers for many decades [Ama00, Bla91, Dud97, McR67]. In particular, oscillations due to position and rate saturation causing several spectacular incidents (see for example [Dor92])
motivated engineers to develop realistic models. Also for railway systems, rate-limit and saturation often occur and combined with time lags result in non-linear and possibly oscillating behaviour. This means, that due to driver inputs, the train may alternatively be too early and then too late in comparison to the reference trajectory. Oscillation rate and maximum deviation thereby depend on the design parameters.

The human behaviour of pilots has been described as a decision process [Bla91]. Based on this description, we have developed a train driver behaviour model in this thesis. After an initial delay, the data visualised is compared with the previous actions and the driver tries to understand the reason for changes or remaining constant deviations (estimator). Based on initial insights, the driver predicts future behaviour (predictor) and determines appropriate measures (controller). After a time delay, the driver's input is finally set. A block diagram, visualising the general train driver model, is illustrated in Figure 5.13.

![Block diagram of a train driver in the loop model](image)

Figure 5.13: Block diagram of a train driver in the loop model

To understand and use train driver behaviour for analysis and evaluation tasks, the decision process is modelled as a flow diagram. The actions taken and the underlying decision process are thus revealed. Where pilots, for whom a decision process was introduced first, have to control 6 degrees of freedom, train drivers can only control the speed (one axis). Omitting the coupling of several variables affecting multiple axes as for aircraft operations simplifies the model for railway systems. But, pilots can detect deviations (e.g. from the horizontal axis) without the aid of a display whereas, in contrast, train drivers must look at the display actively to identify the actual temporal deviation. This leads to longer reaction delays on average because the display will not be continuously monitored. This time delay also argues for an acoustical advice in the event of larger deviations or for incoming new trajectories in the event of a rescheduling.

### 5.2.6.3 Driver-decision process flow diagram

#### 5.2.6.3.1 Train driver-decision process

Based on the train driver model developed, we set-up a detailed train driver-decision process. The detailed decision process, illustrated in Figure 5.14 as a flow diagram, is heavily dependent on the information available to the train driver and is based on the driver in the loop block diagram (Figure 5.13). It is assumed that the temporal deviation at least is available to the
driver. The first step in the decision process basically depends on the availability of additional information (e.g. speed or operation recommendation). If no additional information recommends a change in actual behaviour, the driver will either wait and gather further information or use all available information to define a new speed target and also a strategy (e.g. heavier/lighter braking/acceleration) how to reach his target speed.

Interviews showed that speed is the most important decision variable for train drivers. For displays with adequate supplementary information, pro-active measures by drivers are possible. Of course, safety-related information from other sources (e.g. signals, speed-restriction boards) can be treated as additional DMI information and also causes drivers to change their target speed and driving behaviour directly. However, if indirect information such as target times at specific points is displayed, train drivers must translate this information into a new behaviour including target speed.

As long as supplementary information is either not available or does not request a change in behaviour, the train driver will compare his previously anticipated behaviour with the actual changes in the temporal deviation. In the event of an unexpected development in the temporal deviation and/or rate of change of deviation, the reason for this diverging trend will be identified and, if further information is available, compared with the recommended measures for plausibility.

Future behaviour is predicted based on insights during the estimation process step. Depending on the decisions taken during the estimation phase, three cases are distinguished:

- Is the train’s behaviour actually significantly influenced by the high train inertia?
- Is overshooting foreseeable?
- Is a new target speed required?

After the prediction phase, the entire process is terminated with the control action. Insights from the prediction phase are used to determine the new driving behaviour and the required input.

The time lags in the decision process primarily depend on the driver’s frequency scanning the display as well as the display’s update frequency. However, rounding information (e.g. temporal deviation with an accuracy of 5 seconds) also results in time lags until information is changed in the display. Consequently, one can conclude that more frequent display updates and more precise information help to reduce the time required for stable estimation and improve forecasting accuracy. The decision process that influences the performance is described in greater details below.
Figure 5.14: Flow diagram for the train driver decision process
5.2.6.3.2 Train driver control actions

Trains have low acceleration and braking limits. Inputs set by the drivers consequently often saturate at the braking or acceleration limit. Altogether, the driving mode applied can be:

- maximum acceleration;
- reduced acceleration;
- maximum braking;
- reduced braking (only with electric brake);
- coasting;
- holding actual speed; and
- transition from one driving mode to another.

As visualised in Figure 5.14, there are five different cases to determine the control input $u_{in}(t)$ during the controller process step exist. The case decision is a consequence of the previous estimator and predictor steps of the decision process. The five different cases are:

1. No input changes. Driver proceeds with actual driving mode and input:
   $$u_{in}(t) = u_{in}(t-1).$$

2. Previously defined target speed will be approached soon. Driver reduces actual braking/acceleration input such that constant speed will be achieved as desired:
   $$u_{in}(t) = f\left(\Delta t, \frac{d(\Delta t)}{dt}, v_{target}(t), v_{act}(t)\right) < u_{in}(t-1).$$
   This case describes the transition from an acceleration or braking driving mode to the speed holding mode.

3. Previously defined target speed will not be achieved in the near future. If possible, driver enforces actual braking/acceleration input to reach the target speed more quickly:
   $$u_{in}(t) = f\left(\Delta t, \frac{d(\Delta t)}{dt}, v_{target}(t), v_{act}(t)\right) > u_{in}(t-1).$$

4. Actual target speed has changed or was determined to be incorrect by the driver. If no information about the new target speed is available, the driver must determine a new target speed and, accordingly, a driving strategy (lighter/heavier braking/acceleration input or coasting) as to how the new target speed will be obtained:
   $$v_{target}(t) = f\left(\Delta t, \frac{d(\Delta t)}{dt}, v_{target}(t-1), v_{act}(t)\right)$$
   and
   $$u_{in}(t) = f\left(\Delta t, \frac{d(\Delta t)}{dt}, v_{target}(t), v_{act}(t)\right).$$

5. Information by the DMI seems erroneous or incident/disturbance makes it impossible to follow the reference trajectory. Driver continues to run the train without DMI information and calls traffic management centre.
The following parameters are used to do this:

- previous input \( u_{in}(t - 1) \);
- actual temporal deviation \( \Delta t \);
- rate of change of the temporal deviation \( \frac{d(\Delta t)}{dt} \);
- actual target speed \( v_{target}(t) \);
- previous target speed \( v_{target}(t - 1) \); and
- actual speed \( v_{act}(t) \).

Without providing a target speed or driving mode information, case 4 is of special interest after rescheduling (in particular for abrupt changes as in the 'bathtub' scenario). In this case, drivers must determine the new target speed itself based on the information available. In doing this, the maximum deviation that limits achievable accuracy and controllability occurs.

### 5.2.6.4 Train driver control loop characteristics

Without having additional information beside temporal deviation, the train driver determines control input and target speed primarily according to the train dynamics, on the actual temporal deviation and the rate of change of the deviation. This decision process can be considered as a PD(proportional-derivative)-controller, as described in [Ast95]. PD-controllers can be specified with the input control law function:

\[
u_{pd} = k_{pd}e(t) + T_{pd}\frac{d}{dt}(e(t))
\]

with the actual error \( e(t) \) and the two design parameters \( k_{pd} \) and \( T_{pd} \).

Depending on the train driver’s behaviour (which influences design parameters \( k_{pd} \) and \( T_{pd} \)), displays updating frequency, and accuracy of the information displayed, different developments of the temporal deviation are possible after rescheduling. Particularly for larger speed changes, three different divergent transient effects are possible (see Figure 5.15). The temporal deviation:

- can oscillate with low damping;
- converge slowly (including eventual small overshooting) due to high damping;
- have a remaining offset.

However, PD-controllers are also able to handle dead-time processes (or processes with high inertia) as it is the case with train operations or in general for most path tracking applications. PD-controller also react quickly and overshooting is minimised due to the derivative part. Another characteristic of PD-controller is the small remaining offset. But, for the precise train operation control loop, the remaining offset is negligibly small.
If train drivers are provided with more information, this helps them to determine adequate measures. Without any additional information, train drivers do not know if the new target speed is 10 or 50 km/h slower than originally intended when the temporal deviation becomes increasingly early. Consequently, specifications for the driving mode or information on the target speed allow the drivers to very precisely determine the required measure as part of the decision process as introduced in the section 5.2.6.3.1 above. Using the information given to the driver, deviations in the target speed as well as the point where speed is changed are small, causing deviations of limited size (see Table 5.1 for occurring delays). Having this information, drivers can pay special attention to the development of the temporal deviation and thus take appropriate measures by adjusting their braking or acceleration inputs. So, the resulting temporal deviation can be precisely controlled. Overall, exact information and frequent update of temporal deviation in combination with supporting information on the new driving behaviour or target speed allow a train to be controlled with an accuracy of at least $\pm 15$ seconds. Finally, controllability and short-time reactivity consequently define the limits of rescheduling by means of latest point of time for rescheduling without getting conflicts and thus maximum possible performance.

Extensive experimental test runs must be executed to develop, test, evaluate, validate and calibrate the train driver input control commands. Without these test runs, first assumptions had to be taken and simulations were run only to test initial models. The test runs should also be used to evaluate the middle and maximum frequency update rate of train drivers’ decisions and manageable input changes. In addition to this, the handling with high train inertia can be analysed. Finally, behaviour without a supporting tool and with different displays has to be investigated.

Of course, fully automated systems with adaptive, state feedback or processor based optimisation controller will react within the shortest time, detect trends and deviations early and will thus finally have smaller temporal deviations. A discussion on semi-auto-piloted driving to optimise train traffic flow in conflicts or heavily used areas is thus unavoidable in the future.
5.3 On-time departure

5.3.1 Today’s departure process

5.3.1.1 Procedure of today’s departure process

Precise control of running trains includes their accurate departure. The train departure process is a common reason for delays nowadays. In order to improve quality for an on-time departure, the actual process is analysed and sub-processes are quantified first.

The departure process of a train can be initiated, if the following conditions are fulfilled:

- route is set (signal is green);
- scheduled departure and minimum dwell time is passed;
- driver and guards are ready;
- main boarding and alighting completed; and
- train preparation is finished.

The procedure and conditions thereby depend on the type of station. The train preparation process especially is different for dead-end or terminal stations in comparison to a through station. When the last condition is satisfied, closing and locking of the doors is possible and the subsequent departure process, illustrated in Figure 5.16 initiated. The departure process after departing conditions have been fulfilled differs, on principle, between trains running with or without a guard.

On conducted trains, the guard is responsible for closing the doors and giving the final departing permission. This means, that as soon as the closing and locking of the doors is possible, he has to announce the imminent departure with a whistle and a hand signal or in the darkness with a pocket lamp. After that, he moves to the switch box to grant permission for departure. Subsequently, he enters the train and activates door locking by using the UIC switch. After that, all train doors will be closed and locked. As soon as the driver gets the information "doors locked" transmitted to the driver's cab, he can accelerate the train. After a short delay due to the control system's reaction, the vehicle begins to accelerate. Depending on the rolling stock or the station, the processes described can be executed with some minor changes. For example, the dispatch of a train can be handled by the station staff and not by the guard. In this case, the time required to move to the switch box and back to the door is omitted.

On trains without a guard, the driver is fully responsible for the departure process. As soon as departing conditions are fulfilled, the 'close doors' command can be activated. Flashing lights or audio warnings may be activated depending on the train's equipment. The entire door locking
process is thus automated. When the doors are locked, the information is shown in the driver's cab and the driver can accelerate the vehicle.

Figure 5.16: Setup of today's departure process
5.3.1.2 Time required for today’s departure sub-processes

To identify any weak spots in the process, manual measurements of the time required for all sub-processes were taken in both, through and terminal stations for trains with and without guards. Based on the results, concentrated in Table 5.2, it can be seen that trains in through stations have, on average, a smaller total departure delay in comparison to a terminal station. Another insight from the measurements is, that the departure process for conducted trains takes, on average, 10 seconds longer than for trains without a guard. This difference is mainly caused by the final permission by the guard. A final observation is that trains without a guard have more statistical outliers for longer duration. Problems are, that drivers have a bad overview of the doors and intervention to passengers blocking doors is almost impossible by them within a reasonable time.

Table 5.2: Overview of departing delays (time taken after all conditions are satisfied until the train departs)

<table>
<thead>
<tr>
<th></th>
<th>Through station</th>
<th>Terminal station</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With guard</td>
<td>Without guard</td>
</tr>
<tr>
<td>Percentile P10</td>
<td>19.4 s</td>
<td>11.2 s</td>
</tr>
<tr>
<td>Median</td>
<td>30.0 s</td>
<td>23.0 s</td>
</tr>
<tr>
<td>Percentile P10</td>
<td>43.0 s</td>
<td>48.4 s</td>
</tr>
</tbody>
</table>

The results also show that an improvement in departure accuracy is required in order to reduce scheduled buffer times. The measurements indicated that the basic requirement for on-time departure remains: all departing conditions must be satisfied at the planned departing time. So far, these requirements were only fulfilled on-time in 40 percent of all measurements. A detailed description of measurements for all sub-processes and results of the today’s departure process are given in [Joh07a, Joh07b].

5.3.2 Design for new departure process

The measurements of today’s departure showed that after departing conditions are satisfied, the departure process can cause delays of up to one additional minute. Technical improvements and changes in the process are necessary to ensure a maximum delay which corresponds to driving accuracy.

The idea of a parallelisation of the processes to reduce the time taken was discussed by the SBB in 2004 [Lau04]. As a result, some actions have to be executed in parallel and not sequentially. Therefore, we describe and propose a new departure process (illustrated in Figure 5.17). An example of possible parallelisation is that the door locking process is initiated before the route is set. To avoid passengers waiting in front of closed doors, which, in fact, would be unpopular, the
5.3. On-time departure

train drivers and guards need accurate and up-to-date information when the route for departure is set.

Fixed times for the latest point in time for preparation completed, communicated departure time and scheduled departure time must be given and made available to all actors involved. The communicated departure time for passengers differs thereby for the scheduled departure times provided to all other actors involved.

Figure 5.17: Design of the departure process with improved precision
Another necessity is a reduction in the duration and variation of all sub-processes. Staff on platforms or new technologies - for example dynamic passenger information systems, handhelds for guards or advanced door closing systems - are needed to achieve greater accuracy and reduced delays. Of course, for stations with dominant (long) boarding and alighting times, accurate data is needed (demand, passenger distribution, rolling stock, etc.) to calculate the dwell duration precisely [Buc08]. Applying a knowledge of earlier measurements of variations also allows us to determine meaningful (and thus also not overlong) buffer times for the boarding and alighting process in order to improve the number of on-time departures significantly. Initial tests by SBB showed, that departure within a range of 15 seconds is possible [San07].

5.4 Time-driven infrastructure operations

Nowadays, routes are often set automatically by the interlocking system. A predefined route is thereby reserved if this section is released and the head of the train has passed a selected detection point. However, an event may result in a pre-reserved section not being used by this train. The route release task, which has to be done consequently, is a safety-critical action and results in a significant time delay.

Routes should be set as late as possible to avoid time-consuming and safety-critical measures. A late allocation of the route to a train also allows trains to be rerouted as an efficient rescheduling measure at the last moment. Nevertheless, routes have to be set early enough in order to prevent trains being influenced by a signal. Therefore, today's event-driven route-setting actions should be executed comparable to the driving of a train in a time-driven way. Consequently, a route has to be set when the train is following the specified trajectory and within the production plan defined time window.

Time-driven infrastructure operations require movement inspectors and signallers to be continuously updated with the actual production plan giving information on the routes and times of all trains in their area. In particular for the departure process, the change to time-driven operations is of major importance and will help to reduce variations in the departure process.

To speed-up the entire procedure, direct access and control in the traffic management center of safety-related route settings and further movement tasks seem to be beneficial because intermediate and time-consuming steps are omitted. Enhancements by suppliers are in progress resulting in the integration of remote control and traffic management functionalities.
5.5 Conclusions

Precise production following a dynamically changeable reference trajectory must be guaranteed by the inner feedback control loop. Today’s operation principles without a supporting tool for train drivers lead to large variations in the train running process and emerging conflicts cannot be avoided. The specified requirements of new reference trajectories after a rescheduling can thus not be fulfilled without a supporting display tool. Consequently, investigations focused mainly on the system architecture and design of the DMI.

Temporal deviations in the train running process occur for deviating speeds, excessively heavy or light braking or acceleration inputs, or differing points where braking or accelerating is initiated. It was demonstrated that even small deviations in one of these actions might cause larger consequential temporal deviations amounting to several 10 seconds. Especially for abrupt changes in the reference trajectory to avoid a conflict, e.g. a train will be delayed for two minutes just before a bottleneck area is entered, temporal deviations due to late and wrong reactions might be up to one minute. This highlights the need to have a DMI supporting the train driver in all possible situations in order to enable him to follow a reference trajectory within predefined limits.

Interviews with train drivers have shown that displaying the actual deviation is of utmost importance. The temporal deviation displayed should thus have an accuracy of one second. Actual behaviour can also be checked for traceability by the driver. Monitoring the rate of change of the temporal deviation, too, allows drivers to identify divergent or wrong behaviour and helps to take appropriate counteraction. This implicitly requires a frequent display update, preferably every second. However, precise determination of the actual temporal deviation requires:

- a precise and feasible reference trajectory;
- the precise determination of actual train position; and
- synchronous times for all actors and the rescheduling system.

For a significant improvement of the infrastructure utilisation, an accuracy of at least $\pm 15$ seconds must be achieved for trains running on open track. Test runs with different displays must be executed in order to prove daily applicability and achievable accuracy. Due to the high inertia of trains, oscillations of the temporal deviation or high damping (slow convergence towards the reference trajectory) are possible if only the temporal deviation is displayed. To reduce these effects and to improve driving accuracy, in particular to support drivers in cases with abrupt reference changes due to rescheduling, additional information such as recommended speed or driving mode including a possible preview might be displayed. Without such additional information, it is doubtful that an accuracy of $\pm 15$ seconds can be achieved.

Special attention is required for station areas and the departure process. Imprecise position information and varying stopping points can lead to an overall temporal deviation of 15 seconds
without any faulty operation by the driver. Therefore, distance indications for accurate stopping and adjusted reference trajectories (based on the actual stop position) are required to avoid the resultant errors. Also, the departure process itself needs special attention. Variations in departure delays of up to one minute are possible nowadays with conventional departure processes. To achieve an accuracy of the departure process comparable to the one of running trains, process enhancements such as a parallelisation of tasks, display of activities to be completed for the actors or the application of information systems in order to advise passengers are needed. Initial tests showed that a departure accuracy within a range of 15 seconds can be achievable.
Chapter 6

Integrated real-time rescheduling framework

6.1 Combining rescheduling and accurate production control loops

The models and insights of the previous chapters for the inner and outer control loops are combined in this chapter. The outcome is a detailed model of the integrated real-time rescheduling framework described with a data flow diagram. Interrelations and dependencies are identified with the aid of this model and subsequently described in this chapter. Special attention is given to the tolerance band principle, which is a central aspect in the integration of the control loops and a decisive parameter for the performance of the system.

In addition, the potential benefits obtained by applying the new framework are derived with the focus on improvements in stability, capacity and energy consumption. Finally, attention is drawn to aspects concerning the new roles of human beings in railway operations and standardisation challenges with integrated real-time rescheduling.

6.2 Integrated real-time rescheduling data flow diagram

6.2.1 Data flow diagram principles

The data flow diagram (DFD) is a network representation of an automated, manual or mixed system [DeM78]. Each component piece with all its interfaces and connections of the system is portrayed with the DFD. Possible proceedings, elements and ideas of the real system may be modelled with DFD and can then be used to analyse how a system will react in reality. Critical elements and in particular interfaces between elements may thus be identified.
The visualisation of the data processing and data flow through the system is a useful means. Connections, dependencies and their effects can be recognised. The basic elements of DFD are data flows (represented as vectors), processes (linked by vectors) and data sources and sinks.

### 6.2.2 Data flow diagram model for railway operations

Based on the cognitions described in the previous chapters, we developed a DFD including all the relevant actors and elements for rail operations (see Figure 6.1). The coherences in railway operations, the interactions between the different control loops, and the enhancements of an integrated real-time rescheduling framework required can be visualised.

The model developed has a generic configuration. Depending on the technology or operating method applied, some of the processes do not exist. For example, the detection and transmission of a deviation is modelled in three different ways: frequent transmission by the train, interlocking data and oral transmission by players. By omitting processes or connections, different standards or rescheduling strategies may be visualised and analysed. The model is particularly useful to identify the basic technical requirements for enhancements as well as to analyse various scenarios and operating methods.

One speciality of rail operations is that many processes handle data and coinstantaneously may be sinks or sources of data flows. The process detection of an exceeded threshold for example has all three possibilities:

- incoming data is processed, e.g. identifying a deviation exceeding a given tolerance bandwidth, resulting in the continuation of data flow;
- the incoming data may be within the limits and data flow is stopped; or
- an internal supervision event is triggered because data is missing resulting in the start of a data flow.

The model introduced illustrates regular processes. Unexpected disturbances can occur almost everywhere (from outside, inside a process or also during the transmission) and are not illustrated in the model if they are not especially addressed by a specific process.
Figure 6.1: Detailed process and dataflow model of the integrated real-time rescheduling system
Page intentionally left blank
6.3 Interaction of rescheduling and train operation control loops

6.3.1 Time synchronisation problem

The interaction of the inner and outer control loops leads to several problems. In the following, the problems will be presented together with suggested solutions.

Precise production and interaction of different control loops requires implicitly accurate and correct time information. The time taken by all players involved must be synchronised. Nowadays, clocks are used in trains, traffic management centres, stations etc. Time comparisons in Switzerland showed that differences of up to 10 seconds can be observed between clocks used by different players even if calibration is carried out regularly.

This inaccuracy and variations mean that larger tolerance bandwidths may have to be selected and wrong decisions may be taken by the rescheduling system. Thus, the possible performance of integrated real-time rescheduling systems can be restricted if times cannot be synchronisation. Also manual time calibration must be avoided whenever possible. Satellite navigation systems or radio controlled clocks have the highest accuracy (average time error for GPS systems is below 40ns [HW08]). Based upon this, we recommend that it should therefore be used by all players.

6.3.2 Differing loop frequencies problem

Time synchronisation is a necessary prerequisite for an integrated real-time rescheduling system. However, this does not imply that all rescheduling and train control loops are also synchronous. Figure 6.2 shows, that in the outer rescheduling and the inner precise train control loop, two different time delays and two different loop frequencies can occur.

![Figure 6.2: Differing loop frequencies for inner and outer control loop](image)

Short sample times are required for a high production accuracy (see chapter 5). Therefore, no interaction effects will appear as long as the train operation control loop frequency is substantially greater than the frequency of the rescheduling loop. The frequent display update required ensures that the inner precise operation control loop has a sufficiently high frequency.
6.3.3 Concurrence of rescheduling and train operation loops through the design of the tolerance band

6.3.3.1 Overview of design possibilities for the tolerance band

The decision if a continuous tolerance band or only a time slot at reference points are transmitted to the train has a crucial impact on the design and effectiveness of the rescheduling loop and DMI (see section 5.2.2). Using a tolerance band results in exact trajectories having to be calculated centrally for all trains and having to be submitted to all trains involved after rescheduling. On the other hand, applying a tolerance band results in deviations possibly being detected as early as possible and trajectory conflicts can be avoided due to the central generation of the production plans (see also section 4.3.1.1). Particularly for series of trains, interferences and detractions may result in knock-on delays from one train to another being drastically increased. An exemplary headway conflict for two consecutive trains in the case of non-continuous reference trajectory and missing tolerance bandwidth is illustrated in Figure 6.3. Therefore, transmission of continuous trajectories and tolerance bands are preferable.

![Figure 6.3: Comparison of continuous tolerance band (left) to tolerance slot at reference points (right) causing possible headway conflicts (red)](image)

Various strategies for defining the size of the tolerance bandwidth are imaginable. An initial strategy would be to assign a constant tolerance band to all trains. However, it is possible that some trains are more difficult to handle and operate precisely. Consequently, an adjusted tolerance bandwidth might be better to minimise unused tolerances and assure rescheduling stability coinstantaneously. Possible examples of trains difficult to operate are trains following a disturbance or several freight trains.

The data analysis described in chapter 5 showed that variations in the dwell process are significantly larger than for running trains. Thus, adjustments in the size of the tolerance bandwidth can take this aspect into account in a way that stopping trains may be assigned a larger tolerance bandwidth. However, not only the size, but also a change in size over time is possible. If possible (running time supplements allows speed-up and no other trains are influenced), the size of the tolerance band may be large as long as trains are further away from capacity...
bottleneck areas and may narrow the closer the train comes to the capacity bottleneck area. Constriction may thus be continuous (preferable for running trains) or gradually (for trains at stations) as illustrated in Figure 6.4. However, the result of changeable sizes in the tolerance bandwidth is that in addition to the reference trajectories, tolerance bands also have to be calculated and transmitted and includes train drivers having to be informed about their tolerance bandwidth (visualisation on the DMI). However, it is important that the size of the tolerance bands is minimised in the capacity critical bottleneck areas (condensation areas).

For sequences of equal trains each following the other, changeable sizes in the tolerance bandwidth are inappropriate. Capacity and performance are then limited by the maximum and not the minimum size of the tolerance bandwidth. To summarise, a constant tolerance bandwidth is easier to manage. However, specifically adjusted (larger) tolerance bandwidths may be assigned to trains with special conditions (e.g. after an event or no DMI) in order to ensure stable production.

6.3.3.2 The rescheduling stability performance contradiction

The largest variations and uncertainties appear after a deviation or event. Behaviour varies for each single, specific case. The two-step prediction of future behaviour (see also chapter 4.3.3.2) is thus of high importance. Optimistic predictions, resulting in no or only limited reductions compared to the original schedule, enable the best possible performance because the largest amount of schedule solutions are available. More defensive predictions, resulting in subsequently predicted earliest passing times at reference points, reduce the number of rescheduling measures and consequently, performance in these rescheduling cases is also reduced (see also Figure 6.5 for a visualisation of the different prediction strategies).
Contradictory to defensive predictions optimistic predictions involve a higher probability of not being maintained. As a consequence, the rescheduling procedure will be initiated more often for optimistic predictions resulting in a nervous and less stable rescheduling system. Many changes in the operation behaviour due to frequent rescheduling unnecessarily disturb the players and may lead to the problem that new production plans will be disregarded by them. Consequently, a trade-off between maximum performance and high stability is unavoidable. The conflict of objectives may at least be reduced by an accurate database for prediction including precise and regular data from trains.

Consequently, besides the handling of the prediction strategy, the management and size of the tolerance bandwidth also have a significant impact on rescheduling stability. Nervous rescheduling can, for example, be avoided by assigning larger tolerance bandwidths after rescheduling to trains with a deviation or affected by an event. Uncertainty during prediction and consequently excesses of the threshold resulting in new rescheduling may thus be reduced. In a following second step, adjustments to the production plan through rescheduling may be made. However, a larger tolerance bandwidth results in the planned headway between two trains being larger and thus, capacity usage and finally the performance are reduced.

### 6.3.4 Unscheduled state problem

The time period when a train exceeds a threshold (the tolerance bandwidth) until a new, feasible production plan is generated and transmitted to all players leads to an unscheduled phase. The behaviour of the rescheduling initiating train may cause other trains to be influenced and affected by knock-on delays before the new production plan is available. Consequently, the interaction of the inner and the outer feedback controller must be harmonised based on given rules.
6.3. Interaction of rescheduling and train operation control loops

Figure 6.6 illustrates the case of two following trains where the first train is delayed. Two predictions are possible in this case:

- the first train will not lose further time or only a limited degree and will therefore not affect the second train; or

- the first train will continue to be delayed even more and thus cause a headway conflict with the second train which will have to slow down.

Assuming a wrong prediction therefore leads directly to a new rescheduling process.

The main problem of the unscheduled state is that drivers of the second train may unintentionally approach a closed signal, which in a scheduled state should never occur. This will not only cause further knock-on delays, also prediction of future train run is more difficult and erroneous due to the non-linear behaviour (see section 6.4.2.2 below for more details).

For departing trains, where exact timing is needed as described in chapter 5.3.2, the unscheduled state may result in a train closing and locking the doors regardless of the fact that the route will not be set immediately and the train must wait for a longer time. This leads to passengers waiting in front of a closed train, which annoys them. In a worst case, passengers may try to open the doors by brute force or with the aid of an emergency switch.

During the unscheduled state, other trains may be affected in the event that the scheduled buffer time between the rescheduling initiating train and the other trains is small. This is the case for following trains in a sequence as well as for joining or crossing trains on open track or in station areas. The possibility of immediate impacts on other trains increases for longer rescheduling calculation durations. An interruptible event-driven rescheduling policy must therefore be avoided. Also, two-step algorithms with quick initial measures followed by detailed calculations may be used to reduce the number of trains approaching closed signals unintentionally.
6.3.5 Summary of rescheduling and precise operation loop interconnections

Several properties as part of the control loops in the new integrated real-time rescheduling framework impact each other. The dependencies were introduced in the previous sections and chapters. The interconnections are summarised below. The relevant properties are:

- rescheduling performance;
- rescheduling time;
- rescheduling nervousness;
- rescheduling policy;
- size of the tolerance bandwidth;
- optimistic/defensive prediction; and
- transmission of non-feasible schedules.

All the interconnections between the properties are illustrated in Figure 6.7. To find a balance between all the interconnections, an optimisation target must firstly be defined. This strategic task has to be followed, secondly, by extensive evaluations based on test runs and surveys.

Figure 6.7: Overview of interconnections in the rescheduling and precise operation control loop
6.4 Optimisation potential with integrated real-time rescheduling

6.4.1 Overview of optimisation opportunities

The potential benefits of the new integrated real-time rescheduling framework will be derived and visualised in the following. The new integrated real-time rescheduling concept offers optimisation potential in different fields:

- Traffic flow: unnecessary and time-consuming stops or influences by signals in bottleneck areas and on open tracks are minimised. Two goals can thus be achieved:
  - In combination with advanced real-time rescheduling algorithms, the amount of knock-on delays and delay propagation can be minimised and the network will be operated more stably.
  - Energy consuming acceleration can be reduced if unintended slow-downs are avoided. Similarly, wear of track and wheels are reduced.

- Buffer times: a combination of integrated real-time rescheduling with smart network decomposition allows buffer times in bottleneck areas to be minimised. Consequently, relevant capacity limited areas can be saturated for maximum usage.

- Further benefits: changes in operation principles offer potential improvements in other areas such as, for example, for customer information, or last-minute slot-selling.

The different optimisation possibilities are explained below in detail.

6.4.2 Traffic flow optimisation

6.4.2.1 Principle of avoiding unnecessary signal influences

In conventional rail operations, train drivers have no direct information besides the initial schedule, knowledge of train behaviour, actual speed, and signal aspects. In general, drivers do not know if or how they should adjust their driving behaviour (speed) to prevent conflicts or improve traffic flow. This means that they may inadvertently cause additional delays due to their slowing down and speeding up after a closed signal.

Figure 6.8 illustrates schematically the topology for a train running towards a capacity bottleneck area with two time-reference points. The first time-reference point (TRP1) is located in front of the last distant signal outside the capacity critical area, and the second time-reference point (TRP2) lies within the capacity critical area.

Figure 6.9 shows the consequences of the moment a train approaches the capacity critical area in combination with a fixed, given point of time when the signal turns green in a time-distance...
Figure 6.8: Schematic topology to analyse the delay effects for trains approaching a possible conflict situation

diagram for five different TRP1 passing times. In 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 comes to a standstill in front of the closed main signal whereas in cases B and C, trains do not completely stop. In cases D and E, the train passes the distant signal indicating a movement authority allowing them to proceed with their maximum permitted track speed. Figure 6.9 shows that in case D, where the train passes the first reference point late, the earliest passing time at the second reference point is measurable whereas, in case A, the earliest passing of the first reference point, has a later passing time than most other cases due to the standstill. The target of the new integrated real-time rescheduling framework is to guide trains in a way that they are able to pass the critical section as closely as possible to case D, but not any earlier.

6.4.2.2 Temporal effects due to signal influences

Trains that have to slow down or fully stop because of a route conflict (delayed previous train), lose additional time in conventional railway operations because of braking and reacceleration. Assuming that the order of trains is fixed, the resulting delay for a second train $t_{final-detrain2}$ can be expressed as:

$$ t_{final-detrain2} = \begin{cases} 
  t_{init-detrain2} & \text{for } t_{init-detrain2} > t_{init-detrain1} - t_{buffer} \\
  t_{init-detrain1} - t_{buffer} + t_{add-detrain2} & \text{for } t_{init-detrain2} \leq t_{init-detrain1} - t_{buffer} 
\end{cases} $$

with the initial delays for the first and the second train $t_{init-detrain1}$ and $t_{init-detrain2}$, the scheduled headway buffer time between these two trains $t_{buffer}$ and the additional time lost for the second train due to braking and reacceleration $t_{add-detrain2}$. 
Consequently, the knock-on delay is, as derived and proven in details in appendix C, not only a direct transfer of the first train’s delay. The factors causing major impacts on the transfer function $t_{\text{add}} - \text{del}_{\text{train2}}$ are:

- the type of signalling system;
- signal positions;
- operation rules;
- infrastructure including desired track speed; and
- train dynamics.

Overall, a non-linear transfer function can be derived for the rescheduling delay. Figure 6.10 shows this non-linear transfer function with the influence on delay (ordinate: the moment the 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) exemplarily for a signalling system without direct reacceleration. Depending on the signalling system, 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). Calculations show that larger variations...
and discontinuities (of up to several minutes) are possible in the time-time transfer function due to braking and acceleration in particular for high-speed trains, heavy freight trains or on gradients. Therefore, the new integrated real-time rescheduling framework reduces knock-on delays for one train substantially.

6.4.2.3 Improved stability by integrated real-time rescheduling

The integrated real-time rescheduling framework allows us to minimise the number of signal influences and signal stops. Time-consuming braking and acceleration actions are therefore reduced. In particular, knock-on effects in bottleneck and station areas, where delays are passed from one train to another, can be avoided with integrated real-time rescheduling. This effect alone would improve the stability of the rail network significantly.

The tolerance band principle as part of the integrated real-time rescheduling framework helps to identify deviations at an early stage. This data is used by rescheduling algorithms in order to generate new production plans based on predefined optimisation objectives. In contrast to today’s heuristic, manual dispatching measures, all consequences are foreseeable and possible solutions can be compared. Early measures including network-wide optimisation therefore ensure that delay propagation is reduced to a minimum.

The integration of precise railway operations into the rescheduling process allows us to exploit the maximum possible solution space. Potential benefits including the stability criteria of dispatching algorithms may therefore be improved with the new integrated real-time rescheduling framework.
6.4.2.4 Energy usage minimisation

6.4.2.4.1 Energy reduction strategies

Increasing prices for resources leads us to use energy as efficiently as possible. Research on minimising the train’s traction energy subject to physical (rolling stock and infrastructure) and temporal (timetable) constraints is being carried out by various groups [Alb04, Fra02, How95, Liu03, Mey07]. Nevertheless, current state-of-the-art solutions for energy-saving driving are only effective for systems with static timetables. Changing traffic conditions and particularly delays causing trains to stop unnecessarily are not taken into consideration in these models. Since integrated real-time rescheduling reduces unnecessary slowing down or stopping, it could also reduce energy consumption. Four different strategies to save energy are possible through integrated real-time rescheduling [Lüt08]:

- reducing the number of unnecessary stops or influences by signals;

- reducing maximum speeds by using running time supplements;

- reducing variations (unnecessary consumption) due to driver behaviour by supporting them with tools; and

- applying additional energy optimisation potentials such as electric braking (avoidance of pneumatic brakes), or coasting by using infrastructure data such as the position of tunnels and gradients. These potential savings can be applied by additional algorithms and visualised in the driver’s supporting tool.

Running time has a significant impact on energy consumption. Running time supplements could be used to extend the travel time. Thus, trains arrive at a precisely specified time at a given reference point and traction energy is reduced. Longer travel times could be obtained by:

- coasting;

- reducing maximum speed; and

- lower grades of acceleration and deceleration (earlier braking and use of electric, regenerative brakes).

Research has shown that initiating early braking allows us to use (almost) exclusively electric brakes instead of late and heavier braking with both mechanical and electric brakes together. Early braking enlarges overall running time slightly (several seconds), however energy consumption can be reduced by up to 20 percent [Mey00, Mey07]. Consequently, if a signal stop is unavoidable, integrated real-time rescheduling helps to inform drivers in advance in order to initiate early braking, which results in less energy loss.
6.4.2.4.2 Energy reduction calculation

The traction force of a train is the sum of electrical forces for acceleration and regenerative braking and mechanical forces for braking. Additional energy consumption for supporting systems, such as, for example, lighting, heating or air conditioner have been ignored. The total electrical energy consumed for traction $E_{el,\text{total}}$ over a given time period (e.g. day) in a selected network and a given number of trains ($n_{\text{trains}}$) with their specific ride between $s_0$ and $s_e$ is determined [Fra02] by:

$$E_{el,\text{total}} = \sum_{j=1}^{n_{\text{trains}}} \int_{s_0}^{s_e} \{F_{el,j}(x)v_j(x) + P_{\text{loss}}(F_{el,j}, v_j)\} \, dx$$  \hspace{1cm} (6.1)

with the train’s actual speed $v_j$, electrical tractive force $F_{el,j}$ and propulsion loss $P_{\text{loss}}$. However, for a first approximation, a constant efficiency factor $\eta_j$ (which is at around 85% for nominal power) could be assumed for each train. Consequently, energy usage for one train can be approximated by:

$$E_{el,\text{total}} = \int_{s_0}^{s_e} \frac{1}{\eta_j} F_{el}(x)v(x) \, dx$$  \hspace{1cm} (6.2)

The greatest energy savings through integrated real-time rescheduling can be expected by reducing the number of unnecessary signal influences. A driver recognising a closed signal has to use mechanical brakes in combination with electric brakes to stop the train safely in front of a closed main signal. Therefore, only a minor part of the energy can be used for recuperation. The amount of energy lost due to unnecessary braking is also heavily dependent on initial and target speeds, reduced speeds (or standstills in worst case), train types, train weight and track characteristics. The total amount of possible energy savings is the difference between the energy used for braking and reacceleration minus the energy used to pass the same track section.

A derivation and an overview of potential energy savings for different train types and speed states is in appendix D. It is shown, that rescheduling, and in particular the point of time when the rescheduling is executed, has a significant impact on energy consumption (see also Figures 6.11 and 6.12). It also shows that having one avoidable signal stop for an Intercity train between Rotkreuz and Lucerne (18.5 kilometres distance) results in an increase in additional relative energy consumption of 50%.

To estimate total energy savings through integrated real-time rescheduling in an entire network over a given time period (e.g. day, year), the effects of minimising train stops, avoiding usage of pneumatic brakes and longer travel times must be combined. On track sections with capacity bottlenecks, many trains are forced to make one or more unscheduled stops. For example up to 40% of all Intercity trains from Zug to Lucerne have to stop at least once before enter-
6.4. Optimisation potential with integrated real-time rescheduling

Figure 6.11: Influence of driving strategies and rescheduling point of time on energy consumption based on an Intercity train from Rotkreuz to Lucerne

Figure 6.12: Influence of the rescheduling point of time on energy consumption based on an Intercity train from Rotkreuz to Lucerne without coasting

Consequently, a new integrated real-time rescheduling framework offers a significant potential for optimising energy consumption by both avoiding unnecessary halts and intelligent usage of running time supplements by the generation of reference trajectories. An initial analysis and extrapolations showed that 5 - 10 % of overall traction energy may be saved with integrated real-time rescheduling for the SBB network.
6.4.3 Capacity improvements with integrated real-time rescheduling and network decomposition

6.4.3.1 Buffer time reduction strategies

Surveys have shown (see section 2.3.4.3) that with today’s railway operation methods, the distribution of arriving and departing times of trains is in most cases log-normal or log-logistic. Conflicts and thus knock-on delays occur when the delays of two consecutive trains using the same infrastructure overlap (Figure 6.13). Conflicts are minimised if either buffer times between two consecutive trains are enlarged (which would contradict the intention to improve capacity) or the width of delay distribution is reduced.

![Figure 6.13: Sample delay distribution for two consecutive trains in today’s operation](image)

The scheduled headway between two consecutive trains $t_{\text{head-plan}}$ can be expressed as:

$$t_{\text{head-plan}} = t_{\text{head-min-tech}} + t_{\text{buffer}} + t_{\text{unused}}$$

with $t_{\text{head-min-tech}}$ as the minimum technical headway, the planned buffer times $t_{\text{buffer}}$, and $t_{\text{unused}}$ as the time-span in which no train and buffer are planned. The unused (or unplanned) time period could be used as a planned buffer for maintenance or as an additional buffer in the event of a delay.

Introducing the integrated real-time rescheduling framework implies the application of the tolerance band principle. Instead of (large) buffer times, the size of the tolerance bandwidth can be added to the minimum technical headways. The amount of unused times may thus also be reduced and buffer times can be allocated more systematically and traceably. However, not all signal stops can be avoided and maintaining the tolerance band may not either be possible for all trains. Consequently, small unused time spans will still be required for such eventualities.

6.4.3.2 Determination of possible capacity improvements in bottleneck areas

Various methods exist to determine the number of trains running over a given infrastructure during a defined time period. One typical approach is to condense train movements over a
6.4. Optimisation potential with integrated real-time rescheduling

certain time [Bur05, UIC04]. The different parts of capacity usage (occupation times, buffer times, maintenance buffers and unused capacity) are identified. Instead of non-systematically allocated buffer times, well-defined tolerance bandwidths are used with integrated real-time rescheduling.

Applying the condensation and compensation network division principle allows us to saturate specific bottleneck areas. The compensation areas are used to control trains in order to enter the condensation area at the specified point of time at the correct speed. Buffer times beside the tolerance band may thus be reduced to a minimum in condensation areas. Nevertheless, some eventualities (e.g. due to signal halts and subsequent time-consuming accelerations) have to be taken into account and additional buffers times are required.

The relative capacity gain \( \text{cap}_{\text{gain}} \) through integrated real-time rescheduling over a given observed time period \( t_{\text{observed\,total}} \) with \( n \) scheduled trains is determined as:

\[
\text{cap}_{\text{gain}} = \frac{\sum_{j=1}^{n} t_{\text{buf\,orig\,j}} - \sum_{j=1}^{n} t_{\text{buf\,irtr\,j}}}{\sum_{j=1}^{n} t_{\text{occ\,infra\,j}} + \sum_{j=1}^{n} t_{\text{buf\,maint\,j}} + t_{\text{cap\,unused}}} = \frac{\sum_{j=1}^{n} t_{\text{buf\,orig\,j}} - \sum_{j=1}^{n} t_{\text{buf\,irtr\,j}}}{t_{\text{observed\,total}}}
\]

with the duration of a train occupying the infrastructure \( t_{\text{occ\,infra\,j}} \), the buffer time for maintenance for each train \( t_{\text{buf\,maint\,j}} \) and the total duration of unused capacity over the observed period \( t_{\text{cap\,unused}} \). The three parts \( t_{\text{occ\,infra\,j}} \), \( t_{\text{buf\,maint\,j}} \) and \( t_{\text{cap\,unused}} \) remain identical for both conventional and integrated real-time rescheduling cases (see also Figure 6.14).

The possible relative capacity gain is therefore high for areas with small headways and large scheduled buffer times. This is nowadays the case for station areas, where buffer times are added to reduce the impact of unnecessarily stopped trains or departing delays. It is precisely these buffer times that can be significantly reduced by integrated real-time rescheduling and replaced by tolerance bands. The achievable capacity improvement is therefore heavily dependent on actual planning principles and the achievable production accuracy, which determine the tolerance bandwidth.

Figure 6.15 illustrates a conventional and integrated real-time rescheduling approach for planning and operation effects. The figure shows that planned buffer times between two trains are smaller for integrated real-time rescheduling and are equal to their tolerance bandwidth. In the event of a delay (temporal deviation) during operation, conventionally operated trains must slow down in the bottleneck area. In an integrated real-time rescheduling case, trains are slowed down intentionally before entering the capacity critical condensation area or depart later but without being hindered. Consequently, knock-on delays are reduced with integrated real-time rescheduling even if smaller buffer times are applied.
Figure 6.14: Comparison of capacity usage for conventional (left) and integrated real-time rescheduling operations (right)

Figure 6.15: Comparison of planning approaches and operational (delay propagation) effects in the event of a train arriving late in a capacity critical condensation area for conventional and integrated real-time rescheduling operations

6.4.4 Further optimisation opportunities through integrated real-time rescheduling

Apart from efficiency, capacity and stability problems, which are the main focus of this thesis, railways face with many more challenges. Some of them are closely related to the operational principle and its technology applied. An initial field for improvements concerns the distribution and sale of free slots (for short term). Compared to road traffic, free slots are often handled
inflexibly and thus limit the efficiency of freight train operating companies as well as the railway infrastructure. The new integrated real-time rescheduling framework will also tackle this area by using real-time rescheduling algorithms to find free slots, which can be reserved shortly before being used.

Automated real-time rescheduling based on predefined and harmonised optimisation objectives helps to minimise possible discrimination conflicts. In addition, decisions with integrated real-time rescheduling are, in contrast to heuristic dispatching solutions, more comprehensible and traceable. This fact is gaining in importance as the number of rail operators involved increases. Integrated real-time rescheduling reduces the number of unnecessary signal stops. This not only reduces energy consumption, but wear can also be reduced. In particular, the number of strong braking actions due to closed signals causing the greatest wear can be minimised.

More precise production also allows for more precise information. This could be used to inform passengers about their (missed) connections and alternative options. It will thus be possible to offer passengers tailor-made information concerning their trip automatically (e.g. by mobile devices). This information is not limited to passenger traffic. Freight goods can be predicted better and more accurately, which will strengthen the position of rail freight transport significantly.

6.5 Human players in the integrated real-time rescheduling framework

6.5.1 Challenges and new duties for players in the integrated real-time rescheduling framework

Technical challenges as well as new processes and methods have been in the focus of this thesis. However, railway operations are based on the interactions and tasks of many persons. Adjustments in the train operation process with integrated real-time rescheduling gives all the players involved new tasks and responsibilities. Consequently, requirements and functions are also modified, which may result in new job descriptions. Good collaboration between the players involved with the new systems and processes is a fundamental prerequisite for successful migration and application. Persons in charge have to be responsive to the fears and denegation of the players involved which is a significant challenge. Players undermining or refusing new operation process will cause the project to collapse. Otherwise, if one succeeds in motivating the players, cooperation and supporting assistance in development and realisation is assured. The new duties and responsibilities with integrated real-time rescheduling of the players involved are described below.
6.5.2 Players affected by the integrated real-time rescheduling framework

6.5.2.1 Train drivers in the integrated real-time rescheduling framework

Besides respecting safety-related signalling aspects, train drivers are used to a fair amount of freedom of how trains may be operated nowadays. Of course, standards in terms of regulations as well as instructions describe how trains should be operated. Nevertheless, measurements have shown, that there are major variations. Giving a temporal limit in which trains should remain as part of the integrated real-time rescheduling framework is thus a major restriction of their freedom. It also requires high attention to ensure that trajectories are followed accurately.

On the other hand, trajectories calculated and transmitted to trains minimise the number of closed signals approached. Unnecessary braking actions before of closed signals or even passing closed signals may thus be significantly reduced thereby improving safety. To minimise the preparation workload at the beginning of a trip, the setup and handling of the DMI must be simple and intuitive.

6.5.2.2 Guards in the integrated real-time rescheduling framework

The precise departure of trains in bottleneck areas is an important element in the integrated real-time rescheduling framework. To assure precise departures, special attention by the guards will be required. Supporting elements as handhelds for guards with exact time information or fixed installations on platforms will simplify the task. The integrated real-time rescheduling framework also results in an improved volume and quality of information which can be used by guards to inform passengers with more details and greater reliability.

6.5.2.3 Dispatchers in the integrated real-time rescheduling framework

Today, dispatchers take and communicate measures based mainly on heuristic decisions. With integrated real-time rescheduling, more dispatching actions have to be taken because deviations are detected earlier and thus measurements are taken more in advance. However, support tools including powerful rescheduling algorithms will help dispatchers to take their decisions. Automated information transfers to drivers, guards and movement inspectors reduce the required communication workload for dispatchers as well. However, it is important, that dispatchers realise that they are still important in the operation process and will not be replaced by computers.

6.5.2.4 Movement inspectors in the integrated real-time rescheduling framework

The separation of dispatching and traffic control nowadays may be softened in future. However, measures taken by movement inspectors cannot be independent from dispatching decisions
as occasionally nowadays and must imperatively follow the given production plans from the rescheduling. Technical enhancements are under development and will enable a direct remote control from the traffic management system on the interlocking system.

6.5.2.5 Passengers in the integrated real-time rescheduling framework

Punctuality and the transfer of information will be significantly improved with integrated real-time rescheduling. Passengers are the ones who will profit most. However, accurate on-time departures in bottleneck areas will mean that late passengers running for trains will no longer be picked up. Precise departures are therefore also dependent on passengers participating and not intentionally blocking doors.

6.6 Standardisation challenge for integrated real-time rescheduling systems in Europe

Ambitious efforts are being made in the European rail sector (e.g. ERTMS) to offer and guarantee barrier-free operations between bordering countries. Developments for rescheduling and DMI systems should therefore also be regarded within an European context and not by nation. Recent DMI developments with varying focus and solutions within the Netherlands (RouteLint [Alb07]), Germany (FreeFloat for main lines [Oet08,DB 08] or ENflex-S for secondary local railways [Alb09,Gas08]) or Switzerland (FARE-PULS [Wüs08]) are therefore not meaningful, if a general standard for interoperability is missing. An extension of the ERTMS specification with operation and traffic management tasks (e.g. enhancement of the DMI functionalities) is desirable particularly to avoid multiple DMI’s but is not foreseeable in the near future.

Consequently, specific standards are necessary. As is the case with ETCS, various levels can be developed. Level 0 represents the actual state with no automated data exchange between the train and the traffic management centre (visualised in Figure 6.16).

Level 1, illustrated in Figure 6.17, would require a specification for down-link data transfer from the traffic management system to the train. This would also implicitly require a decision if entire trajectories (and thus no on-board calculations) or only specific reference times for given points (with on-board calculation and optimisation of the trajectory) is to be transmitted and also if additional information (e.g. occupation state of sections ahead) is given by the traffic management system to the DMI. Also, a synchronised time using GPS or radio clock would be required at the first level.

At a second level, see Figure 6.18 the up-link from the train to the traffic management system can be defined. At this level, not only position and speed information but also coded event data or notifications from the driver or the train may be automated. The entire rescheduling process
can thus be speeded-up and misunderstandings due to linguistic problems may be reduced with coded messages.

For both levels 1 and 2, the layout and design of the DMI would remain arbitrary, as long as the accuracy achieved during operations satisfies the quality requirements and safety restrictions desired. Consequently, national regulatory authorities are requested to define consistent safety requirements and guidelines for DMIs as part of the integrated real-time rescheduling developments.
6.7 Conclusions

A detailed analysis and model of the integrated real-time rescheduling framework for railway operations including all data flows and processes has been developed. The fundamental inter-relations were visualised and could be verified with it.

The main advantage of the integrated real-time rescheduling framework is the combination of real-time rescheduling including the generation of conflict-free trajectories and the possibility of accurately following these dynamically changeable reference set-points. Unnecessary time, energy, and capacity-consuming signal stops or influences can be reduced. It was demonstrated that avoiding one signal stop by integrated real-time rescheduling may reduce delays on this train by more than one minute. Consequently, avoiding a multiplicity of signal stops, in particular in the conflicting bottleneck areas, results in savings on delays being even significantly higher, thus resulting in improved stability.

The integrated real-time rescheduling framework allows deviations or potential conflicts to be detected and solved at an early stage. Rescheduling algorithms based on predefined optimisation objectives instead of today's heuristic, manual dispatching measures allow all the consequences to be identified. Rescheduling-based optimisation as part of the integrated real-time rescheduling framework therefore ensures that the propagation of delays is reduced to a minimum resulting which also results in more stable production.

However, the impacts and effects of the integrated real-time rescheduling need to be analysed by means of simulations and practical tests. Also, the achievable improvements during daily operations needs to be determined in a long-term survey.
It was demonstrated that the size of the tolerance bandwidth is a key-element in the integrated real-time rescheduling framework. The tolerance band defines interactions between the inner operations feedback control loop and the outer, superimposed rescheduling loop. The tolerance band is used by players to react to variations and small disturbances and provides space to complete their tasks. An inadequately large tolerance band would thus result in frequent exceeding of the tolerance band and consequently nervous rescheduling behaviour. However, the size of the tolerance band directly reduces possible capacity. Extensive bandwidth must therefore be avoided implicitly.

Instead of fairly imprecise, non-systematic buffer times as allocated today, systematic tolerance bands are used in combination with blocking time theory for scheduling in the integrated real-time rescheduling framework. Thereby, several strategies for the designing of the tolerance bandwidth are possible. It is demonstrated that special care is required after a deviation or an event. Enlarging the size of the tolerance band or defensive prediction thus help to avoid nervous rescheduling.

In combination with smart network decomposition in condensation areas as well as compensation areas, it is shown that an integrated real-time rescheduling framework allows the systematical saturation of selective capacity bottlenecks. Trains on the feeder lines towards the bottleneck areas are assigned specific (dynamically changeable) trajectories and are controlled in order to get trains passing the condensation area without unnecessary time- and capacity-consuming stops. The resulting possible improvement in capacity depends on the achievable accuracy and the tolerance bandwidth assigned as well as the actual size of buffer times.

However, the principle of condensation and compensation areas in combination with integrated real-time rescheduling ensures that planned latitude (by running time supplements and buffer times) in compensation areas is specifically used to make sure that trains enter condensation areas with minimum variations and at defined point of time and precisely specified speed. This approach exclusively allows planned buffer times to be minimised and, consequently, capacity usage of the bottleneck areas to be maximised.
Chapter 7

Case studies

7.1 Introduction

Two kind of tests were done to evaluate and analyse the added value and benefits as well as the limits and challenges of the integrated real-time rescheduling system: simulation tests and test runs with different set-ups and technical components.

In a first step, micro-simulation was used to analyse the maximum possible performance and limits of the integrated real-time rescheduling system. The main factors that influence the system’s performance and their effects were identified. Fundamental relations can be illustrated.

In a second step, various types of test runs were executed to evaluate the system and to identify the technical problems and challenges. The proof-of-concept phase was also used to identify the practical limits of the integrated real-time rescheduling application. Focus was not only put on technical problems and solutions. In particular, the actors involved (drivers, guards, movement inspectors, dispatchers) were monitored specially to identify weak points and to improve the system.

The case studies focussed on passenger trains. Operational behaviour, in particular, the controllability of freight trains can therefore not yet be predicted. However, due to their low dynamics and heavy weights, freight trains offer major optimisation potential for reducing both secondary delays as well as energy consumption.

Simulation evaluations were done for areas around Lucerne, Bern and Winterthur. Initial test runs with SBB’s FARE-DMI were done between Zurich and Chur. Extended test runs were executed in the area around Lucerne in order to utilise and transfer insights and cognitions from simulations to real operations. For example, rescheduling decisions were transferred from pre-calculated simulation tests to online applications. Simulation results can also be used to compare real behaviour with the theoretically maximum possible performance as well as to determine deviations and possible reasons.
Both, simulation studies as well as test-runs, focus on improving rail traffic flow and reducing delays within a given area. Network wide delay propagation effects and strategies for holding or breaking connections were thus not analysed. This meant that coordination impacts with neighbouring condensation areas could be ignored. Large delays and big disturbances or incidents were not the main focus of research because handling these events is different to small deviations. In particular, for trains with larger delays, this information is normally available in advance in contrast to an upcoming minor deviation. Also, if a longer delay arises due to an incident or disturbance, the time during which the delay is generated could be used to predict and calculate new solutions. An overview of the case studies and the areas is given in Table 7.1 and Figure 7.1.

Table 7.1: Overview of case studies and areas

<table>
<thead>
<tr>
<th>Area</th>
<th>Characteristics</th>
<th>Studies evaluated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lucerne</td>
<td>Feeder lines merge on two heavily used tracks connecting the dead-end station</td>
<td>Simulation studies</td>
</tr>
<tr>
<td></td>
<td>Multiplicity of single line sections on feeder lines</td>
<td>FARE-DMI test runs</td>
</tr>
<tr>
<td></td>
<td>Limited impact on network stability</td>
<td></td>
</tr>
<tr>
<td>Bern</td>
<td>Main hub with integrated fixed-interval timetable</td>
<td>Simulation studies</td>
</tr>
<tr>
<td></td>
<td>Major station for important long-distance Intercity trains</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High infrastructure usage especially in eastern part</td>
<td></td>
</tr>
<tr>
<td>Winterthur</td>
<td>Station with integrated fixed-interval timetable</td>
<td>Simulation studies</td>
</tr>
<tr>
<td></td>
<td>Major bottleneck in the western part towards Zurich</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Few interactions in the eastern part</td>
<td></td>
</tr>
<tr>
<td>Chur - Zurich</td>
<td>Line with sufficient running time supplements</td>
<td>Initial test runs with</td>
</tr>
<tr>
<td></td>
<td>Intercity trains have few interactions with other trains except in Zurich area</td>
<td>FARE-DMI</td>
</tr>
</tbody>
</table>

The studies were used to investigate the interaction effects between the two superimposed control-loops. Important requirements for both control loops could also be derived based on the insights gained in the studies.

Firstly, the different areas are described in the following. Thereafter, the simulation studies are introduced and the results evaluated. Finally, SBB’s test runs with the new system are presented and analysed.
7.2 Test areas

7.2.1 Lucerne area

Both, case studies and test runs with specific FARE-DMI equipped trains were carried out on the SBB network approximately 15-25 kilometres around Lucerne, located in the centre of Switzerland. Figure 7.2 illustrates the topology of the area observed. This area was selected by the SBB for research on rescheduling methods and adjusting railway operation processes since it is a critical bottleneck with one of the most heavily used double track sections in the Swiss rail network. However, trains to and from Lucerne only have a minor impact on the total network stability. The critical bottleneck area in Lucerne extends over about 4 kilometres and links standard gauge trains heading in 5 main directions.

Infrastructure extensions are almost impossible within this area. In particular, the single line section between Fluhmühle and Rotsee will remain for a lengthy period of time. However, there is increasing demand in the surrounding area of Lucerne for additional services. Therefore, Lucerne is an excellent area to test the possibilities and limits of the new systems.
Four switch areas are along the condensation area in Lucerne: Fluhmühle (FMUE), Gütsch (GTS), Heimberg (HEIM) and Vorbahnhof (VBHF). To increase infrastructure usage, all tracks can be used in both directions. Especially in the event of a delay, using alternative routes by changing tracks in one of the switch areas is a useful measure to reduce delay propagation. The station has 10 tracks for standard gauge passenger trains, which are connected to the national network by only 2 tracks.

In Lucerne station, regular off-peak hours accommodate 30 standard gauge passenger train services per hour in a terminal dead-end station configuration. During peak-hours, additional train services are offered. Some freight trains are also operated in the Lucerne area. Lucerne is thus not an integrated fixed-interval station, instead trains arrive and depart distributed throughout every hour. Train intervals in the periodic timetable are either 30 or 60 minutes depending on the lines. Table 7.2 presents an overview of the timetable of the busier off-peak half-hour. The analysis included shunting movements, but neglected the narrow gauge trains serving Lucerne since they do not significantly impact the standard gauge network tracks.

A conventional fixed block track signalling system (mainly ZUB only the fewest of which have loops and some Signum train protection systems [Exe92, Sta87]) is used in Lucerne's station area. Using partial route release groups, train headway is between 90 and 130 seconds (depending on train category and direction). The maximum speed is 40 km/h on the approximately last/first 800 m from the station's buffer stop and 80 km/h within the rest of the condensation area. On feeder lines, the maximum track speed permitted varies between 70 and 140 km/h. The basic timetable including the blocking time stairways are illustrated in Figure 7.3.
Table 7.2: Basic timetable of Lucerne used for simulation studies

<table>
<thead>
<tr>
<th>Train ID</th>
<th>Type of train</th>
<th>Departing condensation area</th>
<th>Arriving (next) condensation area</th>
<th>Platform used in Lucerne</th>
<th>Planned arrival or departure time in Lucerne</th>
</tr>
</thead>
<tbody>
<tr>
<td>2518</td>
<td>IR</td>
<td>Lucerne</td>
<td>Sursee</td>
<td>6</td>
<td>9:55</td>
</tr>
<tr>
<td>3320</td>
<td>IR</td>
<td>Lucerne</td>
<td>Wolhusen</td>
<td>5</td>
<td>9:57</td>
</tr>
<tr>
<td>21933</td>
<td>RE</td>
<td>Hochdorf</td>
<td>Lucerne</td>
<td>11</td>
<td>9:58</td>
</tr>
<tr>
<td>21938</td>
<td>RE</td>
<td>Lucerne</td>
<td>Hochdorf</td>
<td>10</td>
<td>10:00</td>
</tr>
<tr>
<td>3311</td>
<td>IR</td>
<td>Wolhusen</td>
<td>Lucerne</td>
<td>4</td>
<td>10:02</td>
</tr>
<tr>
<td>2517</td>
<td>IR</td>
<td>Sursee</td>
<td>Lucerne</td>
<td>6</td>
<td>10:04</td>
</tr>
<tr>
<td>3568</td>
<td>RX</td>
<td>Lucerne</td>
<td>Sursee</td>
<td>3</td>
<td>10:05</td>
</tr>
<tr>
<td>21135</td>
<td>RE</td>
<td>Rotkreuz</td>
<td>Lucerne</td>
<td>8</td>
<td>10:07</td>
</tr>
<tr>
<td>2328</td>
<td>IC</td>
<td>Lucerne</td>
<td>Rotkreuz</td>
<td>7</td>
<td>10:10</td>
</tr>
<tr>
<td>21393</td>
<td>RE</td>
<td>Lucerne</td>
<td>Küsnacht</td>
<td>11</td>
<td>10:12</td>
</tr>
<tr>
<td>111</td>
<td>IC</td>
<td>Sursee</td>
<td>Lucerne</td>
<td>3</td>
<td>10:12</td>
</tr>
<tr>
<td>2410</td>
<td>RX</td>
<td>Küsnacht</td>
<td>Lucerne</td>
<td>5</td>
<td>10:14</td>
</tr>
<tr>
<td>21638</td>
<td>RE</td>
<td>Lucerne</td>
<td>Wolhusen</td>
<td>9</td>
<td>10:15</td>
</tr>
<tr>
<td>21838</td>
<td>RE</td>
<td>Lucerne</td>
<td>Sursee</td>
<td>2</td>
<td>10:18</td>
</tr>
<tr>
<td>111</td>
<td>IC</td>
<td>Lucerne</td>
<td>Rotkreuz</td>
<td>3</td>
<td>10:21</td>
</tr>
<tr>
<td>90814</td>
<td>Freight</td>
<td>Lucerne</td>
<td>Rotkreuz</td>
<td>1</td>
<td>10:23</td>
</tr>
<tr>
<td>2321</td>
<td>IC</td>
<td>Rotkreuz</td>
<td>Lucerne</td>
<td>11</td>
<td>10:24</td>
</tr>
</tbody>
</table>

Figure 7.3: Time distance diagram of the planned schedule for the Lucerne condensation area
7.2.2 Bern area

Bern with 13 station tracks is one of the largest railway stations in Switzerland and a significant bottleneck in SBB’s network. The station of Bern is planned as a main hub in the Swiss integrated fixed-interval timetable and Intercity trains meet every 30 minutes.

The critical bottleneck area (visualised in Figure 7.4) extends over 4 kilometres in the west and 5 kilometres to the east of the station and trains are heading to 8 different main directions. The narrow gauge Regionalverkehr Bern-Solothurn (RBS) trains ending in Bern are operated independently of the standard gauge trains and thus not regarded in the analysis. Also freight trains, mainly passing the condensation area tangentially in the east, are not considered in the studies.

In the 2009 timetable, 40 passenger train services are operated in Bern main station per regular half hour. The aggregation of Intercity and Interregio trains at the symmetric time is clearly visible from the schedule (see Table 7.3 for the trains analysed in the simulation studies). Commuter trains have planned stops within the condensation area at stations Wankdorf (WKD) Ausserholli- gen (BNAH) and Stöckacker (BNST).

The train headway is between 85 and 110 seconds for all trains without a scheduled stop in the condensation area. The maximum permitted track speed is 40 km/h in the switching zone close to the station and varies between 80 and 125 km/h on the feeder lines.
### 7.2. Test areas

#### Table 7.3: Basic timetable of Bern used for simulation studies

<table>
<thead>
<tr>
<th>Train ID</th>
<th>Type of train</th>
<th>Point entering condensation area</th>
<th>Platform used in Bern</th>
<th>Point leaving condensation area</th>
<th>Planned arrival time in Bern</th>
<th>Planned departure time in Bern</th>
</tr>
</thead>
<tbody>
<tr>
<td>15164</td>
<td>S</td>
<td>OST</td>
<td>2</td>
<td>BNBS</td>
<td>16:43</td>
<td>16:46</td>
</tr>
<tr>
<td>15165</td>
<td>S</td>
<td>BNBS</td>
<td>3</td>
<td>OST</td>
<td>16:43</td>
<td>16:46</td>
</tr>
<tr>
<td>16164</td>
<td>S</td>
<td>-</td>
<td>12C</td>
<td>BNBS</td>
<td>-</td>
<td>16:49</td>
</tr>
<tr>
<td>15264</td>
<td>S</td>
<td>OST</td>
<td>1</td>
<td>BNBS</td>
<td>16:48</td>
<td>16:50</td>
</tr>
<tr>
<td>16465</td>
<td>S</td>
<td>BNFI</td>
<td>10</td>
<td>ZOL</td>
<td>16:48</td>
<td>16:50</td>
</tr>
<tr>
<td>3064</td>
<td>RE</td>
<td>-</td>
<td>12A</td>
<td>BNBS</td>
<td>-</td>
<td>16:54</td>
</tr>
<tr>
<td>1864</td>
<td>IR</td>
<td>ZOL</td>
<td>9</td>
<td>-</td>
<td>16:48</td>
<td>-</td>
</tr>
<tr>
<td>3230</td>
<td>IR</td>
<td>ZOL</td>
<td>10</td>
<td>-</td>
<td>16:53</td>
<td>-</td>
</tr>
<tr>
<td>982</td>
<td>IC</td>
<td>OST</td>
<td>7</td>
<td>MAT</td>
<td>16:52</td>
<td>17:04</td>
</tr>
<tr>
<td>835</td>
<td>IC</td>
<td>OST</td>
<td>4</td>
<td>MAT</td>
<td>16:54</td>
<td>17:02</td>
</tr>
<tr>
<td>2535</td>
<td>IR</td>
<td>BNBS</td>
<td>8</td>
<td>MAT</td>
<td>16:56</td>
<td>17:00</td>
</tr>
<tr>
<td>979</td>
<td>IC</td>
<td>MAT</td>
<td>5</td>
<td>OST</td>
<td>16:56</td>
<td>17:04</td>
</tr>
<tr>
<td>830</td>
<td>IC</td>
<td>MAT</td>
<td>6</td>
<td>OST</td>
<td>16:57</td>
<td>17:07</td>
</tr>
<tr>
<td>2530</td>
<td>IR</td>
<td>MAT</td>
<td>3</td>
<td>BNBS</td>
<td>17:00</td>
<td>17:04</td>
</tr>
<tr>
<td>3235</td>
<td>IR</td>
<td>-</td>
<td>10</td>
<td>ZOL</td>
<td>-</td>
<td>17:07</td>
</tr>
<tr>
<td>1867</td>
<td>IR</td>
<td>-</td>
<td>9</td>
<td>ZOL</td>
<td>-</td>
<td>17:12</td>
</tr>
<tr>
<td>15565</td>
<td>S</td>
<td>BNBZ</td>
<td>13A</td>
<td>-</td>
<td>16:52</td>
<td>-</td>
</tr>
<tr>
<td>15665</td>
<td>S</td>
<td>BNBZ</td>
<td>13C</td>
<td>-</td>
<td>16:54</td>
<td>-</td>
</tr>
<tr>
<td>15365</td>
<td>S</td>
<td>BNBZ</td>
<td>12A</td>
<td>ZOL</td>
<td>16:58</td>
<td>17:00</td>
</tr>
<tr>
<td>15364</td>
<td>S</td>
<td>ZOL</td>
<td>1</td>
<td>BNBZ</td>
<td>17:00</td>
<td>17:02</td>
</tr>
<tr>
<td>15065</td>
<td>S</td>
<td>-</td>
<td>2</td>
<td>OST</td>
<td>-</td>
<td>17:02</td>
</tr>
<tr>
<td>15666</td>
<td>S</td>
<td>-</td>
<td>12A</td>
<td>BNBZ</td>
<td>-</td>
<td>17:06</td>
</tr>
<tr>
<td>3067</td>
<td>RE</td>
<td>BNBZ</td>
<td>13A</td>
<td>-</td>
<td>17:06</td>
<td>-</td>
</tr>
<tr>
<td>15566</td>
<td>S</td>
<td>-</td>
<td>12A</td>
<td>BNBZ</td>
<td>-</td>
<td>17:12</td>
</tr>
<tr>
<td>15267</td>
<td>S</td>
<td>BNBS</td>
<td>4</td>
<td>OST</td>
<td>17:10</td>
<td>17:12</td>
</tr>
<tr>
<td>16466</td>
<td>S</td>
<td>ZOL</td>
<td>7</td>
<td>BNBZ</td>
<td>17:10</td>
<td>17:12</td>
</tr>
<tr>
<td>16167</td>
<td>S</td>
<td>BNBZ</td>
<td>13C</td>
<td>-</td>
<td>17:14</td>
<td>-</td>
</tr>
</tbody>
</table>

#### 7.2.3 Winterthur area

Winterthur is a typical mid-size SBB station with an integrated fixed-interval timetable. Around 60 passenger train movements are scheduled per regular hour in the area of Winterthur (visualised in Figure [7.5](#)). Commuter trains are predominant and operated as Intercity trains mostly in a half-hour interval. Due to the intensive usage of the S-Bahn network, connections are often broken in case of a delay. Also planned dwell times are often very short (2 minutes or less) and can thus not be used to recover from delays.
The station of Winterthur has 9 tracks and is linked with 7 different directions. Interactions between trains are limited in the eastern section. Crossings and thus possible delay propagation occur normally in the western part, where in addition the condensation area of Winterthur is connected with the line to Zurich, representing one of the major bottlenecks in the Swiss railway network. At the boarder of the condensation area, the stations Winterthur-Grüze (GWR), Oberwinterthur (OWT), Winterthur Töss (WTOE), Hettlingen (HET) and Kemptthal (KE) are used as stopping points by some commuter trains.

The condensation area of Winterthur extends 5 kilometres in the western part and around 3 kilometres to the east. The maximum permitted speed is 70 or 80 km/h through the station and 80 to 125 km/h on the feeder lines. The headway is between 75 and 135 seconds, depending on the line and train type.

<table>
<thead>
<tr>
<th>Train ID</th>
<th>Type of train</th>
<th>Point entering condensation area</th>
<th>Platform used in Winterthur</th>
<th>Point leaving condensation area</th>
<th>Planned arrival time in Winterthur</th>
<th>Planned departure time in Winterthur</th>
</tr>
</thead>
<tbody>
<tr>
<td>831</td>
<td>IC</td>
<td>KE</td>
<td>5</td>
<td>OWT</td>
<td>17:33</td>
<td>17:35</td>
</tr>
<tr>
<td>531</td>
<td>ICN</td>
<td>KE</td>
<td>4</td>
<td>WGR</td>
<td>17:35</td>
<td>17:37</td>
</tr>
<tr>
<td>8170</td>
<td>S</td>
<td>-</td>
<td>9</td>
<td>WTOE</td>
<td>-</td>
<td>17:36</td>
</tr>
<tr>
<td>18773</td>
<td>S</td>
<td>-</td>
<td>6</td>
<td>KE</td>
<td>17:36</td>
<td>17:36</td>
</tr>
<tr>
<td>20664</td>
<td>S</td>
<td>WGR</td>
<td>2</td>
<td>-</td>
<td>17:38</td>
<td>-</td>
</tr>
<tr>
<td>19269</td>
<td>S</td>
<td>KE</td>
<td>5</td>
<td>OWT</td>
<td>17:39</td>
<td>17:41</td>
</tr>
<tr>
<td>18873</td>
<td>S</td>
<td>OWT</td>
<td>8</td>
<td>KE</td>
<td>17:39</td>
<td>17:41</td>
</tr>
<tr>
<td>20564</td>
<td>S</td>
<td>WGR</td>
<td>1</td>
<td>-</td>
<td>17:42</td>
<td>-</td>
</tr>
<tr>
<td>19673</td>
<td>S</td>
<td>HET</td>
<td>6</td>
<td>KE</td>
<td>17:42</td>
<td>17:43</td>
</tr>
<tr>
<td>30199</td>
<td>F</td>
<td>KE</td>
<td>4</td>
<td>WGR</td>
<td>17:45</td>
<td>17:45</td>
</tr>
<tr>
<td>20370</td>
<td>S</td>
<td>-</td>
<td>7</td>
<td>HET</td>
<td>-</td>
<td>17:42</td>
</tr>
</tbody>
</table>
7.3 Simulation studies

7.3.1 Microscopic railway simulation

7.3.1.1 Basic requirements for microscopic railway simulations

A microscopic railway simulation tool has to model all relevant process elements (including infrastructure, rolling stock, timetable), detailed behaviour as well as their interactions accurately. The simulation tools can be combined with optimisation tasks or run independently. Two simulation tools were used in the studies: the commercial train simulation program OpenTrack for the Lucerne and Winterthur area, and a specific microscopic simulation and scheduling tool including optimisation developed by IFOR for the Bern area. The two tools are described below.

7.3.1.2 OpenTrack

OpenTrack is a synchronous, event-driven micro-simulation application that precisely models track topology and train characteristics [Hür02, Nas04a]. OpenTrack also offers the possibility of defining distribution curves and multiple simulations in combination with events and incidents in order to evaluate different scenarios.

During a simulation run, OpenTrack allows trains to be selectively controlled. For example, exact speed limits for pre-defined sections or operational behaviour like coasting or reduced acceleration factors can be set. This makes it possible to imitate the train driver using DMI information, which directly represents the inner accurate production feedback control loop.
OpenTrack consists of basic dispatching rules. For example, based on train priorities, track segments can be pre-reserved for more important trains earlier. Also predefined alternative routes may be selected by trains. However, an optimisation-based dispatching algorithm is not implemented. Nevertheless, breaking a simulation run manually also allows the operational procedure including dispatching and route setting decisions to be influenced. This functionality allows the integrated real-time rescheduling principles to be imitated. Therefore, manually calculated rescheduling measures can be implemented and simulated with OpenTrack.

7.3.1.3 IFOR simulation tool

Rescheduling for larger areas such as Bern with roughly 600 switches and dense rail traffic is very hard to be completed manually. To evaluate the performance of the new integrated real-time rescheduling framework, the resource constrained space/time multicommodity flow train scheduling tool developed by the IFOR is used (see [Cai09b] for a more detailed description). The tool can construct detailed and conflict-free train schedules by considering locally precise topologies, the corresponding safety system as well as approximated train dynamics, given connections, arriving speed windows, and train sequences. For each train, discrete time windows for arriving and departing are defined. A multi-objective function is used in order to minimise delays, broken connections and number of reroutings in comparison to a basis schedule. In contrast to OpenTrack, this tool integrates calculation of train movements and infrastructure usage with mathematical optimisation. The tool at its current stage is not as evaluated and accurate as OpenTrack and has a complex data handling. However, output comparisons has shown that the simulation with the IFOR tool is sufficiently precise for the analysis in this thesis.

7.3.2 Simulation objectives

Simulation research was used mainly to evaluate two aspects of the proposed real-time rescheduling approach:

- Firstly, focus was set on the outer rescheduling loop. The impact of the point in time when rescheduling is initiated and time taken for the rescheduling process on rescheduling performance was investigated.

- Secondly, focus was set on the inner rescheduling loop. The impact of production accuracy and its consequences on punctuality was investigated.

For both simulation analysis objectives, the first step of the simulation study was to identify the most common delay scenarios as well as typical delay patterns and statistical characteristics including distribution curves for running, departing and arriving trains in the case study areas.
This was done using detailed operational data provided by the SBB and the railway data analysis tool OpenTimeTable [Nas04b, Ull05]. Based on data analysis, single trains were assigned a delay of between 2 and 4 minutes in the first rescheduling process where the focus of our investigation was on the outer rescheduling loop. In the second case, where the impact of driving accuracy was analysed, all trains were assigned a random delay of a scenario-dependent varying size.

7.3.3 Impact of rescheduling point of time

7.3.3.1 Simulation procedure to evaluate the impact of rescheduling point of time

The first set of simulations tested the impacts of the point of time when the rescheduling process is completed. The evaluation was done by comparing the total knock-on delays as performance indicator in three cases:

- without rescheduling;
- with rescheduling (including rerouting, reordering and delaying of trains) but without the ability to speed up trains; and
- with rescheduling including the ability to speed up trains.

To evaluate the impact of the outer rescheduling loop and, in particular, the rescheduling point of time, different scenarios listed in Table 7.5, Table 7.6 and Table 7.7 were evaluated. In each scenario, a single train was assigned a small original delay that was found to be common in the analysis of the SBB operational data.

Delays of around 2 to 4 minutes occur for both trains entering the bottleneck area or leaving the station. These delays are critical because delays occurring close to the bottleneck area or when trains should be leaving the station must be detected and solved within a short time.

These delays also directly impact the decisions dispatchers make regarding whether connections should be hold or broken. The operational analysis showed that the rescheduling measures for such delays vary significantly between dispatchers.

Consequently, there is major potential for optimisation-based rescheduling systems. In contrast, delays of more than 5 minutes are usually identified earlier in the process and therefore give the dispatchers more time to react.
Table 7.5: Rescheduling scenarios to evaluate the impact of rescheduling point-of-time for the Lucerne area

<table>
<thead>
<tr>
<th>Scenario ID</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initially delayed train (ID)</td>
<td>2518</td>
<td>3311</td>
<td>3311</td>
<td>2517</td>
<td>2517</td>
<td>111</td>
<td>3320</td>
<td>21933</td>
</tr>
<tr>
<td>Direction of delayed train Dep</td>
<td>Arr</td>
<td>Arr</td>
<td>Arr</td>
<td>Arr</td>
<td>Arr</td>
<td>Arr</td>
<td>Arr</td>
<td>Arr</td>
</tr>
<tr>
<td>Initial delay [s]</td>
<td>120</td>
<td>120</td>
<td>240</td>
<td>120</td>
<td>240</td>
<td>120</td>
<td>120</td>
<td>180</td>
</tr>
</tbody>
</table>

Table 7.6: Rescheduling scenarios to evaluate the impact of rescheduling point-of-time for the Bern area

<table>
<thead>
<tr>
<th>Scenario ID</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initially delayed train (ID)</td>
<td>2535</td>
<td>15365</td>
<td>1864</td>
<td>16465</td>
<td>979</td>
<td>15264</td>
<td>15065</td>
<td></td>
</tr>
<tr>
<td>Appearance of delay Dep</td>
<td>Arr</td>
<td>Arr</td>
<td>Arr</td>
<td>Arr</td>
<td>Dep</td>
<td>Arr</td>
<td>Arr</td>
<td>Dep</td>
</tr>
<tr>
<td>Initial delay [s]</td>
<td>240</td>
<td>240</td>
<td>240</td>
<td>180</td>
<td>240</td>
<td>240</td>
<td>120</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.7: Rescheduling scenarios to evaluate the impact of rescheduling point-of-time for the Winterthur area

<table>
<thead>
<tr>
<th>Scenario ID</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initially delayed train (ID)</td>
<td>831</td>
<td>531</td>
<td>18773</td>
<td>19673</td>
<td>831</td>
</tr>
<tr>
<td>Direction of delayed train Arr</td>
<td>Arr</td>
<td>Dep</td>
<td>Arr</td>
<td>Arr</td>
<td>Arr</td>
</tr>
<tr>
<td>Initial delay [s]</td>
<td>240</td>
<td>120</td>
<td>180</td>
<td>180</td>
<td>120</td>
</tr>
</tbody>
</table>

The simulation case studies only considered the rescheduling of train operations. Thus, the following actions could be taken to address a delay:

- trains heading to the bottleneck area could be retimed (this means that trains could be operated either more slowly or quickly as originally planned in order to arrive earlier or later in comparison to the original schedule at the border of the condensation area);
- train departure times from stations could be delayed;
- trains could be rerouted locally within the condensation area; and
- trains could be reordered within the condensation area.

Station platforms were fixed and could not be changed by the rescheduling process and crew or rolling stock rescheduling was not taken into account. This was allowed because only minor delays were taken into account.
The number of trains that can be speeded up depends on the stopping patterns and the running time supplements (which for the SBB are about 10 percent in this area although they vary for all lines). This meant that Interregional and Intercity trains could be speeded up by a maximum of 1 to 2 minutes. Increasing train speed to minimise total delay is only possible up to a certain moment. Thereafter, trains can only be rerouted, reordered and/or delayed.

The simulation analysis included several constraints:

- Firstly, all actions taken to address delays were required to generate conflict-free train paths; this enabled non-linear delay effects, which would be caused by conflicts, to be ignored. Fixed discrete time slots (as with the pulsing method described in section 4.3.4.4.10) were therefore used.

- Secondly, it was assumed that trains run within their given tolerance bandwidth of 15 seconds in all cases including speeding up or slowing down.

- Thirdly, the delayed train could not be speeded up (prediction constraint).

The objective function of the simulation analysis was to minimise the total knock-on delay of all trains. If multiple solutions with identical amount of knock-on delays were possible, the one with fewest route changes is chosen. The negative values of early trains as well as shunting delays were not added to the objective function. Also, all trains were weighted with the same value.

For the Lucerne area, the rescheduling optimisation process, which corresponds to the external rescheduling loop, was modelled manually by performing repeated simulations. Due to the limited number of routing possibilities in Lucerne in combination with time discretisation, the number of possible rescheduling measures is relatively small and consequently manually executable. Daily experiences from movement inspectors were also used. Using the original schedule with the time-space discrete pulsing method for rescheduling, rerouting possibilities and conflicts are immediately obvious. Figure 7.7 illustrates the planned schedule in the pulsing model. The similarities between the pulsing model and the time-distance diagram (see Figure 7.3) can be determined for both, train movements and blocking times.

The detailed pulsing method for the Lucerne condensation area was introduced in 2006 [Roo06]. Thereby, time-discrete pulses of one minute were defined and each train is assigned two pulses. Having a train headway of, for example, 90 seconds thus means that 30 seconds of capacity are not used. Of course, these buffer times are used to absorb regular variations in the driving and departing process as long as they remain within the predefined tolerance bandwidth. Also, the high flexibility and the replaceability of each train as well as the discrete times at the border of the condensation area with the pulsing method made the entire rescheduling process manageable for initial tests.
For the Bern scenario, calculation of the new schedule with given constraints included optimisation and was executed automatically by the IFOR tool. Thereby, each train was assigned with an arrival or departure time window of 5 to 7 minutes during which the train had to enter the simulation area or depart at the station. Connections between Intercity and Interregio trains were defined to hold up to a delay of 5 minutes. To reduce the solution space, discrete time slots of 30 seconds were defined for each train. However, around 1,000 routes per train, 10 or more time slots per train, 41 trains (resulting in $10^{160}$ possible schedules), and 50 connections had to be considered in the calculations. The elapsed time during the rescheduling can be modelled with this tool by fixing routes and train sequences as well as by reducing the time windows for the trains.

7.3.3.2 Results of impacts of rescheduling point of time

The results of the first simulation study with the focus on the outer rescheduling loop are shown in Table 7.8. It shows the initial delay, the total knock-on delay without rescheduling, and the percentage reduction in knock-on delay with the application of rescheduling (rerouting, reordering and retiming), firstly with speeding up trains, secondly without speeding up trains and thirdly
with rescheduling at the point of time when the delayed train enters the condensation area or departs from the station.

The results show that propagation and number of knock-on delays can be significantly reduced with rescheduling. The simulations demonstrated, in addition, that the rescheduling point of time has a significant impact on the final overall delay.

Table 7.8: Reduction of knock-on delays with rescheduling (rerouting, reordering and speeding up trains) for a single initially delayed train for the simulation studies in Lucerne (LZ), Bern (BN) and Winterthur (W) areas

<table>
<thead>
<tr>
<th>Scenario ID</th>
<th>Total knock-on delays without rescheduling [s]</th>
<th>Reduction knock-on delays with rescheduling including speed up of trains [%]</th>
<th>Reduction knock-on delays with rescheduling but no speed up of trains [%]</th>
<th>Reduction knock-on delays with rescheduling at the point of time when the delayed train enters the condensation area or departs at the station [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>LZ 1</td>
<td>660</td>
<td>82</td>
<td>55</td>
<td>45</td>
</tr>
<tr>
<td>LZ 2</td>
<td>600</td>
<td>70</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>LZ 3</td>
<td>1500</td>
<td>72</td>
<td>72</td>
<td>56</td>
</tr>
<tr>
<td>LZ 4</td>
<td>180</td>
<td>67</td>
<td>67</td>
<td>0</td>
</tr>
<tr>
<td>LZ 5</td>
<td>960</td>
<td>94</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>LZ 6</td>
<td>240</td>
<td>100</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>LZ 7</td>
<td>540</td>
<td>89</td>
<td>57</td>
<td>44</td>
</tr>
<tr>
<td>LZ 8</td>
<td>780</td>
<td>54</td>
<td>23</td>
<td>15</td>
</tr>
<tr>
<td>BN 1</td>
<td>960</td>
<td>100</td>
<td>100</td>
<td>3</td>
</tr>
<tr>
<td>BN 2</td>
<td>300</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>BN 3</td>
<td>150</td>
<td>100</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>BN 4</td>
<td>960</td>
<td>81</td>
<td>81</td>
<td>40</td>
</tr>
<tr>
<td>BN 5</td>
<td>540</td>
<td>100</td>
<td>100</td>
<td>33</td>
</tr>
<tr>
<td>BN 6</td>
<td>390</td>
<td>50</td>
<td>0</td>
<td>43</td>
</tr>
<tr>
<td>BN 7</td>
<td>300</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>W 1</td>
<td>1920</td>
<td>67</td>
<td>67</td>
<td>67</td>
</tr>
<tr>
<td>W 2</td>
<td>120</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>W 3</td>
<td>360</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>W 4</td>
<td>240</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>W 5</td>
<td>750</td>
<td>68</td>
<td>52</td>
<td>68</td>
</tr>
</tbody>
</table>

The effects of the rescheduling point of time on the performance (total delay) are visualised for the two cases in Figure 7.8, Figure 7.9 and Figure 7.10. The gradual growth in delay with respect to elapsed time, derived in chapter 4.4, is shown for all delay scenarios. The detailed results and rescheduling measures for all scenarios listed in appendix E.1 show that the earlier the rescheduling is applied, the more measures and thus better performance are possible. In particular, speeding up trains, if possible, is a powerful method of reducing delays.
Figure 7.8: Total relative delay depending on the rescheduling point of time for all rescheduling scenarios for the Lucerne area.

Figure 7.9: Total relative delay depending on the rescheduling point of time for all rescheduling scenarios for the Bern area.

Figure 7.10: Total relative delay depending on the rescheduling point of time for all rescheduling scenarios for the Winterthur area.
The following conclusions can be drawn from the simulation results:

- The moment when exceeding a threshold 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 are reduced.

- The gradual increase in the total knock-on delay means that even small differences in elapsed time can have a significant impact on performance.

- Applying rescheduling at the point of time when the delayed train departs or enters the condensation area implies that the rescheduling measures are already limited and thus performance is substantially reduced. This highlights the importance of initiating and applying rescheduling at an early stage. However, this also increases the possibility that another incident or delay will occur in the meantime and further rescheduling will be needed.

- Speeding up trains can be an 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 to allow trains to be speeded up.

- The effectiveness of the integrated real-time rescheduling system is highly dependent 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.

- Lucerne’s specific topology where the condensation area is linked to three short single-track line sections causes some trains to intentionally slow down or even stop in the condensation area. In contrast to the original strategy of preventing trains from slowing down or stopping within condensation areas because of losing additional time, in the specific case of Lucerne this can sometimes increases the overall flow and results in a lower total knock-on delay.

- Slowing down trains heading towards the condensation area to avoid unnecessary signal stops is often applied the later the time elapses in rescheduling. Without slowing down trains before entering the condensation area would result in even further knock-on delays.

- The size of the original delay, even small differences of 2 minutes, can have a major impact on the difference in the resulting total knock-on delay for both with and without rescheduling.
• The rescheduling system must handle freight trains in a comparable way to passenger trains. However, it is difficult to accurately predict and control freight trains. In the future, adaptive identification methods and supporting tools will be needed to help operate and plan freight trains nearly as accurately as passenger trains.

• The simulation studies were carried out for only one train assigned a delay. Having more than one delayed train would initiate a new rescheduling whenever a new delay is identified. However, the gradual characteristics would remain. The size and position of the steps would thereby heavily depend on the specific circumstances.

7.3.4 Impact of production accuracy

7.3.4.1 Simulation procedure to evaluate the impact of production accuracy

The second set of simulations tested the influence of production accuracy on delays and their propagation. Production accuracy refers to how close to the exact schedule a train is operated and thus represents the performance and consequences of the inner control loop. Small time deviations from the schedule occur frequently in normal operations. In order to analyse the impact of production accuracy and the resulting delay propagation, areas with dense rail traffic and various interdependencies are relevant. The areas around Lucerne and Winterthur were used therefore. In order to demonstrate the effects of production accuracy on heavily used areas, the original timetable of the two areas was condensed by reducing the scheduled headway buffer times between two trains down to 15 seconds.

Three different scenarios with a uniform distribution width of 30, 60 and 90 seconds respectively were defined for the original delay. Each train was assigned a random delay based on the distribution at the beginning of the simulation run. For arriving trains, the random delay was therefore added to the beginning of their simulation runs (e.g. at the station before the portal). The delay for departing trains had to be added to the time after the signal turned green and not only in addition to the scheduled departure time. Departure delays are consequently the sum of the knock-on delays until the route is set and the original random delay. So, delays from daily operations are modelled more realistically (see also section 5.3.1).

Running time variations within the condensation area could be modelled although they were not assumed in the simulation runs. Trains were thus unable to speed up and regain lost time. Also, variations in the point of time at which the reference point is passed or the train departs the station are sufficient to identify the delay propagation effects due to inaccurate production. In addition, rescheduling measures were not taken since the objective was to analyse the effects of the production accuracy.

Ten simulations with different original random delays were run for each delay distribution width scenario and area. OpenTrack generates random departure delays based on predefined distri-
7.3. Simulation studies

butions. However, the departure delay is only added to the planned departure time and does not take into consideration or include late route reservations which have to be added to the final departure delay. Consequently, simulations and, in particular, the dispatch of the departure had to be completed manually in OpenTrack.

7.3.4.2 Results of the impact of production accuracy

The propagation of knock-on delays for different production accuracy is visualised in Figure 7.11 for the Lucerne area and in Figure 7.12 for the Winterthur area. It shows the average knock-on delay of all ten simulation runs for each distribution width scenario for all trains ordered by their planned arrival or departure time at the platform at Lucerne’s terminal.

![Figure 7.11: Average consecutive delay in Lucerne area for each train for different temporal delay distribution width variations](image1)

Figure 7.11: Average consecutive delay in Lucerne area for each train for different temporal delay distribution width variations

![Figure 7.12: Average consecutive delay in Winterthur for each train for different temporal delay distribution width variations](image2)

Figure 7.12: Average consecutive delay in Winterthur for each train for different temporal delay distribution width variations

The results show that for small variations (scenario with a maximum of 30 seconds), the knock-on delays are almost negligible and the timetable remains stable. For larger variations (assign-
ing trains a stochastic delay of maximum 60 seconds), a limited increase of the knock-on delay is observed. The knock-on delay for all trains is 22 seconds on average for Lucerne and 8 seconds for Winterthur. For even larger variations (scenario with maximum delay of 90 seconds), the knock-on delays increase significantly to an average value of 55 seconds in Lucerne and 19 seconds in Winterthur. Consequently, inaccurate production may result in a considerable growth of knock-on delays. Nevertheless, these simulation studies with a condensed timetable proved that the infrastructure utilisation can be improved with integrated real-time rescheduling without having more delays than today.

In addition, the simulations provided the following insights:

- In general, departing trains have larger knock-on delays in comparison to arriving trains. The reasons are that arriving trains are only assigned the random delays at their beginning whereas departing trains are, at the beginning, assigned both the random original delay and the knock-on delay until the route is free. In addition, train headway is comparatively large for the first block after the station. Incoming trains can therefore use the partial route release whereas departing trains have to wait until the entire route is released and can be set.

- The non-linear delay effects, introduced in section 6.4.2, are limited. In addition, most trains influenced by closed signals causing them to slow down arrive at the final reference point earlier than trains without a signal influence. Sample graphs for incoming trains with changing passing times of portal time are illustrated in Figure 7.13. Consequently, the propagation of knock-on delays is reduced. The reasons for the low non-linear delay effects are:
  - low permitted speed in the condensation area of Lucerne where the conflicts occur;
  - good train dynamics; and
  - signalling system (ZUB with loops and SIGNUM) allows reacceleration before the signal is passed.

- Instability is observable for the delay variation scenario with a maximum distribution width of 90 seconds. Average delays may thus be larger than origin delay. This behaviour is possible, because trains have to stop more often for large variations. The non-linear delay effects for trains slowing down only partially, resulting in little time lost, are thus less common and stopping requires more time and capacity.

- The train mix has a significant impact on the delay propagation. Areas with high share of commuter trains (e.g. Winterthur) are more stable than areas with significant amount of Intercity and freight trains.

Accurate production is a fundamental precondition for effective real-time rescheduling systems. As shown in the simulations with the focus on the rescheduling point of time, the rescheduling
measures have a significant impact on the overall delay. However, if these measures are not implemented precisely, they can result in suboptimum rescheduling. Furthermore, conflict-free operations (avoiding the non-linear delay effects causing knock-on delays) can only be achieved with accurate production. Consequently, trains running on open tracks and leaving stations have to be operated within the given tolerance bandwidth.

7.4 SBB test runs

7.4.1 SBB’s PULS 90 and FARE projects

PULS 90 (Produktorientierte Umsetzung der Leistungs-Steigerung) was a program initiated and executed by SBB AG’s infrastructure traffic management department in 2001 to improve the capacity of rail infrastructure with as less additional infrastructure as possible [Sta04b, Wüs06]. PULS 90 thereby combines planning and operating tasks to improve efficiency, quality and customer satisfaction for minimum investment.

As part of the PULS 90 program, five strategic directions of impact were identified by SBB AG on which changes for the future railway planning and operation are focussed [Lau07]:

1. unification of planning and operation methods;
2. reduction of reaction times to handle events more efficiently as well as more precise planning and production;
3. alignment of all processes to a more customer-oriented method;
4. intelligent network separation for improved and simplified dispatching and planning; and
5. change from event-driven to time-driven operations.
The integrated real-time rescheduling method developed and presented in this thesis was part of this PULS 90 program. The FARE-DMI (see section 5.2.5.7) was also part of PULS 90 considerations to close the inner train operation control loop and was developed in cooperation with the SBB’s passenger operation’s division. The introduction of the integrated real-time rescheduling loop with the DMI was one of the key elements of the PULS 90 program.

Of course, during the proof-of-concept phase, many processes were not under full operation and had to be completed manually. For example, rescheduling tasks were executed manually at the traffic control center with a specifically developed tool based on the time-space discretised pulsing method and pre-calculated solutions from earlier experiences and data-analysis. Changed routes were communicated verbally directly to the movement inspectors and adjusted trajectories were sent to the FARE-DMI on the trains initiated by a human event.

7.4.2 Overview of SBB’s test runs

The SBB executed three types of test runs with different targets and varying set-ups. Thereby, the proof-of-concept test-runs had to be executed on dedicated trains on specific track sections or selected areas in order to minimise the impacts on other trains. The test runs should prove the following:

- the integrated real-time rescheduling is conceptional and technical feasible;
- departure of trains and running trains can be rescheduled and operated precisely following a dynamically changeable schedule; and
- dispatchers are able to reschedule trains within reasonable time in order to avoid conflicts within a given area.

During the first phase with test runs on regular Intercity trains between Zurich and Chur, technical components and fundamental principles of the integrated real-time rescheduling were evaluated. Rescheduling was not tested thereby which means that reference trajectories were fixed. However, the test runs were used to gain varying insight: feedback by drivers, identification of technical problems, verification of reference trajectories. In parallel to the test runs, accurate departing procedures were evaluated in Lucerne in order to identify the achievable accuracy.

In a second phase, Interregio trains from Zurich heading for Lucerne and entering the single-line conflict area at Rotsee were used as a test scenario to evaluate all the components of the integrated real-time rescheduling methods together in operation for a single train. Thereby, the trains were equipped with a FARE tool and dispatchers used a specific rescheduling tool based on the pulsing method. Depending on all the trains delay in the Lucerne area, a new trajectory was sent by GSM-R to the DMI in order to achieve a conflict-free run into the station of Lucerne. These test runs and results with the FARE tool are described in detail in the following section below.
Finally, the integrated real-time rescheduling was tested for feasibility when multiple trains equipped with the FARE-DMI were departing or arriving in Lucerne around the same time. Thereby, dispatchers in Lucerne used the pulsing method based tool to generate conflict-free schedules for the entire duration of the tests. New schedules were sent by GSM-R to the FARE tools or by cell phone to station staff in order to guarantee departure on time. Tests with multiple trains using FARE were executed four times. The tests have shown that integrated real-time rescheduling with a multiplicity of trains is possible and feasible. However, the test runs could not be used to compare them with the results gained through simulation.

Altogether, technical difficulties, availability challenges and the required number of well-trained staff limited the number of test runs. Time synchronisation problems meant that several test runs were not evaluable and also disconnected data transfers between trains and traffic management center led to test run's being aborted. The limited number of possible test runs also made it impossible to test and evaluate various versions of the DMI. However, the FARE-DMI is being developed in cooperation with train drivers. Nevertheless, a comparison of possible performance (driving accuracy) with different DMIs would be useful.

### 7.4.3 The Rotsee test case

#### 7.4.3.1 Interregio train operations at Rotsee without integrated real-time rescheduling

The single-line section between the Rotsee and Lucerne (schematically illustrated in Figure 7.14) is a significant source of conflict between departing and arriving Interregio trains. In addition, Interregio trains heading for Lucerne have no scheduled trains around themselves causing knock-on conflicts between Rotkreuz and Rotsee and can therefore be used to test precise train operation principles (including speed-up or slow-down of the train) with FARE.

![Figure 7.14: Detailed topology of the Rotsee single-line section](image-url)
Limited data was available to identify conflicts and influences at the Rotsee crossing point. The control system only allowed the passing times of trains at main signals to be obtained with an accuracy of 1 second. However, the (partial) route release times, the route reservation times and the signal green times were not available for post-processing. Unique identification of a conflict or influence was therefore not possible.

Instead, conflicts were estimated based on the trains’ signal passing times. Therefore, three reference points (main signal positions) were defined:

- Reference point 1: signal EBI-F94: First main signal after the single line section where the conflict-causing train heading for Rotkreuz has completely released to conflict point at the Rotsee.
- Reference point 2: signal EBI-C3: First main signal for a train heading for the single-line section in the direction of Lucerne where a reduced speed can be signalised.
- Reference point 3: signal EBI-60P: First main signal after the Rotsee conflict point where full track speed is achieved.

The reference point 1 for trains leaving the Lucerne area in the direction of Rotkreuz and reference point 2 for trains heading in the direction of Lucerne are used to estimate a conflict or influence. Four different cases can be identified:

- there is a major likelihood of an influence: $t(tr_1[RP_1]) \leq t(tr_2[RP_2]) + t_{c1}$;
- there is a possible likelihood of an influence: $t(tr_2[RP_2]) + t_{c2} < t(tr_1[RP_1]) < t(tr_2[RP_2]) + t_{c3}$ and $t(tr_2[RP_3]) - t(tr_2[RP_2]) > t_{c3}$;
- there is a potential likelihood of an influence: $t(tr_2[RP_2]) + t_{c2} < t(tr_1[RP_1]) < t(tr_2[RP_2]) + t_{c3}$ and $t(tr_2[RP_3]) - t(tr_2[RP_2]) \leq t_{c3}$;
- there is a major likelihood of no influence: $t(tr_2[RP_2]) + t_{c2} \leq t(tr_1[RP_1])$;

with the point of time when the train heading for Rotkreuz passes reference point 1 $t(tr_1[RP_1])$, the point of time when the train heading for Lucerne passes reference point 2 $t(tr_2[RP_2])$ and reference point 3 $t(tr_2[RP_3])$ and the constant time factors $t_{c1}$, $t_{c2}$ and $t_{c3}$ which were determined based on measured data.

Consequently, the combined analysis for both the running time and relative positions of the two crossing trains help to improve the correctness of detecting influences. However, some uncertainty remains. The time taken for the trains to run from reference point 2 to reference point 3 depending on the relative position with the crossing trains is illustrated in Figure 7.15.
7.4. SBB test runs

Figure 7.15: Running time of Interregio trains towards the single line section at Rotsee in dependence of the relative passing time with opposite train (green: major likelihood of no influence, blue: potential likelihood of an influence, magenta: possible likelihood of an influence; red: major likelihood of an influence)

The data analysis at the Rotsee using $t_{c1} = -10s$, $t_{c2} = 10s$ and $t_{c3} = 95s$ for weekdays 7:00 to 20:00 between July and October 2008 with 2390 train runs yielded to the following results:

- influence with major likelihood exist in 24.9%;
- a possible likelihood of an influence occur in 4.0%;
- in 5.3% of all cases an influence is potentially possible; and
- 65.8% had a major likelihood of no influence.

The four cases are also illustrated using different colours in Figure 7.15. A rather small variation around a constant value of the running time with several statistical outliers is observable for the case with no influence. For trains with an influence, the linear ascent combined with the larger variations of the running time is also visible for trains having an influence. In the uncertain range in-between, a discontinuity or step can be identified for the running time resulting from the non-linear delay effects. The detailed extract of the uncertain range is visualised in Figure 7.16. Summing up, the post-analysis of railway operation data with conventional operations showed that there is a large potential for improvement of rail traffic flow with integrated real-time rescheduling at the Rotsee conflict point.
7.4.3.2 Interregio train operations at Rotsee with integrated real-time rescheduling

7.4.3.2.1 Setup of integrated real-time rescheduling Rotsee test runs

Between July and November 2008, 28 test runs from Rotkreuz to Lucerne with FARE were conducted by the SBB. Chief train drivers accompanied the tests and were responsible for their installation and initiation on the trains. Movement inspectors and dispatchers specified the target passing times which resulted in input trajectories in the FARE tool.

The FARE tool, used for the Rotsee tests, archived position, time, temporal deviation and speed among other things every four seconds. Additional communication data between the FARE server and the FARE tool as calibration position or reference trajectories were also logged with an accurate time stamp. The driving behaviour, discrepancies, abnormalities and accuracy achieved can be analysed in detail. An additional test journal was used to describe all events and incidents.

The objective of the Rotsee test runs was to gain insights into driving behaviour and to identify the driving accuracy achievable with the supporting FARE tool during regular operations and after rescheduling. However, a limited number of test runs were used to transmit a reference trajectory to the FARE tool based on the temporal requirements of the test train and the opposite train. The reference trajectory could imply a deviant departure time at Rotkreuz and reduced maximum speed between Rotkreuz and the Rotsee. Rescheduling were completed if the test train’s temporal deviation was too large but not in the case of a delay of another train. Consequently, the interaction of inner precise operation with the outer rescheduling control loop was tested only to a limited degree.
Technical problems were identified during the initial test runs between Zurich and Chur. Availability and reliability were also improved based on the insights gained in these tests. However, technical problems as communication interrupts, wrong position calibration, login problems or unknown reasons caused 8 out of the 28 Rotsee test runs to be cancelled. Furthermore, wrong or improper reference trajectories were transmitted in 5 cases. 15 test runs therefore remained and were used for the detailed analysis.

Depending on the conditions during the test runs, the test train’s delay, and the delay of other trains, the test runs can be divided into three categories:

- the test train was much too late and thus had to be operated with minimum running time as reference trajectory;
- unexpected events and special conditions (such as bad adhesion conditions due to heavy rain falls or temporary slow speed area) required specific reference trajectories or caused larger temporal variations; and
- regular conditions, which would normally cause a conflict at Rotsee, with a reduced-speed reference trajectory for the test train.

In 4 of the 15 remaining test runs, the reference trajectory was equal to the minimum running time due to train delays at Rotkreuz. During 5 test runs, special conditions were measurable or unexpected events occurred. Consequently, only 6 test runs were completed under regular conditions.

Together with the speed profile, the temporal deviation in comparison to the actual valid reference trajectory is the relevant analysis measurement during the test runs. The accuracy achievable and the driving behaviour is evaluated for the different test runs in the following.

### 7.4.3.2.2 Results of Rotsee FARE test runs with minimum running time reference trajectories

Firstly, the test runs with minimum running time as a reference are analysed. The temporal deviations for these four test runs are illustrated in Figure 7.17. The following effects are recognisable:

- The train running at August 19 was influenced by a closed signal in Ebikon (position 9,750m). However, the increase in delay of 11 seconds is marginal.
- Determining a reference trajectory with minimum running time is difficult. Train runs on August 26 and November 3 show that faster runs compared to the calculated minimum running time are partially possible, in particular in sections with accelerations. Also significant deviations are measurable in slow speed areas (e.g. Lucerne station and Vorbahnhof), in particular, on October 28 and November 3.
• An error in transmission of the reference trajectory occurred on October 28 shortly after position 4,000m. This led to the step shape of the temporal deviation.

![Figure 7.17: Development of the temporal deviation for trains running with FARE from Rotkreuz to Lucerne with minimum running time as scheduled reference trajectory](image)

7.4.3.2.3 Results of Rotsee FARE test runs with special conditions

Special conditions or unusual events occurred during five test runs. The temporal deviation for these runs is visualised in Figure 7.18. The special circumstances and insights from these test runs are described subsequently:

- July 14: Bad adhesion conditions due to heavy rainfalls cause time to be lost during acceleration and braking. Smart reference trajectories may take such conditions into account.
- July 22: A shunting movement in Ebikon (position 8,500) caused the train to slow down. Consequently, shunting movements also have to be taken into account when generating conflict-free reference trajectories.
- July 28: A step in the temporal deviation (position 2,000m) due to an unknown technical problem was followed by inappropriate driving actions where speed was too high for unnecessarily long time resulting in a maximum earliness of 81 seconds. This overshooting (from 30 seconds delayed to 81 seconds in advance) effect shows that an unusually large differing maximum speed as reference input (80km/h instead of 140km/h) may cause significant deviations for tools without any speed information as FARE.
- September 19: Maintenance work around position 4,000m required a speed restriction. The reference speed used for the FARE tool was slower than the maximum allowed restricted speed. The higher speed with which the train was operated in this area resulted in...
the deviation quickly changing from a delay of 11 seconds to an earliness of 20m seconds. Once again, overreaction by the driver resulted in too slow speed after the maintenance area, which finally ended in a delay of 35 seconds and a rescheduling action due to inappropriate behaviour.

- September 22: An adjusted reference speed equal to the maximum speed permitted in the section subject to speed restriction resulted in fewer deviations during the entire passing procedure.

![Figure 7.18: Development of the temporal deviation for trains running with FARE from Rotkreuz to Lucerne under special conditions or occurring events](image)

**7.4.3.2.4 Results of Rotsee FARE test runs under regular conditions**

Finally, the six test runs with regular conditions including a reduced speed reference trajectory are analysed. The temporal deviation of these test runs is visualised in Figure 7.19. The following behaviours can be observed:

- Four out of the six test runs remained within a bandwidth of ±15 seconds for the entire test run.

- Increases in delays are measured at speed changing sections, in particular between position 2,000m and 4,000m (for trains on August 4, August 5 and October 6), where drivers are unsure of which speed they have to accelerate to.

- A slow reaction (initial overspeed of 20 km/h followed by late braking) resulted in a maximum earliness of 25 seconds after an initial delay of 10 seconds for the train run on October 7.
• No appropriate reaction to the temporal deviation can be observed for the train run on July 15. Late rescheduling by shifting the reference trajectory with 30 seconds made the train on-time.

• A crossing train caused the test train to stop on August 4 shortly before the Lucerne station (position 15,500m). Sending a new trajectory with an updated arrival time caused a jump in the temporal deviation at this point.

Figure 7.19: Development of the temporal deviation for trains running with FARE from Rotkreuz to Lucerne

7.4.3.2.5 Results of Rotsee FARE test runs with minimum running time reference trajectories

Comparable to the tests without FARE, the running time between Ebikon and Rotsee is visualised in dependence of the relative crossing times (see Figure 7.18). The test run from July 22 (a conflict with a shunting movement in Ebikon caused a slowdown) shows the excessively long travel time in this section and also illustrates that the train would have been influenced by the opposite train. Using the conflict classification parameters introduced in section 7.4.3.1, the second train run with a reported conflict (dated August 19) is classified as 'potential likelihood of an influence'. Using the maximum speed as a reference trajectory on August 19 resulted in an approximately similar time used to pass the critical section in comparison with trains without an influence but lower reference speed at Ebikon. The analysis with the temporal lag of the opposite train also shows that too many time margins between the two trains were used resulting in an unnecessary capacity loss.
To summarise, the test runs with FARE for the Rotsee conflict point yield the following insights:

- It seems that an accuracy of $\pm 15$ seconds is achievable under normal circumstances.
- The rate of trains having an influence could be reduced from around 30 percent down to 13 percent with FARE.
- Major differences in the reference speed cause larger temporal deviations due to the lack of missing information for the drivers.
- Behaviour after a larger deviation differs: both slow reactions as well as overshooting were observed. Missing supporting information would reduce variations and speed-up the time required to return within a given tolerance bandwidth.
- The limited amount of information provided with FARE (especially the missing reference speed information) is not sufficient to guarantee an accuracy of $\pm 15$ seconds under all conditions. In particular, late rescheduling actions causing reference trajectories with bathtub shapes (see section 5.2.6.1) are supposably not manageable. However, integrated real-time rescheduling requires late reschedulings in order to improve train traffic flow.
- Tests with rescheduling during a train run are required. Resulting effects of temporal shifting and distorting the reference trajectory as well as consequences by providing different information to the driver must be determined.
• Reference trajectories must be precise taking into account all details including restricted speed sections, insulated sections, or braking tests after departure.

• To improve infrastructure usage, the amount of buffer times between the two opposite trains at Rotsee should be reduced to a minimum.

• Well-trained supervisors and staff were used for the test runs in order to guarantee the high reliability of the tools and the correct application of information provided.

• High reliability and the robustness of the tools are required for efficient testing and daily usage.

7.5 Conclusions

The simulation studies for the areas around Bern and Lucerne have shown that the integrated real-time rescheduling framework can improve traffic flow resulting in significantly less knock-on delays. In particular, the combination of precise train operations and real-time rescheduling allows selective measures to be taken at an early point of time. Consequently, time-consuming conflicts were avoided. However, the non-linear delay effects caused by a signal influence were shortened in the simulation areas due to slow track speed in the station area and the signalling system which often allowed reacceleration directly at the point of time when the signal turned green.

In addition to improved traffic flow, the simulation studies were also used to prove that an increased usage of the infrastructure in condensation areas is possible. The effect of precise train operations (e.g. with the help of supporting advice tools) was demonstrated. Altogether, the simulation studies showed that improvements over conventional railway operations are possible with integrated real-time rescheduling for both stability and capacity.

The test runs have proved that the framework of the integrated real-time rescheduling system is technically feasible. However, robustness and reliability problems in particular for data connections as well as time synchronisation challenges need to be solved. The test runs also showed that the staff involved must be well trained, familiarised with the system, and motivated. Under regular conditions, the FARE tool developed allowed an accuracy of $\pm$ 15 seconds to be achieved following a reference trajectory. Nevertheless, abrupt changes in the reference trajectory were not tested. With the actual FARE system where no speed information is displayed, it is supposed that accuracy will worsen significantly under such conditions and with reschedulings involving larger speed changes. Consequently, the impacted of the data displayed should be investigated with extensive test runs using different DMIIs including speed information.
Chapter 8

Conclusions

8.1 Summary

Rail traffic has grown rapidly within the last few years and further continuous growth is projected. Consequently, rail networks are being increasingly operated at the limits of their capacity and stability. The focus of the present thesis consists of developing a new framework in order to improve existing performance. The new framework combines dynamic traffic management solving events, conflicts and delays in real-time in combination with advice tools supporting drivers and guards to operate trains precisely following dynamically changeable schedules. This research combines recent developments in operations research and enhancements to technical systems. The thesis shows that feedback control methods can be adapted and applied for railway operations.

The integrated real-time rescheduling framework developed and introduced was designed as a superimposition of two control loops. A precise description of the framework with the outer rescheduling loop and the inner precise operation loop was used to identify all the relevant elements and processes within the framework. Based on today’s operation principles and technological implementations, three main components to improve were identified for an integrated real-time rescheduling framework:

- Up-link to transmit train state and position as well as delay, incident or event information frequently and accurately. This can be achieved by using GSM-R or other communication channels.Implicitly precise position determination and synchronised times for all actors must be ensured.

- Software to handle event and delay information, identify primary delay reasons, produce adequate predictions, and generate new production plans based on enhanced operation research techniques in real-time.
Down-link (e.g. by GSM-R) to transmit new schedules to drivers, guards and infrastructure operators as well as advice tools displaying the updated recommendations to the actors involved.

The analysis of the rescheduling loop has shown that delays and time lags may occur for all processes. Enhancements were evaluated in order to minimise the overall duration. Along with promising development for rescheduling algorithms, the new framework will make it possible that schedules can be generated and transmitted to all actors involved for a local area within 30 seconds. However, a balance must be found between nervous rescheduling due to vague data, simplifications for the algorithms generating a new schedule, and short rescheduling duration.

Visualised recommendations based on an updated schedule are at the focus of the inner accurate production control loop. It was shown how deviations arise and that supporting tools are required for all the actors involved in order to ensure accurate production. Architectures and designs of the driver-machine interface (DMI) were at the focus of the evaluations. Models showed that an operational accuracy of ± 15 seconds (which is required at least for a substantial efficiency improvement) may achievable as long as suitable information is provided. However, test-runs with trains using a DMI must be run in order to evaluate the accuracy achievable during daily operations.

The thesis demonstrated that the major added benefit of the new framework with the superposition of the two feedback control loops adapted for railway operations is the opportunity to avoid conflicts by adjusting driving behaviour to precisely following dynamically changeable trajectories. In contrast to conventional rescheduling systems, where trains can only be delayed, rerouted or reordered, being able to retine and control running trains precisely is a significant improvement.

The potential benefits of the new integrated real-time rescheduling framework on stability, energy consumption and capacity usage were derived and subsequently determined by simulation evaluations and field tests for the specific areas around the main stations of Lucerne, Bern and Winterthur. The simulation tests proved that the integrated real-time rescheduling framework can significantly increase rail traffic flow and thus reduces delay propagation. In particular, early rescheduling and avoiding conflicts due to precise train operations are major benefits of the new system. Test runs showed that, in particular, reliability and robustness are considerable challenges in the development for the new system. However, it was proven that the integrated real-time rescheduling is technically feasible. With the FARE-DMI developed, an accuracy of ± 15 seconds was achieved for test runs under regular conditions. Nevertheless, abrupt changes in the reference trajectory after rescheduling and specific circumstances were not tested. It seems that accuracy is reduced in these cases with the actual FARE display, and evaluations with other information including speed advices are recommended.

The tolerance band principle, a key element for a fast and consistent rescheduling process, was introduced. Each train is assigned a temporal tolerance in which disturbances and variations can be absorbed by the actors. Whenever a train exceeds the tolerance limit, rescheduling will be initiated. In addition, applying the tolerance band helps to make the planning process more
8.2. Migration possibilities for an integrated real-time rescheduling system

systematical. An excessively small tolerance band results in nervous behaviour whereas an excessively large size unnecessarily reduces capacity. The tolerance band principle is comparable to the green band used in Zurich’s dispatching center in order to identify possible conflicts. However, green bands were not systematically used by dispatchers.

The thesis demonstrated that clever network decomposition in condensation (capacity bottlenecks) and compensation areas (track sections with sufficient capacity) help to improve the benefits of the integrated real-time rescheduling framework. Running time supplements on feeder lines in compensation areas are used to schedule and control trains in order to enter condensation areas at the precisely specified time and at the given speed. Time- and capacity-consuming signalling influences can thus be minimised in condensation areas, which allows a seamless sequence of trains with minimum buffer times in between. It was demonstrated that combining the integrated real-time rescheduling framework with the specific operation of trains in condensation areas consequently results in the maximised capacity usage of a given infrastructure.

Further challenges can be solved by applying the new integrated real-time rescheduling framework. The prediction of train movements and connections will be more accurate. Passengers can thus be informed more precisely and at an earlier point in time about their further journey. Terminal planning and just-in-time production for freight transports are also improved because prediction is more accurate and reliable. Real-time rescheduling algorithms can be used in addition for planning long- and mid-term schedules. Various timetable and infrastructure enhancement scenarios can thus be compared. Finally, safety will be improved due to the fact that the number of closed signals approached will decrease.

8.2 Migration possibilities for an integrated real-time rescheduling system

Based on actual operation principles and available technologies, intermediate steps in the direction of a fully integrated real-time rescheduling system are conceivable. Migration steps are possible for implemented functions and architecture, for geographical limited areas and for selected train types.

Firstly, possible steps for the implementation and functions as well as their consequential characteristics are described below:

- Only advanced rescheduling algorithms: control loop principles and accurate production are not implemented yet. Retiming of trains is consequentially only possible to a limited degree, in particular, train departures can be delayed. Time and capacity consuming conflicts are unavoidable. Altogether, the benefits of precise train control following the dynamically changeable trajectories resulting in reduced energy consumption, increased stability and capacity are not applicable if only algorithms or rail traffic management systems are enhanced.
• Integrated real-time rescheduling framework without the implementation of the tolerance band principle: a predefined, consistent rule to initiate the rescheduling process does not exist yet. Early detection of deviations or events is therefore restricted, resulting in reduced rescheduling performance.

• Integrated real-time rescheduling framework without active data up-link from trains to the rescheduling system: the time taken for rescheduling can be significantly reduced through periodic state transmission including train position. In addition, past behaviour can be better understood and retraced with dense recorded data which consequently improves prediction quality. Coded messages about events also help to speed up and improve the quality of the entire rescheduling process. Consequently, a more nervous rescheduling system and reduced rescheduling performance are expected without the active data up-link.

• Integrated real-time rescheduling framework without network decomposition in condensation and compensation areas: applying the integrated real-time rescheduling framework in its full functionality helps to reduce significantly knock-on delays and their propagation in the network. However, improving capacity as well as stability requires new strategies for buffer time allocations. Omitting the condensation and compensation principle of network decomposition therefore causes more buffer times to be allocated in bottleneck areas. Consequently, capacity usage may not be maximised in these areas.

To summarise, integrated real-time rescheduling and clever network decomposition together strongly increase performance. The combination of these two principles results in added benefits. Implementing only some of the functions therefore have a significantly negative effect on stability and capacity.

Secondly, migration questions addressing the difficulty, if the new integrated real-time rescheduling framework with all function is to be implemented network-wide and for all trains or if reduced implementation is possible. Geographical restrictions and limited train categories can be distinguished:

• Integrated real-time rescheduling framework only applied in partial network sections: it is certainly meaningful to focus initially on delay-prone capacity bottlenecks and their linked feeder lines and to apply the rescheduling and precise train control framework there. Coordination between different areas can also be reduced on information transfers as today. Complex algorithms ensuring network-wide optimisation and coordination can be implemented subsequently. However, trains equipped with supporting tools may use incoming data in the entire network. A gradual implementation of network sections around condensation areas thus seems a promising approach.

• Integrated real-time rescheduling framework only applied by selected train categories: the analysis showed that Intercity and freight trains, in particular, produce more knock-on delays compared to commuter trains due to their reduced dynamics. Due to the longer
8.3. Future research and developments

Distances between scheduled stops or entering the next condensation area, these trains can be handled better for a retiming measure. However, equipping not all trains means that deviations may be detected late and appropriate rescheduling measures are not possible in all cases. Also, the systematic saturation of condensation areas is not possible when trains are not equipped accordingly. Larger tolerance bandwidths or buffers have to be planned for trains without a DMI in order to absorb the comparable larger variations in their operations. Sporadic, isolated trains without a supporting tool are possible in order to minimise the amount of non-utilisable buffer times. Therefore, after a preliminary equipping of Intercity and freight trains with a DMI, the aim should be to equip as many trains passing capacity bottlenecks as possible.

For a successful implementation of the integrated real-time rescheduling framework, a gradual implementation is recommended. Various designs and implementations of DMI have to be tested on their performance in a first step. After the evaluation, selected Intercity and Interregio trains should be equipped in a second step with the DMI. Thereby, the selected trains should be controlled (have conflict-free reference trajectories) on several well-known bottleneck areas which could be either around a station or on an open track area with single line section, crossing or joining. For this stage, an automated one way connection from traffic management to train and manual development of conflict-free schedules is sufficient. This migration phase can be used to analyse and evaluate interdependencies and open questions introduced in this thesis. In a third stage of development, automated solving of conflicts (amount of inputs by dispatchers limited to final decisions) can be implemented. Finally, in a fourth step, all trains can be equipped with a DMI for more accurate running of trains including precise departure at stations. Nevertheless, main corridors and bottleneck areas can be operated with the integrated real-time rescheduling framework first and then, if required, the implementation can be extended to the entire network.

8.3 Future research and developments

A multiplicity of developments, tests, and research is still needed for a successful and efficient implementation of the new integrated real-time rescheduling framework. Major developments have to be done on the following topics:

- Fast rescheduling algorithms: A key position within the integrated real-time rescheduling framework consist in the generation of new production plans. So far, promising results for real-time rescheduling have been achieved recently. Nevertheless, research is still needed to obtain algorithms proving that accurate and realistic schedules can be generated within seconds for complex networks with dense traffic and coordination over several areas. The algorithms must also prove their applicability in the event of larger disturbances or major events.
• Prediction of future behaviour: The handling and manipulation of incoming data used for prediction was so far unattended. However, this data is of high importance and a balance between rescheduling performance and rescheduling nervousness must be found. Thereby, rescheduling method and policy, as well as the transmission of non-feasible production plans must be taken into account. Consequently, a specific survey is recommended.

• DMI design and achievable accuracy: The driving accuracy achievable was demonstrated with initial test-runs and a specific DMI. Nevertheless, the accuracy achievable during daily operations must be demonstrated in a long term survey. Different DMI designs and displays can be compared for their effectiveness. In particular, the display of speed information for managing abrupt changes in the reference trajectory should be studied and discussed.

• Dwell process: In parallel to driving test, also long term surveys on departing accuracy achievable during daily operations must be investigated. An analysis including quantification of the entire dwell process to identify and solving weak points is needed.

• Utilisation limit: Based on the accuracy achievable for running and departing trains, the improved level of infrastructure usage during daily operations and thus the available capacity gain with the new framework must be analysed. The utilisation limit, at which the system gets unstable, is thereby of highest interest. Research about the basic interrelations of capacity, infrastructure utilisation, buffer and supplementary time allocations, operation accuracy and stability is suggested.

• International standards for railway operations: Developments for new rail traffic management systems and DMIs are undertaken in different countries. To avoid new barriers, specific standards must be defined. The thesis suggested different levels. In particular, the data transferred between traffic management system and trains must be specified at least. This challenge has to be identified and solved in order to establish successfully new integrated real-time rescheduling systems. Consequently, standards defined have to be part of a detailed migration strategy to be developed which must satisfy the requirements specified and the budget available.

• Energy optimal reference trajectories: The thesis proved that applying an integrated real-time rescheduling system will reduce traction energy consumption. To improve overall energy savings, new trajectories to be developed must combine calculation of conflict-free slots with the consolidated findings of earlier research in the field of energy optimal train control.

• Data handling: Exceptions and distinctions are common in railway systems. Handling and translation into rules for a dynamic traffic management system should not be underrated and will require an immense effort. A comparable effort is needed to provide accurate infrastructure and train data based on actual conditions. Precise schedules are
useless when changes, for example due to maintenance work, are not updated continuously. Consequently, appropriate system architectures must be developed in order to handle exceptions, database and information flows efficiently.

- **Short term slot selling:** The revenue and efficiency of a rail infrastructure is maximised, when train slots can also be sold in the short term. The rescheduling algorithms may therefore be extended to find possible train paths through the network. Reserved but unused track slots can then be sold efficiently and the entire path can be scheduled and guaranteed.

- **Automation and role of human in operations:** In long term, when infrastructure has to be operated even more efficiently, fully automated systems with adaptive, state feedback or processor based optimisation controller may be required. Thereby, these systems will react within the shortest time, detect trends and deviations early and will thus finally have smaller temporal deviations. A discussion on semi-auto-piloted driving or other replacements of humans during operations to improve train traffic flow in conflicts or heavily used areas is thus unavoidable in the future.

Building a detailed migration plan is thereby the basis for all development and research activities. Consequently, a strong coordination between all actors and researchers involved must be aspired in order to obtain the best possible efficiency improvement with the new framework.
Appendix A

Abbreviations

ATC   Automatic train control
ATO   Automatic train operation
ATP   Automatic train protection
BAV   Swiss Federal Office of Transport
DB AG German Railways
DFD   Data flow diagram
DMI   Driver-machine interface
ERTMS European rail traffic management system
ETCS  European train control system
ETH   Swiss Federal Institute of Technology
FARE  DMI developed by SBB AG
GPS   Global positioning system
GSM-R Global system for mobile communication - railway
IFOR  Institute for Operations Research
iRTR  integrated Real-Time Rescheduling
IVT   Institute for Transport Planning and Systems
LZB   Type of continuous ATP system applied in Germany
MPC   Model predictive control
PULS 90 Program by SBB AG to improve the capacity and efficiency of its rail infrastructure
PZB   Type of intermittent ATP system applied in Germany
RAMS  Reliability, availability, maintainability and safety
SBB AG Swiss Federal Railways
UIIC  International Union of Railways
ZEB   Future development strategy for the Swiss rail infrastructure
ZUB   Type of intermittent ATP system applied in Switzerland
Appendix B

Evaluation of running times

B.1 Analysis of train driver behaviour

Measured infrastructure data was used in order to get a more detailed insight of the running behaviour and temporal aspects of running trains. Only limited amount of data and statistical evaluations are available concerning train driving behaviours nowadays. This is in contrast to the dwell process where a larger data volume is available.

The focus of the analysis of the running time data was get a more detailed comprehension of the train drivers behaviours. In particular:

- Does the actual delay impact the driving behaviour? Are early trains operated slower than late trains?
- Do train drivers hold their driving behaviour constant? When a train is operated quickly in a first section, does it also run fast in a following section?
- How large are the running time variations and what shape of density curves fit best?

Infrastructure data from the entire SBB network from January 2003 to November 2003 was used for the ex-post analysis (see also section 2.3.4). To identify the train driver behaviour, trains and track sections have to be selected in a way to ensure that trains to analyse are not influenced by other trains. Consequently, only trains with sufficient buffer times (at least 10 minutes) to the previous train were selected. Sections with heavy usage of freight trains were also not selected because conflicts can not be excluded. Also larger station areas were not suitable for the analysis: trains may have changing routes with varying maximum permitted speed as well as other trains may influence the train to analyse.

The ensure comparability of the data, only one given train operated at weekdays around 9:00 in the morning on the selected track section was used in each case for the analysis. Consequently,
the data basis for the analysis consist of around 230 train runnings per section and train group. Intercity (IC) or Interregio (IR) as well as commuter (R) or suburban (S) trains were used in the analysis. Altogether, 17 trains on 11 track sections were analysed (see table B.1).

Table B.1: Trains and track sections used for analysis of train driver behaviour

<table>
<thead>
<tr>
<th>Track section analysed</th>
<th>Geographical location</th>
<th>Trains analysed</th>
<th>Planned running time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotkreuz - Gisikon</td>
<td>Zug - Luzern</td>
<td>IR 1935 R 6929</td>
<td>156 s 174 s</td>
</tr>
<tr>
<td>Goldach - Moerschwil</td>
<td>Rorschach - St.Gallen</td>
<td>IR 2820 R 8130</td>
<td>228 s 264 s</td>
</tr>
<tr>
<td>Cornaux - Le Landeron</td>
<td>Neuchatel - Biel</td>
<td>IC 615</td>
<td>132 s</td>
</tr>
<tr>
<td>Mels - Sargans</td>
<td>Zurich - Chur</td>
<td>IC 759</td>
<td>96 s</td>
</tr>
<tr>
<td>Lengnau - Biel Matt</td>
<td>Biel - Solothurn</td>
<td>IC 510</td>
<td>246 s</td>
</tr>
<tr>
<td>Dagmersellen - Nebikon - Wauwil</td>
<td>Olten - Luzern</td>
<td>IC 345 R 5121</td>
<td>252 s 144 s / 186 s</td>
</tr>
<tr>
<td>Sulgen - Erlen - Oberaach</td>
<td>Winterthur - Romanshorn</td>
<td>IC 909 R 7035</td>
<td>216 s 234 s / 156 s</td>
</tr>
<tr>
<td>Raron - Gampel - Turtmann</td>
<td>Brig - Lausanne</td>
<td>IR 1714 R 3732</td>
<td>204 s 192 s / 174 s</td>
</tr>
<tr>
<td>Chenens - Cottens</td>
<td>Lausanne - Bern</td>
<td>IC 2719 R 3525</td>
<td>102 s 150 s</td>
</tr>
<tr>
<td>Uzwil - Flawil - Gossau</td>
<td>Winterthur - St.Gallen</td>
<td>R 8123</td>
<td>264 s / 222 s</td>
</tr>
<tr>
<td>Bubikon - Wetzikon</td>
<td>Zurich S-Bahn</td>
<td>S 18526</td>
<td>294 s</td>
</tr>
</tbody>
</table>

B.2 Results of train driving behaviour analysis

Figures B.1 to B.21 illustrate the density of the measured running times as well as the measured running time in dependency of the train’s delay at the beginning of the measured section for each train and track section analysed. As exemplarily introduced in section 2.3.4.3.1, the figures show that the log-logistic density function fits best for all kind of trains and track sections. The figures also show that the actual delay of the train has in general no influence on the train drivers behaviour and thus the running time. An adapted driving behaviour, longer running times for early trains, can only be observed for a few data points (train drivers) and exclusively for IC or IR trains (for example between Rotkreuz and Gisikon figure B.1, between Cornaux and Le Landeron figure B.3 between Dagmersellen and Wauwil figure B.6 or between Chenens and Cottens figure B.9).

The running time for two consecutive track sections with stops at the beginning, the middle and the end was analysed for commuter trains (see figures B.22 and B.23). The figures show that the
running time of the first section has in general no significant correlation with the running time on
the second section. An exception is the section Dagmersellen - Nebikon - Wauwil (figure B.22)
at which a trend of identical driving behaviour (and thus running times) in the two consecutive
sections is observable.

These fundamental insights may not only be used for the rescheduling evaluations. In addition,
these findings should be basis as input for stochastic models for planning including robustness
analysis as well as online for prognosis calculations. However, the analysis may be extended
by evaluating the influence of several additional parameters such as running time supplements,
time of day as well as behaviour in station or conflict areas.

Figure B.1: Running times for IR 1935 Rotkreuz - Gisikon

Figure B.2: Running times for IR 2820 Goldach - Moerschwil
Figure B.3: Running times for IC 615 Coraux - Le Landeron

Figure B.4: Running times for IC 759 Mels - Sargans

Figure B.5: Running times for IC 510 Lengnau - Biel Matt
B.2. Results of train driving behaviour analysis

Figure B.6: Running times for IC 345 Dagmersellen - Wauwil

Figure B.7: Running times for IC 909 Sulgen - Oberaach

Figure B.8: Running times for IR 1714 Raron - Turtmann
Figure B.9: Running times for IR 2719 Chenens - Cottens

Figure B.10: Running times for R 6929 Rotkreuz - Gisikon

Figure B.11: Running times for R 8130 Goldach - Moerschwil
B.2. Results of train driving behaviour analysis

Figure B.12: Running times for R 5121 Dagmersellen - Nebikon

Figure B.13: Running times for R 5121 Nebikon - Wauwil

Figure B.14: Running times for R 7035 Sulgen - Erlen
Appendix B. Evaluation of running times

Figure B.15: Running times for R 7035 Erlen - Oberaach

Figure B.16: Running times for R 3732 Raron - Gampel

Figure B.17: Running times for R 3732 Gampel - Turtmann
B.2. Results of train driving behaviour analysis

Figure B.18: Running times for R 3525 Chenens - Cottens

Figure B.19: Running times for R 8123 Uzwil - Flawil

Figure B.20: Running times for R 8123 Flawil - Gossau
Appendix B. Evaluation of running times

Figure B.21: Running times for S 18526 Bubikon - Wetzikon

Figure B.22: Running times of consecutive sections: R 3732: Raron - Gampel - Turtmann (left), R 5121: Dagmersellen - Nebikon - Wauwil (right)

Figure B.23: Running times of consecutive sections: R 7035: Sulgen - Erlen - Oberaach (left), R 8123: Uzwil - Flawil - Gossau (right)
Appendix C

Influence of signals on train delays and capacity usage

C.1 Introduction

One goal during the planning and rescheduling process is to calculate conflict-free routes and track slots. In reality, delays occur and trains can approach closed signals and thus be influenced by them. Consequently, additional delays could occur, depending on the state of the signal and the signalling system. In planning, this fact is considered by adding running time supplements and buffer times between two trains. For rescheduling, the influence of the signal state on the running time is neglected in most models or imprecisely approximated.

The model presented in this section illustrates, that the delay effects due to signal impacts cannot be neglected. It also visualises the potential added value of an integrated rescheduling system where trains can be controlled precisely in order to pass a given reference point (signal) at a specific and optimum point of time at a predefined speed.

The model is used to describe the transfer function of the running time \( t_{\text{run}} \) and the passing time of a second reference point \( t_{\text{pass}} \) in dependence of the point of time a first reference point is passed \( t_x \), the signal green time \( t_{\text{sa}} \) and also the infrastructure and train characteristics:

\[
\begin{align*}
  t_{\text{pass}} &= f(t_x, t_{\text{sa}}) \\
  t_{\text{run}} &= f(t_x, t_{\text{sa}})
\end{align*}
\] (C.1)

The following section describes the influence on running time and arrival or passing time for a train for two basic and common signalling systems. The impact of various factors is shown. Delay effects occurring for related signalling systems could be derived from the two basic systems presented. Based on the running time evaluation, the impacts and consequences of the signalling system on capacity usage are derived.
**Assumptions 1:** Several assumptions for simplification are assumed to exemplify the basic principles of the influence of a signalling system on a train’s travel time or arrival delay:

- The maximum permitted speed within the observed section (between two reference points around the main signal) is constant.
- The signal watching time, which is part of the regular blocking time and usually at least three seconds, is assumed to be zero.
- Braking actions are not taken before a distant signal is passed.
- Direct reacceleration of a train after a braking action is considered possible without a temporal lag. Regularly, the delay could take several seconds. The lag could last even longer, in particular, for freight trains.

### C.2 Notation

- $t_x$: Point of time train passes the first reference point (initial braking point or distant signal)
- $t_{sa}$: Point of time signal turns green
- $t_{pass}$: Point of time train passes the second time reference point (the point where a train at a complete standstill at a main signal reaches maximum speed)
- $t_{run}$: Time needed to pass the total track length $s_{total}$ considered between the two reference points
- $t_{wait}$: Time when the train waits in front of a closed main signal
- $t_{brake}(v_1 \rightarrow v_2)$: Braking time from speed $v_1$ to $v_2$
- $t_{acc}(v_1 \rightarrow v_2)$: Acceleration time from speed $v_1$ to $v_2$
- $t_{hold}(v_x)$: Time when the speed $v_x$ is held constant
- $t_{pass:min}$: Earliest point of time when a train passes a second reference point
- $t_{x:pass:min}$: Point of time train passes first reference point such that the second reference point is passed at the earliest point of time $t_{pass:min}$
- $\Delta t_{pass}(t_{x1}, t_{x2})$: Time difference at second reference point for two different passing times $t_{x1}$ and $t_{x2}$ at first reference point
- $v_{max}$: Maximum track speed permitted
- $v_{appr}$: Target speed permitted after passing a closed distant signal for the signalling system at a fixed approach speed
- $v_{min}$: Minimum speed obtained during the braking phase
- $v_{hold}:x$: Speed $v_x$ that is held constant
- $v_{bd}$: Boundary speed used to distinguish two different cases for the analysis of the signalling system at a fixed approach speed
- $a_{brake}$: Braking factor
- $a_{acc}$: Acceleration factor
C.3 Analysis for a signalling system which allows direct reacceleration

C.3.1 Derivation of running time and passing time equations for a signalling system with direct reacceleration

Firstly, the running time and delay effects are derived for a signalling system with information update during the approach of a main signal and permitted reacceleration before passing the main signal. Examples of such continuous train protection systems with the possibility of direct reacceleration are ETCS Level 2 or LZB (Germany) [Cat95].

Three cases, illustrated in Figure C.1, are differentiated to identify the running time and delay effects:

- Case 1: The signal turns to green after the train has come to a standstill in front of the signal (yellow speed distance curve in Figure C.1).
- Case 2: The signal turns to green while the train is braking and when approaching the signal (red).
- Case 3: The signal turns to green before the train passes the first reference point (initial braking point for the first two cases) and thus the train is not influenced by the signal (blue).

The variable which could be influenced in the case of integrated rescheduling is the point of time $t_x$ when the train passes a given first reference point. For simplification, the reference point is set at the initial braking point and the train’s behaviour before the braking point is identical in all cases.

The three cases are distinguished by the point of time when the reference point is passed:

- Case 1: $t_x < t_{sa} - t_{brake}(v_{max} \rightarrow 0)$
- Case 2: $t_{sa} - t_{brake}(v_{max} \rightarrow 0) \leq t_x < t_{sa}$
- Case 3: $t_{sa} \leq t_x$
Figure C.1: Speed-distance visualisation for the three different cases when a train passes a first reference point for a signalling system with continuous information and the possibility of direct reacceleration.

with the point of time when the signal turns to green \( t_{sa} \) and the time required to brake from the initial track speed down to a complete standstill \( t_{brake}(v_{max} \rightarrow 0) \).

The total track distance considered \( s_{total} \) between the two reference points is determined by case 1, where trains are coming to a standstill in front of a closed main signal. The total distance considered can therefore be determined by the summation of two parts. The first part of the distance includes braking down the train to a halt. The second part of the total length considered consists of accelerating the train from a halt at the signal to maximum speed. The overall length can thus be expressed as:

\[
s_{total} = s_{brake}(v_{max} \rightarrow 0) + s_{acc}(0 \rightarrow v_{max})
\]  

(C.3)

First, the time taken \( t_{run} \) to pass the total track length considered \( s_{total} \) is calculated for the three cases.

Case 1 consists of three phases. First the train brakes down to a halt. Afterwards, the train waits at the closed signal until the signal turns to green and finally reaccelerates to its maximum speed permitted. The running time for case 1 \( t_{run_{case1}} \) can thus be expressed as:

\[
t_{run_{case1}} = t_{brake}(v_{max} \rightarrow 0) + t_{wait} + t_{acc}(0 \rightarrow v_{max})
\]
the train halts in front of the main signal. The waiting time can thus also be expressed as:

\[ t_{\text{wait}} = t_{\text{sa}} - (t_x + t_{\text{brake}}(v_{\text{max}} \to 0)) \]

Consequently, running time for case 1 is:

\[ t_{\text{run case 1}} = -t_x + t_{\text{sa}} + t_{\text{acc}}(0 \to v_{\text{max}}) \] (C.4)

Case 2 also consists of three phases. First, the train brakes until the signal turns to green at which point the train has an actual speed \( v_{\text{min}} \). After the signal turns to green, the train reaccelerates directly up to its maximum speed permitted. The speed is then held during the third phase until the second reference point is passed. The running time in case 2 \( t_{\text{run case 2}} \) can therefore be expressed as:

\[ t_{\text{run case 2}} = t_{\text{brake}}(v_{\text{max}} \to v_{\text{min}}) + t_{\text{acc}}(v_{\text{min}} \to v_{\text{max}}) + t_{\text{hold case 2}}(v_{\text{max}}) \]

The braking time from maximum speed to the reduced speed \( v_{\text{min}} \) is equal to the difference between the signal green time \( t_{\text{sa}} \) and the point of time when the braking point is passed \( t_x \). Thus, the braking duration can be expressed as:

\[ t_{\text{brake}}(v_{\text{max}} \to v_{\text{min}}) = t_{\text{sa}} - t_x \]

Consequently, running time for case 2 can be rewritten as:

\[ t_{\text{run case 2}} = -t_x + t_{\text{sa}} + t_{\text{acc}}(v_{\text{min}} \to v_{\text{max}}) + t_{\text{hold case 2}}(v_{\text{max}}) \] (C.5)

In case 3, the train passes the total distance considered \( s_{\text{total}} \) constantly at a maximum speed permitted \( v_{\text{max}} \). Consequently, the running time in case 3 \( t_{\text{run case 3}} \) can be expressed simply as:

\[ t_{\text{run case 3}} = t_{\text{hold case 3}}(v_{\text{max}}) = \frac{s_{\text{total}}}{v_{\text{max}}} \] (C.6)

**Assumption 2:** For the evaluation of the impact of the signalling system on running time and delay, a constant braking factor \( a_{\text{brake}} \) and constant acceleration factor \( a_{\text{acc}} \) are assumed.

Under assumption 2, the distance and time for acceleration or braking can be expressed as:

\[ s(v_1 \to v_2) = \frac{v_2^2 - v_1^2}{2a} \]
\[ t(v_1 \to v_2) = \frac{v_2 - v_1}{a} \]
Using assumption 2 for equation (C.3), the total track distance considered can be expressed as:

\[ s_{\text{total}} = \frac{v_{\text{max}}^2}{2a_{\text{brake}}} + \frac{v_{\text{max}}^2}{2a_{\text{acc}}} = \frac{v_{\text{max}}^2}{2} \left( \frac{1}{a_{\text{acc}}} - \frac{1}{a_{\text{brake}}} \right) \]  

(C.7)

Under a condition of constant braking and acceleration factors, three cases can be identified:

- **Case 1:** \( t_x < t_{sa} + \frac{v_{\text{max}}}{a_{\text{brake}}} \)
- **Case 2:** \( t_{sa} + \frac{v_{\text{max}}}{a_{\text{brake}}} \leq t_x < t_{sa} \)
- **Case 3:** \( t_{sa} \leq t_x \)

Under assumption 2, the time taken to pass the total track distance considered \( s_{\text{total}} \) is determined for each case.

For case 1, the running time, introduced in equation (C.4), can be rewritten as:

\[ t_{\text{run,case1}} = -t_x + t_{sa} + \frac{v_{\text{max}}}{a_{\text{acc}}} \]  

(C.8)

For case 2, first the minimum speed \( v_{\text{min}} \) and the remaining distance where the maximum speed is held constant \( s_{\text{hold,case2}}(v_{\text{max}}) \) were calculated.

The braking time with constant braking factor down to the minimum from the maximum speed takes to the difference of the point of time when the signal is passed and the point of time when the signal turns to green. Consequently, the minimum speed can be represented as follows:

\[ v_{\text{min}} = v_{\text{max}} + a_{\text{brake}}(t_{sa} - t_x) \]  

(C.9)

The distance \( s_{\text{hold,case2}}(v_{\text{max}}) \), during which the maximum speed after reacceleration is held until the second reference point is passed, is determined:

\[
\begin{align*}
\quad s_{\text{hold,case2}}(v_{\text{max}}) &= s_{\text{total}} - s_{\text{brake}}(v_{\text{max}} \rightarrow v_{\text{min}}) - s_{\text{acc}}(v_{\text{min}} \rightarrow v_{\text{max}}) \\
&= \frac{v_{\text{max}}^2}{2a_{\text{acc}}} \left( \frac{1}{-a_{\text{acc}}} - \frac{1}{-a_{\text{brake}}} \right) - \frac{v_{\text{min}}^2 - v_{\text{max}}^2}{2a_{\text{brake}}} - \frac{v_{\text{min}}^2 - v_{\text{max}}^2}{2a_{\text{acc}}} \\
&= \frac{v_{\text{max}}^2(a_{\text{brake}} - a_{\text{acc}}) - a_{\text{acc}}(v_{\text{min}}^2 - v_{\text{max}}^2) - a_{\text{brake}}(v_{\text{max}}^2 - v_{\text{min}}^2)}{2a_{\text{brake}}a_{\text{acc}}} \\
&= \frac{v_{\text{min}}^2(a_{\text{brake}} - a_{\text{acc}})}{2a_{\text{brake}}a_{\text{acc}}} \\
&= \frac{(v_{\text{max}} + a_{\text{brake}}(t_{sa} - t_x))^2(a_{\text{brake}} - a_{\text{acc}})}{2a_{\text{brake}}a_{\text{acc}}} \\
&= \frac{t_x^2a_{\text{brake}}(a_{\text{brake}} - a_{\text{acc}})}{2a_{\text{acc}}} + \frac{(v_{\text{max}}^2 + 2v_{\text{max}}a_{\text{brake}}t_{sa} + a_{\text{brake}}^2t_{sa}^2)(a_{\text{brake}} - a_{\text{acc}})}{2a_{\text{brake}}a_{\text{acc}}} \\
&= \frac{(v_{\text{max}}^2 + 2v_{\text{max}}a_{\text{brake}}t_{sa} + a_{\text{brake}}^2t_{sa}^2)(a_{\text{brake}} - a_{\text{acc}})}{2a_{\text{brake}}a_{\text{acc}}} \\
\end{align*}
\]  

(C.10)
Using assumption 2, the acceleration time \( t_{\text{acc}}(v_{\text{min}} \rightarrow v_{\text{max}}) \) can be rewritten as

\[
t_{\text{acc}}(v_{\text{min}} \rightarrow v_{\text{max}}) = \frac{v_{\text{max}} - v_{\text{min}}}{a_{\text{acc}}} = \frac{v_{\text{min}} - v_{\text{max}}}{a_{\text{brake}}} \frac{a_{\text{brake}}}{a_{\text{acc}}}
\]

\[
= - t_{\text{brake}}(v_{\text{max}} \rightarrow v_{\text{min}}) \frac{a_{\text{brake}}}{a_{\text{acc}}} = -(t_{\text{sa}} - t_x) \frac{a_{\text{brake}}}{a_{\text{acc}}}
\]

\[
= (t_x - t_{\text{sa}}) \frac{a_{\text{brake}}}{a_{\text{acc}}}
\]

(C.11)

Using equations [C.10] and [C.11], the running time for case 2 from equation (C.5) can be rewritten as:

\[
t_{\text{run}}_{\text{case}2} = (t_{\text{sa}} - t_x) + (t_x - t_{\text{sa}}) \frac{a_{\text{brake}}}{a_{\text{acc}}} + s_{\text{hold\_case2}}(v_{\text{max}}) \frac{a_{\text{brake}}}{v_{\text{max}}}
\]

\[
= t_x^2 a_{\text{brake}} \frac{a_{\text{brake}} - a_{\text{acc}}}{2 v_{\text{max}} a_{\text{acc}}}
\]

\[
+ t_x \left[ \frac{u_{\text{max}} + a_{\text{brake}} t_{\text{sa}} (a_{\text{acc}} - a_{\text{brake}})}{v_{\text{max}} a_{\text{acc}}} + \left( \frac{a_{\text{brake}}}{a_{\text{acc}}} - 1 \right) \right]
\]

\[
+ \frac{(v_{\text{max}}^2 + 2 v_{\text{max}} a_{\text{brake}} t_{\text{sa}} + a_{\text{brake}}^2 t_{\text{sa}}^2) (a_{\text{brake}} - a_{\text{acc}})}{2 v_{\text{max}} a_{\text{brake}} a_{\text{acc}}}
\]

\[
+ t_{\text{sa}} \left( 1 - \frac{a_{\text{brake}}}{a_{\text{acc}}} \right)
\]

(C.12)

In case 3, the running time from equation (C.6) can be expressed as:

\[
t_{\text{run\_case3}} = l_{\text{total}} v_{\text{max}} = \frac{v_{\text{max}}^2}{2} \left( \frac{1}{a_{\text{acc}}} - \frac{1}{a_{\text{brake}}} \right)
\]

(C.13)

The point of time when a train passes the second reference point \( t_{\text{pass}} \) is the summation of the point of time when the train passes the first reference point \( t_x \) and the running time \( t_{\text{run}} \):

\[
t_{\text{pass}} = t_x + t_{\text{run}}
\]

(C.14)

Consequently, in case 1, the point of time when the second reference point is passed is:

\[
t_{\text{pass\_case1}} = t_{\text{sa}} + \frac{v_{\text{max}}}{a_{\text{acc}}}
\]

(C.15)
In case 2, the passing time is determined as:

\[
\begin{align*}
    t_{\text{pass,case2}} &= t_x \frac{a_{\text{brake}}(a_{\text{brake}} - a_{\text{acc}})}{2v_{\text{max}}a_{\text{acc}}} \\
    &+ t_x \frac{a_{\text{brake}} t_{\text{sa}}(a_{\text{acc}} - a_{\text{brake}}) + v_{\text{max}}a_{\text{acc}}}{v_{\text{max}}a_{\text{acc}}} \\
    &+ \frac{(v_{\text{max}}^2 + t_{\text{sa}}^2 a_{\text{brake}}^2)(a_{\text{brake}} - a_{\text{acc}})}{2v_{\text{max}}a_{\text{brake}}a_{\text{acc}}} \\
\end{align*}
\]  
(C.16)

And finally, the passing time in case 3 is:

\[
    t_{\text{pass,case3}} = t_x + \frac{v_{\text{max}}}{2} \left( \frac{1}{a_{\text{acc}}} - \frac{1}{a_{\text{brake}}} \right)
\]  
(C.17)

It can be shown that the time difference at the second reference point \( \Delta t_{\text{pass}}(t_{x1}, t_{x2}) \) between case 1 when a train has to stop at the closed main signal \( t_{x1} = t_{\text{sa}} + \frac{v_{\text{max}}}{a_{\text{brake}}} \) and case 3 when the train passes the first reference point immediately after the signal turns to green \( t_{x2} = t_{\text{sa}} \) is:

\[
    \Delta t_{\text{pass}}(t_{\text{sa}} + \frac{v_{\text{max}}}{a_{\text{brake}}}, t_{\text{sa}}) = t_{\text{sa}} + \frac{v_{\text{max}}}{a_{\text{acc}}} - \left( t_{\text{sa}} + \frac{v_{\text{max}}}{2} \left( \frac{1}{a_{\text{acc}}} - \frac{1}{a_{\text{brake}}} \right) \right)
\]
\[
= \frac{v_{\text{max}}}{2} \left( \frac{1}{a_{\text{acc}}} + \frac{1}{a_{\text{brake}}} \right)
\]  
(C.18)

The point of time when the first reference point is passed \( t_{x:\text{pass:min}} \) so that the second reference is passed at the earliest point of time is part of the second phase. The following condition thus has to be satisfied:

\[
\frac{\delta t_{\text{pass}}}{\delta t_x} = t_{x:\text{pass:min}} \frac{2a_{\text{brake}}(a_{\text{brake}} - a_{\text{acc}})}{2v_{\text{max}}a_{\text{acc}}} + \frac{a_{\text{brake}} t_{\text{sa}}(a_{\text{acc}} - a_{\text{brake}}) + v_{\text{max}}a_{\text{acc}}}{v_{\text{max}}a_{\text{acc}}} = 0
\]

Consequently, the point of time when the first reference point is passed \( t_{x:\text{pass:min}} \) so that the second reference point is passed at the earliest point of time can be expressed by:

\[
    t_{x:\text{pass:min}} = t_{\text{sa}} + \frac{v_{\text{max}}a_{\text{acc}}}{a_{\text{brake}}(a_{\text{brake}} - a_{\text{acc}})}
\]  
(C.19)

Using \( t_x = t_{x:pass:min} \) as input in equation [C.12], the earliest point of time when the second reference point \( t_{\text{pass:min}} \) is passed can be expressed as:

\[
    t_{\text{pass:min}} = t_{\text{sa}} + \frac{v_{\text{max}}(a_{\text{brake}} - 2a_{\text{acc}})}{a_{\text{brake}}(a_{\text{brake}} - a_{\text{acc}})}
\]  
(C.20)
C.3.2 Example of a signalling system with direct reacceleration

One example illustrates the derived equations for a typical situation. The running time from the first reference point (the initial braking point) to the second reference point (point where the train which stopped at the main signal achieves maximum speed) and the passing time at the second reference point in dependence of the point of time when the first reference point is passed, are visualised in an example. The following parameters, which are typical of a Swiss Intercity train, are assumed:

\[ t_{sa} = 0, \quad v_{max} = 140 \text{ km/h}, \quad a_{brake} = -0.6 \text{ m/s}^2 \quad \text{and} \quad a_{acc} = 0.2 \text{ m/s}^2. \]

The total track length considered \( s_{total} \) between the two reference points in this example is 5041 m. The running time in dependence of the point of time when the first reference point is passed is visualised for the track length considered in Figure C.2. Point A in the figure represents the boundary of the first and the second case. Trains arriving earlier than A halt at the closed main signal. The linear relationship between running time and the point of time when the first reference point is passed for trains earlier than A is observable as defined in equation C.8 and can be explained by the waiting at the closed main signal. Point B in Figure C.2 is the boundary between case 2 and 3. Trains passing the first reference point later than B are not influenced by the signal. As proved in equation C.13 the running time is thus constant. Between A and B, the running time is a quadratic function depending on the passing time at the first reference point (as derived in equation C.12).

![Figure C.2: Running time between two reference points depending on the point of time when the first reference point is passed for a signal system with continuous information and permitted direct reacceleration after the signal turns to green](image)

The relevant and decisive parameter for train traffic management and operations is the point of time when the second reference point is passed, which is the summation of the point of
time when the first reference point is passed and the running time (equation \[C.14\]). Figure \[C.3\] illustrates the passing time at the second reference point as a function of the passing time at the first reference point.

![Figure C.3: Point of time when second reference point is passed as a function of the point of time when the first reference point is passed for a signal system with loop information](image)

Once again, three cases can be identified. For the first case (trains pass the first reference point earlier than \(A\) and came to a halt at the closed main signal), the passing time at the second reference point is constant and independent of the point of time when the first reference point is passed (as proved in equation \[C.15\]). Trains passing the first reference point late (after \(B\)) have a constant running time and therefore a linear growth at the passing time as determined in equation \[C.16\]. The earliest point of time when the second reference point is passed (point \(C\)) is part of the second case (see equation \[C.19\]).

In this example, the time difference \(\Delta t_{\text{pass}}(A, B)\) at the second reference point between a train which has come to a halt at a closed main signal (point \(A\)) and a train that can hold its maximum speed and passes the first reference point at the point of time when the signal turns to green (point \(B\)) is 64 seconds. Consequently, if a train can be controlled in a way that the first reference point is passed later and without being stopped, the passing time at a second reference point (usually a capacity critical area) could be more than one minute earlier.

Time optimum train control (resulting in the earliest possible passing at the second reference point) requires minor intentional deceleration. In this example, speed has to be reduced to 105\(km/h\) until the signal turns to green and the train can immediately reaccelerate (this corresponds to point \(C\)). Choosing this strategy, the train would pass the second reference point 8 seconds earlier than in case \(B\). Using equation \[C.19\] the first reference point must be passed 16 seconds before the signal turns to green.
Figure C.4 illustrates the consequences of the three different cases A (yellow), B (red) and C (blue) in the time-distance diagram.

![Time Distance Diagram](image)

**C.4 Analysis for a signalling system with fixed approach speed**

**C.4.1 Derivation of running time and passing time equations for a signalling system with fixed approach speed and no direct reacceleration**

In this section, the impacts and effects of a second signalling system in which reacceleration is prevented until the main signal is passed are analysed. This fundamental and common signalling system is widely used by railway companies. Trains have to brake immediately to a predefined approach speed if a closed distant signal is passed. The approach speed has to be held constant until the main signal is passed even if the signal turned green in the meanwhile. Reaccelerating the train with a speed larger than the approach speed before passing the main signal would cause an emergency brake initiated by the train’s safety protection system.

Examples of such an intermittent train protection system are ZUB used by SBB or PZB (DB) [Pac00, Sta87, Suw92, The07]. Additional installations between the distant and the main signal are possible to update the signal information. To analyse the fundamental effects, additional transmission possibilities are not taken into consideration for this basic model. Nevertheless,
the impacts and resulting effects of such additional installations could be derived through the combination of the second and the first signalling system.

For this signalling system, five different cases, illustrated in Figures C.5 and C.6 are possible:

- Case 1: The signal turns to green after the train came to a halt in front of the signal (green).
- Case 2: The signal turns to green while the train is braking from its approach speed to the main signal and after reacceleration, the speed at the main signal is lower than its approach speed (yellow).
- Case 3: The signal turns to green while the train is braking towards the main signal. Reacceleration up to the approach speed is possible before the main signal is passed (orange).
- Case 4: The signal turns to green while the train is braking to its approach speed or when the train holds its approach speed. Therefore, the signal turns to green before the train initiates the final braking to a halt at the main signal (red).
- Case 5: The signal turns to green before the train passes the initial braking point (blue).

Figure C.5: Visualisation of the five different cases for a signalling system without direct reacceleration

The determining influence variable that is analysed by the model for this specific signalling system is the point of time $t_x$ when the train passes the first reference point.
C.4. Analysis for a signalling system with fixed approach speed

Figure C.6: Detailed visualisation of the four different cases influenced by the signal for the slow speed part at the main signal for a signalling system without direct reacceleration

The five cases are distinguished by the point of time when the first reference point is passed:

1. case 1: $t_x < t_{sa} - t_{brake}(v_{max} \rightarrow v_{appr}) - t_{hold}(v_{appr}) - t_{brake}(v_{appr} \rightarrow 0)$

2. case 2: $t_{sa} - t_{brake}(v_{max} \rightarrow v_{appr}) - t_{hold}(v_{appr}) - t_{brake}(v_{appr} \rightarrow 0) \leq t_x < t_{sa} - t_{brake}(v_{max} \rightarrow v_{appr}) - t_{hold}(v_{appr}) - t_{brake}(v_{appr} \rightarrow v_{bnd})$

3. case 3: $t_{sa} - t_{brake}(v_{max} \rightarrow v_{appr}) - t_{hold}(v_{appr}) - t_{brake}(v_{appr} \rightarrow v_{bnd}) \leq t_x < t_{sa} - t_{brake}(v_{max} \rightarrow v_{appr}) - t_{hold}(v_{appr})$

4. case 4: $t_{sa} - t_{brake}(v_{max} \rightarrow v_{appr}) - t_{hold}(v_{appr}) \leq t_x < t_{sa}$

5. case 5: $t_{sa} \leq t_x$

The boundary speed $v_{bnd}$ (which differs in cases 2 and 3) is the speed to which a train is braking and can be reaccelerated in order to pass the main signal exactly at its approach speed. It can be shown, that the speed $v_{bnd}$ must satisfy the following condition:

$$s_{brake}(v_{appr} \rightarrow 0) = s_{brake}(v_{appr} \rightarrow v_{bnd}) + s_{acc}(v_{bnd} \rightarrow v_{appr}) \quad (C.21)$$

The total distance considered is determined by case 1, where trains stop in front of a closed main signal. The distance consists of two parts. The first part is the distance from the distant signal to the main signal $s_{distsig}$. The second part is the distance used to accelerate the train from a halt at the signal to its maximum speed permitted. The overall track length considered can thus be expressed as:

$$s_{total} = s_{distsig} + s_{acc}(0 \rightarrow v_{max})$$
First, the time taken $t_{\text{run}}$ to pass the total track length considered $s_{\text{total}}$ is calculated for the five different cases.

Case 1 consists of five phases. First the train brakes to the approach speed. In the second phase, the approach speed is hold until, third phase, the train brakes to a halt at the main signal. After waiting at the closed signal (phase 4), the train finally reaccelerates in phase 5 to its maximum speed. The position at which the maximum speed is achieved is defined as the second reference point. The running time for case 1 $t_{\text{run}}(\text{case 1})$ can thus be expressed as:

$$t_{\text{run,case 1}} = t_{\text{brake}}(v_{\text{max}} \rightarrow v_{\text{appr}}) + t_{\text{hold,1}}(v_{\text{appr}}) + t_{\text{brake}}(v_{\text{appr}} \rightarrow 0) + t_{\text{wait}} + t_{\text{acc}}(0 \rightarrow v_{\text{max}})$$

The waiting time at the closed signal $t_{\text{wait}}$ can be expressed as:

$$t_{\text{wait}} = t_{\text{sa}} - t_{x} - t_{\text{brake}}(v_{\text{max}} \rightarrow v_{\text{appr}}) - t_{\text{hold,1}}(v_{\text{appr}}) - t_{\text{brake}}(v_{\text{appr}} \rightarrow 0)$$

Consequently, the running time for case 1 is:

$$t_{\text{run,case 1}} = t_{\text{sa}} - t_{x} + t_{\text{acc}}(0 \rightarrow v_{\text{max}})$$ (C.22)

In case 2, the signal turns to green when the train brakes close to the signal. In this case, five subsequent phases are identified. First, the train brakes to its approach speed. The approach speed is then held until in the third phase, the train brakes from its approach speed to a reduced minimum speed $v_{\text{min}}$ until the signal turns to green. In the fourth phase, the train directly reaccelerates and passes the main signal at a speed lower than its approach speed until the maximum speed is reached. The fifth phase consists of the time when the maximum speed is held until the second reference point is passed.

$$t_{\text{run,case 2}} = t_{\text{brake}}(v_{\text{max}} \rightarrow v_{\text{appr}}) + t_{\text{hold,1}}(v_{\text{appr}}) + t_{\text{brake}}(v_{\text{appr}} \rightarrow v_{\text{min}}) + t_{\text{acc}}(v_{\text{min}} \rightarrow v_{\text{max}}) + t_{\text{hold,case 2}}(v_{\text{max}})$$ (C.23)

The most complex case 3 has seven phases. First the train brakes to its approach speed and holds this speed until the target braking for the closed main signal is initiated. The first two phases are therefore equal to cases 1 and 2. The difference to case 2 is that the minimum speed $v_{\text{min}}$ is larger than in case 2 and that in phase 4 reacceleration is stopped when the approach speed is achieved before the main signal is passed. After holding the approach speed until the main signal is passed (phase 5), the train accelerates (phase 6) until the maximum speed is achieved. The last phase, number seven, is the time needed to pass the remaining distance to the second reference point running at maximum speed.
Therefore, the running time for case 3 can be expressed as:

\[
   t_{\text{run case }3} = t_{\text{brake}}(v_{\text{max}} \rightarrow v_{\text{appr}}) + t_{\text{hold case }3}(v_{\text{appr}}) + t_{\text{brake}}(v_{\text{appr}} \rightarrow v_{\text{min}}) + t_{\text{acc}}(v_{\text{min}} \rightarrow v_{\text{appr}}) + t_{\text{hold case }3}(v_{\text{appr}}) + t_{\text{acc}}(v_{\text{appr}} \rightarrow v_{\text{max}}) + t_{\text{hold case }3}(v_{\text{max}})
\] (C.24)

Case 4 consists of four subsequent phases. First the train brakes to its approach speed which is held in a second step until the main signal is passed. After passing the main signal, the train accelerates up to its maximum speed (phase 3) at which point the maximum speed is finally held until the second reference point is passed. The running time for case 4 is thus defined as:

\[
   t_{\text{run case }4} = t_{\text{brake}}(v_{\text{max}} \rightarrow v_{\text{appr}}) + t_{\text{hold case }4}(v_{\text{appr}}) + t_{\text{acc}}(v_{\text{appr}} \rightarrow v_{\text{max}}) + t_{\text{hold case }4}(v_{\text{max}})
\] (C.25)

In case 5, the train is not influenced by the signal and passes the entire section considered at the maximum speed permitted. Consequently, the running time for case 5 is:

\[
   t_{\text{run case }5} = t_{\text{hold case }5}(v_{\text{max}}) = \frac{s_{\text{total}}}{v_{\text{max}}}
\] (C.26)

Assumption 2 is used to specify and identify the influencing parameters for the running and passing times in all five cases.

Firstly, the total observed distance is calculated:

\[
   s_{\text{total}} = s_{\text{distsig}} + s_{\text{acc}}(0 \rightarrow v_{\text{max}}) = s_{\text{distsig}} + \frac{v_{\text{max}}^2}{2a_{\text{acc}}}
\]

Using assumption 2, equation [C.21] to determine \(v_{\text{bnd}}\) can be expressed as:

\[
   -\frac{v_{\text{appr}}^2}{2a_{\text{brake}}} = \frac{v_{\text{bnd}}^2 - v_{\text{appr}}^2}{2a_{\text{brake}}} + \frac{v_{\text{appr}}^2 - v_{\text{bnd}}^2}{2a_{\text{acc}}}
\]

Consequently, the speed \(v_{\text{bnd}}\) is:

\[
   v_{\text{bnd}} = v_{\text{appr}} \sqrt{\frac{a_{\text{brake}}}{a_{\text{brake}} - a_{\text{acc}}}}
\] (C.27)

The speed \(v_{\text{bnd}}\) is used to differentiate the five cases. Using assumption 2, the five cases are:

1. case 1:
   \[
   t_x < t_{sa} + \frac{v_{\text{max}}}{a_{\text{brake}}} - \frac{s_{\text{distsig}} + \frac{v_{\text{max}}^2}{2a_{\text{brake}}}}{v_{\text{appr}}}
   \]

2. case 2:
   \[
   t_{sa} + \frac{v_{\text{max}}}{a_{\text{brake}}} - \frac{s_{\text{distsig}} + \frac{v_{\text{max}}^2}{2a_{\text{brake}}}}{v_{\text{appr}}} \leq t_x < t_{sa} - \frac{v_{\text{appr}} \sqrt{\frac{a_{\text{brake}}}{a_{\text{brake}} - a_{\text{max}}}} - v_{\text{max}}}{a_{\text{brake}}} - \frac{s_{\text{distsig}} + \frac{v_{\text{max}}^2}{2a_{\text{brake}}}}{v_{\text{appr}}}
   \]
3. case 3: 

\[ t_{sa} - \frac{v_{appr} \sqrt{a_{brake}}}{a_{brake}} - v_{max} \leq t_x < t_{sa} - \frac{v_{appr}^2 + v_{max}^2}{v_{appr}} \]

4. case 4: 

\[ t_{sa} - \frac{v_{appr} - v_{max}}{a_{brake}} - \frac{s_{dist} + v_{max}^2}{v_{appr}} \leq t_x < t_{sa} \]

5. case 5: \( t_{sa} \leq t_x \)

In case 1, the running time between the two reference points can be expressed as:

\[ t_{run, case 1} = -t_x + t_{sa} + \frac{v_{max}}{a_{acc}} \] (C.28)

Before the running time is determined for cases 2 and 3, the minimum speed \( v_{min} \), at which the train runs at the point of time when the signal turns to green, is calculated as follows:

\[
\begin{align*}
    v_{min} &= v_{appr} + a_{brake} \left( (t_{sa} - t_x) - \left( t_{brake}(v_{max} \to v_{appr}) - t_{hold,c1}(v_{appr}) \right) \right) \\
    &= v_{appr} + a_{brake} \left( t_{sa} - t_x - \frac{v_{appr} - v_{max}}{a_{brake}} - \frac{s_{dist} - s_{brake}(v_{max} \to 0)}{v_{appr}} \right) \\
    &= v_{appr} + a_{brake} \left( t_{sa} - t_x - \frac{v_{appr} - v_{max}}{a_{brake}} - \frac{s_{dist} + v_{max}^2}{2a_{brake} v_{appr}} \right) \\
    &= -t_x a_{brake} + v_{max} + t_{sa} a_{brake} - \frac{2s_{dist} a_{brake} + v_{max}^2}{2v_{appr}} \\
\end{align*}
\] (C.29)

In an initial step, the running time elements for each phase of case 2 (equation [C.23]) are formulated:

\[
\begin{align*}
    t_{brake}(v_{max} \to v_{appr}) &= \frac{v_{appr} - v_{max}}{a_{brake}} \\
    t_{hold,c}(v_{appr}) &= \frac{s_{dist} - s_{brake}(v_{max} \to v_{appr}) - s_{brake}(v_{appr} \to 0)}{v_{appr}} \\
    &= \frac{s_{dist} + v_{max}^2}{2a_{brake} v_{appr}} = \frac{s_{dist} + v_{max}^2}{2v_{appr} a_{brake}} + \frac{v_{max}^2}{2v_{appr}} \\
    t_{brake}(v_{appr} \to v_{min}) &= \frac{v_{min} - v_{appr}}{a_{brake}} \\
    &= -t_x + t_{sa} + \frac{v_{max} - v_{appr}}{a_{brake}} - \frac{2s_{dist} a_{brake} + v_{max}^2}{2v_{appr} a_{brake}} \\
    t_{acc}(v_{min} \to v_{max}) &= \frac{v_{max} - v_{min}}{a_{acc}} \\
    &= t_x a_{brake} - t_{sa} a_{brake} + \frac{2s_{dist} a_{brake} + v_{max}^2}{2v_{appr} a_{acc}} \\
\end{align*}
\] (C.30)
C.4. Analysis for a signalling system with fixed approach speed

\[ t_{\text{hold case 2}}(v_{\text{max}}) = \frac{s_{\text{brake}}(v_{\text{appr}} \to 0) + s_{\text{acc}}(0 \to v_{\text{max}})}{v_{\text{max}}} \]

\[ - \frac{s_{\text{brake}}(v_{\text{appr}} \to v_{\text{min}}) + s_{\text{acc}}(v_{\text{min}} \to v_{\text{max}})}{v_{\text{max}}} \]

\[ = \frac{v_{\text{min}}^2}{2v_{\text{max}}} \left( \frac{1}{a_{\text{acc}}} - \frac{1}{a_{\text{brake}}} \right) \]

\[ = \left( -t_xa_{\text{brake}} + v_{\text{max}} + ts_{\text{brake}}a_{\text{brake}} - \frac{2s_{\text{dist sig}}a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}}} \right)^2 (a_{\text{brake}} - a_{\text{acc}}) \]

\[ t_{\text{run case 2}} = t_x^2a_{\text{brake}}(a_{\text{brake}} - a_{\text{acc}}) \frac{v_{\text{max}}^2a_{\text{acc}}}{2v_{\text{max}}a_{\text{brake}}} \]

\[ + t_x(a_{\text{brake}} - a_{\text{acc}}) \left( v_{\text{max}}^2 + ts_{\text{brake}}a_{\text{brake}} - \frac{2s_{\text{dist sig}}a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}}} \right) \frac{v_{\text{max}}^2a_{\text{acc}}}{2v_{\text{max}}a_{\text{brake}}} \]

\[ + ts_{\text{brake}} \left( 1 - \frac{a_{\text{brake}}}{a_{\text{acc}}} \right) + \left( \frac{2s_{\text{dist sig}}a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}}a_{\text{acc}}} \right) \frac{v_{\text{max}}^2a_{\text{acc}}}{2v_{\text{max}}a_{\text{brake}}} \]

\[ + \left( v_{\text{max}} + ts_{\text{brake}}a_{\text{brake}} - \frac{2s_{\text{dist sig}}a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}}} \right)^2 (a_{\text{brake}} - a_{\text{acc}}) \] (C.31)

Inserting all running time elements from equations [C.31] into equation [C.23], the running time for case 2 to can be expressed as:

\[ t_{\text{run case 2}} = t_x^2a_{\text{brake}}(a_{\text{brake}} - a_{\text{acc}}) \frac{v_{\text{max}}^2a_{\text{acc}}}{2v_{\text{max}}a_{\text{brake}}} \]

\[ + t_x(a_{\text{brake}} - a_{\text{acc}}) \left( v_{\text{max}}^2 + ts_{\text{brake}}a_{\text{brake}} - \frac{2s_{\text{dist sig}}a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}}} \right) \frac{v_{\text{max}}^2a_{\text{acc}}}{2v_{\text{max}}a_{\text{brake}}} \]

\[ + ts_{\text{brake}} \left( 1 - \frac{a_{\text{brake}}}{a_{\text{acc}}} \right) + \left( \frac{2s_{\text{dist sig}}a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}}a_{\text{acc}}} \right) \frac{v_{\text{max}}^2a_{\text{acc}}}{2v_{\text{max}}a_{\text{brake}}} \]

\[ + \left( v_{\text{max}} + ts_{\text{brake}}a_{\text{brake}} - \frac{2s_{\text{dist sig}}a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}}} \right)^2 (a_{\text{brake}} - a_{\text{acc}}) \] (C.32)

As in case 2, the single elements for each phase of the running time (see equation [C.24]) are defined separately for case 3:

\[ t_{\text{brake}}(v_{\text{max}} \to v_{\text{appr}}) = \frac{v_{\text{appr}} - v_{\text{max}}}{a_{\text{brake}}} \]

\[ t_{\text{hold c1}}(v_{\text{appr}}) = \frac{s_{\text{dist sig}} - s_{\text{brake}}(v_{\text{max}} \to v_{\text{appr}}) - s_{\text{brake}}(v_{\text{appr}} \to 0)}{v_{\text{appr}}} \]

\[ = \frac{s_{\text{dist sig}} + \frac{v_{\text{max}}^2}{2a_{\text{brake}}}}{v_{\text{appr}}} = \frac{s_{\text{dist sig}}}{v_{\text{appr}}} + \frac{v_{\text{max}}^2}{2v_{\text{appr}}a_{\text{brake}}} \]

\[ t_{\text{brake}}(v_{\text{appr}} \to v_{\text{min}}) = \frac{v_{\text{min}} - v_{\text{appr}}}{a_{\text{brake}}} \]

\[ = -t_x + ts_{\text{brake}} \frac{v_{\text{max}} - v_{\text{appr}}}{a_{\text{brake}}} - \frac{2s_{\text{dist sig}}a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}}a_{\text{brake}}} \]
Summing up all elements from equations \[ \text{(C.33)} \]

the total running time for case 3 is:

\[
\begin{align*}
t_{\text{run,case3}} &= t_x^2 \left( \frac{a_{\text{brake}}(a_{\text{brake}} - a_{\text{acc}})}{2v_{\text{appr}}a_{\text{acc}}} \right) \\
+ t_x \left( v_{\text{appr}} - v_{\max} - t_{sa}a_{\text{brake}} + \frac{2s_{\text{dist,sig}}a_{\text{brake}} + v_{\max}^2}{2v_{\text{appr}}a_{\text{acc}}} \right) \left( \frac{a_{\text{brake}} - a_{\text{acc}}}{2v_{\text{appr}}a_{\text{acc}}} \right) \\
+ t_{sa} \left( 1 - \frac{a_{\text{brake}}}{a_{\text{acc}}} \right) + \frac{2s_{\text{dist,sig}}a_{\text{brake}} + v_{\max}^2 - v^2_{\text{appr}}}{2v_{\text{appr}}a_{\text{acc}}} + \frac{v^2_{\text{appr}}}{2a_{\text{acc}}v_{\max}} \\
+ \left( v_{\max} + t_{sa}a_{\text{brake}} - \frac{2s_{\text{dist,sig}}a_{\text{brake}} + v_{\max}^2}{2v_{\text{appr}}a_{\text{acc}}} \right)^2 \left( \frac{a_{\text{brake}} - a_{\text{acc}}}{2v_{\text{appr}}a_{\text{brake}}a_{\text{acc}}} \right) \\
\end{align*}
\]

\[ \text{(C.34)} \]

The running time elements for case 4, introduced in equation \[ \text{(C.25)} \]
can be rewritten as:

\[
t_{\text{brake}}(v_{\max} \rightarrow v_{\text{appr}}) = \frac{v_{\text{appr}} - v_{\max}}{a_{\text{brake}}}
\]
\[ t_{\text{hold}, 4}(v_{\text{appr}}) = \frac{s_{\text{distsig}} - s_{\text{brake}}(v_{\text{max}} \rightarrow v_{\text{appr}})}{v_{\text{appr}}} = \frac{s_{\text{distsig}} - \frac{v_{\text{appr}}^2 - v_{\text{max}}^2}{2a_{\text{brake}}}}{v_{\text{appr}}} \]

\[ t_{\text{acc}}(v_{\text{appr}} \rightarrow v_{\text{max}}) = \frac{v_{\text{max}} - v_{\text{appr}}}{a_{\text{acc}}} \]

\[ t_{\text{hold}, case 4}(v_{\text{max}}) = \frac{s_{\text{acc}}(0 \rightarrow v_{\text{max}}) - s_{\text{brake}}(v_{\text{appr}} \rightarrow v_{\text{max}})}{v_{\text{max}}} = \frac{v_{\text{appr}}^2}{2a_{\text{acc}}v_{\text{max}}} \] (C.35)

The running times in case 4 for each phase defined in equation C.35 are summed up and can be expressed as:

\[ t_{\text{run case 4}} = \frac{v_{\text{appr}} - v_{\text{max}}}{a_{\text{brake}}} + \frac{s_{\text{distsig}} - \frac{v_{\text{appr}}^2 - v_{\text{max}}^2}{2a_{\text{brake}}}}{v_{\text{appr}}} + \frac{v_{\text{max}} - v_{\text{appr}}}{a_{\text{acc}}} + \frac{v_{\text{appr}}^2}{2a_{\text{acc}}v_{\text{max}}} \]

\[ = \frac{s_{\text{distsig}}}{v_{\text{appr}}} + \frac{(v_{\text{appr}} - v_{\text{max}})^2}{2v_{\text{appr}}a_{\text{brake}}} + \frac{2v_{\text{max}}^2 - 2v_{\text{max}}v_{\text{appr}} + v_{\text{appr}}^2}{2v_{\text{max}}a_{\text{acc}}} \] (C.36)

In case 5, the running time is written as:

\[ t_{\text{run case 5}} = \frac{s_{\text{total}}}{v_{\text{max}}} = \frac{s_{\text{distsig}} + \frac{v_{\text{max}}^2}{2a_{\text{acc}}}}{v_{\text{max}}} = \frac{s_{\text{distsig}}}{v_{\text{max}}} + \frac{v_{\text{max}}}{2a_{\text{acc}}} \] (C.37)

The point of time when the second reference point is passed is the summation of the point of time when the first reference point is passed with the running time:

\[ t_{\text{pass}} = t_x + t_{\text{run}} \]

Consequently, the passing time at the second reference point for case 1 is:

\[ t_{\text{pass case 1}} = t_s + \frac{v_{\text{max}}}{a_{\text{acc}}} \] (C.38)

For case 2, the time when the second reference point is passed is:

\[ t_{\text{pass case 2}} = \frac{a_{\text{brake}}(a_{\text{brake}} - a_{\text{acc}})}{2v_{\text{max}}a_{\text{acc}}} \]

\[ + t_x \left[ \frac{(a_{\text{brake}} - a_{\text{acc}})(v_{\text{max}}^2 + 2s_{\text{distsig}}a_{\text{brake}} - 2t_s v_{\text{appr}}a_{\text{brake}})}{2v_{\text{max}}v_{\text{appr}}a_{\text{acc}}} + 1 \right] \]

\[ + t_s \left( 1 - \frac{a_{\text{brake}}}{a_{\text{acc}}} \right) + \left( \frac{2s_{\text{distsig}}a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}}a_{\text{acc}}} \right) \]

\[ + \left( \frac{v_{\text{max}} + t_s a_{\text{brake}} - \frac{2s_{\text{distsig}}a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}}}}{2v_{\text{max}}a_{\text{acc}}a_{\text{brake}}} \right) \] (C.39)
The point of time when the reference point is passed for case 3 can be expressed as:

\[ t_{\text{pass,case3}} = t_x^2 \left( \frac{a_{\text{brake}}(a_{\text{brake}} - a_{\text{acc}})}{2v_{\text{appr}} a_{\text{acc}}} \right) + t_x \left( \frac{v_{\text{appr}} - v_{\text{max}} - t_s a_{\text{brake}}}{2v_{\text{appr}} a_{\text{brake}}} + \frac{2s_{\text{distsig}} a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}} a_{\text{acc}}} \right) + 1 + t_s \left( 1 - \frac{a_{\text{brake}}}{a_{\text{acc}}} \right) + \frac{2s_{\text{distsig}} a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}} a_{\text{acc}}} + \frac{v_{\text{appr}}}{2a_{\text{acc}} v_{\text{max}}} + \left( v_{\text{max}} + t_s a_{\text{brake}} - \frac{2s_{\text{distsig}} a_{\text{brake}} + v_{\text{max}}^2}{2v_{\text{appr}}} \right) \left( \frac{a_{\text{brake}} - a_{\text{acc}}}{2v_{\text{appr}} a_{\text{brake}} a_{\text{acc}}} \right) \]  

(C.40)

The passing time at the second reference point for case 4 is:

\[ t_{\text{pass,case4}} = t_x + \frac{s_{\text{distsig}}}{v_{\text{appr}}} + \frac{(v_{\text{appr}} - v_{\text{max}})^2}{2v_{\text{appr}} a_{\text{brake}}} + \frac{2v_{\text{max}}^2 - 2v_{\text{max}}v_{\text{appr}} + v_{\text{appr}}^2}{2v_{\text{max}} a_{\text{acc}}} \]  

(C.41)

Finally, the point of time when the second reference point is passed for case 5 is:

\[ t_{\text{pass,case5}} = t_x + \frac{s_{\text{distsig}}}{v_{\text{max}}} + \frac{v_{\text{max}}}{2a_{\text{acc}}} \]  

(C.42)

In contrast to the signalling system where reacceleration is allowed after the signal turns to green, discontinuities in the function of the passing time occur for the signalling system without direct reacceleration. Firstly, the discontinuity between passing the observed section immediately after the signal turned to green \((t_x = t_{\text{sa}}^-)\) with the case when the distant signal is just passed shortly before the signal turned to green \((t_x = t_{\text{sa}}^+)\), is analysed. The resulting time difference between these two cases at the second reference point can be expressed as:

\[ \Delta t(t_{\text{sa}}^-, t_{\text{sa}}^+) = t_{\text{pass}}(\text{case4}, t_x = t_{\text{sa}}^-) - t_{\text{pass}}(\text{case5}, t_x = t_{\text{sa}}^+) \]

\[ = \frac{s_{\text{distsig}}}{v_{\text{appr}}} + \frac{(v_{\text{appr}} - v_{\text{max}})^2}{2v_{\text{appr}} a_{\text{brake}}} + \frac{2v_{\text{max}}^2 - 2v_{\text{max}}v_{\text{appr}} + v_{\text{appr}}^2}{2v_{\text{max}} a_{\text{acc}}} + t_{\text{sa}}^- - \frac{s_{\text{distsig}}}{v_{\text{max}}} - \frac{v_{\text{max}}}{2a_{\text{acc}}} - t_{\text{sa}}^+ \]  

Based on assumption 1 with a signal watch time down to zero, the difference between the two assumed passing times \(t_{\text{sa}}^-\) and \(t_{\text{sa}}^+\) is negligible. The difference is thus:

\[ \Delta t(t_{\text{sa}}^-, t_{\text{sa}}^+) = s_{\text{distsig}} \left( \frac{1}{v_{\text{appr}}} - \frac{1}{v_{\text{max}}} \right) + (v_{\text{appr}} - v_{\text{max}})^2 \left( \frac{1}{2v_{\text{appr}} a_{\text{brake}}} + \frac{1}{2v_{\text{max}} a_{\text{acc}}} \right) \]  

(C.43)
The second relevant time difference to be determined is between a train which passes the section without being affected by the signal (case 5) with a train that has come to a halt at the closed main signal (case 1). Consequently, the time difference between these two cases at the second reference point is:

\[
\Delta t \left( t_{sa} + \frac{v_{\text{max}}}{a_{\text{brake}}} - \frac{s_{\text{distsig}} + \frac{v_{\text{max}}^2}{2a_{\text{brake}}}}{v_{\text{appr}}}, t_{sa} \right) = t_{\text{pass}} \left( \text{case 1, } t_x = t_{sa} + \frac{v_{\text{max}}}{a_{\text{brake}}} - \frac{s_{\text{distsig}} + \frac{v_{\text{max}}^2}{2a_{\text{brake}}}}{v_{\text{appr}}} \right)
\]

\[
= -t_{\text{pass}} \left( \text{case 5, } t_x = t_{sa} \right)
\]

\[
= \frac{v_{\text{max}}}{a_{\text{acc}}} \left( \frac{v_{\text{appr}}}{a_{\text{brake}}} - \frac{s_{\text{distsig}} + \frac{v_{\text{max}}^2}{2a_{\text{brake}}}}{v_{\text{appr}}} \right) - t_{sa}
\]

\[
= \frac{v_{\text{max}}}{2a_{\text{acc}}} - \frac{s_{\text{distsig}}}{v_{\text{max}}}
\]

(C.44)

A third interesting case worth analysing is the point of time when the signal turns green to when a train runs at approach speed \(v_{\text{appr}}\) and is just on the point of braking. This case represents the boundary of the two cases 3 and 4. The time difference between this point and the case when a train is not hindered by a signal is determined by:

\[
\Delta t \left( t_{sa} - \frac{v_{\text{appr}} - v_{\text{max}}}{a_{\text{brake}}} - \frac{s_{\text{distsig}} + \frac{v_{\text{max}}^2}{2a_{\text{brake}}}}{v_{\text{appr}}}, t_{sa} \right)
\]

\[
= t_{\text{pass}} \left( \text{case 4, } t_x = t_{sa} - \frac{v_{\text{appr}} - v_{\text{max}}}{a_{\text{brake}}} - \frac{s_{\text{distsig}} + \frac{v_{\text{max}}^2}{2a_{\text{brake}}}}{v_{\text{appr}}} \right)
\]

\[
= -t_{\text{pass}} \left( \text{case 5, } t_x = t_{sa} \right)
\]

\[
= \frac{s_{\text{distsig}}}{v_{\text{appr}}} + \frac{(v_{\text{appr}} - v_{\text{max}})^2}{2v_{\text{appr}}a_{\text{brake}}} + \frac{2v_{\text{max}}^2 - 2v_{\text{max}}v_{\text{appr}} + v_{\text{appr}}^2}{2v_{\text{max}}a_{\text{acc}}} + t_{sa}
\]

\[
- \frac{v_{\text{appr}} - v_{\text{max}}}{a_{\text{brake}}} - \frac{s_{\text{distsig}} + \frac{v_{\text{max}}^2}{2a_{\text{brake}}}}{v_{\text{appr}}}
\]

\[
- \frac{s_{\text{distsig}}}{v_{\text{max}}} - \frac{v_{\text{max}}}{2a_{\text{acc}}} - t_{sa}
\]

\[
= \frac{(v_{\text{max}} - v_{\text{appr}})^2}{2v_{\text{max}}a_{\text{acc}}} - \frac{s_{\text{distsig}}}{v_{\text{max}}} - \frac{v_{\text{appr}}}{2a_{\text{brake}}}
\]

(C.45)

Analysing these time differences, it is first shown that there is a discontinuity which does not occur for a signalling system with direct reacceleration. Also, the point where the second reference point is passed at the earliest point of time is not fixed as for the first signalling system, but is either at the start of case 5 (train is not affected by the signal) or when the train brakes and reaccelerates close to the main signal (case 2 and 3).
C.4.2 Example of a signalling system with a fixed approach speed and without any direct reacceleration

The running time and consequential delay effects for the point of time when a train passes a signal is visualised exemplarily for a signalling system with a fixed approach speed. The following parameters, which are typical of a Swiss intercity train, are assumed:

\[ t_{sa} = 0, \quad v_{\text{max}} = 140 \text{ km/h}, \quad a_{\text{brake}} = -0.6 \text{ m/s}^2, \quad a_{\text{acc}} = 0.2 \text{ m/s}^2, \quad v_{\text{appr}} = 40 \text{ km/h} \text{ and } s_{\text{distsig}} = 1400 \text{ m}. \]

The distance between the two reference points (the section considered) is 5180 m for these specific parameters. The running time for this section depending on the point of time when the first reference point is passed is visualised in Figure [C.7]. Trains passing the first reference point after \( D \) are not influenced by the signal. Consequently, the running time is constant for times passing the first reference point later than \( D \). Trains which pass the first reference point between \( B \) and \( C \) also have a constant running time. This occurs when the signal turns to green and the train brakes or holds its approach speed. Because reacceleration to a speed larger than the approach speed is not allowed until the main signal is passed, there is a constant running time. This impossibility to reaccelerate is the reason for the major discontinuity between \( C \) and \( D \).

![Figure C.7: Running time between two reference points depending on the point of time when the first reference point is passed for a signal system without continuous information and no reacceleration until the main signal is passed.](image)

Trains passing the first reference point so early that a halt at the main signal cannot be avoided (earlier than \( A \)) have a linear dependence to the passing time at the first reference point which can be explained by the waiting time at the closed main signal. Trains passing the first reference
point between $A$ and $B$ have to brake from the approach speed to a lower speed, but these trains will not be forced halt at the main signal because the signal turns to green during braking. As derived, there are two different cases between $A$ and $B$ and both are of quadratic impact depending on the passing time at the first reference point.

The point of time when the second reference point is passed depending on the passing time at the first reference point is illustrated in Figure C.8. Cases with constant running time (between $B$ and $C$ and after $D$) now have a linear growth. All trains stopping at the main signal (passing the first reference point earlier than $A$) have the same arriving time at the second reference point. It can also be observed that there is a point of time a little bit earlier than $B$ at which the train arrives at the second reference point earlier. Usually, its approach speed is low and the temporal difference is only a few seconds or even less.

For this specific example, the discontinuity between $C$ and $D$ causes a time difference of 82 seconds. Thus, a train which passes the distant signal just before the section finally becomes available has an additional delay of 82 seconds. As specified in equation C.44, this delay can even be larger and up to several minutes for heavier or faster trains, longer distances between distant and main signal or slower approach speeds. A train which passes the first reference point more than 78 seconds earlier than the signal turns to green (earlier or equal to $A$) will halt at the main signal. Such a train will finally pass the second reference point 61 seconds later compared to the train which passes the first reference point just after the signal turned to green ($D$). The unaffected train ($D$) also reaches the second reference point 20 seconds earlier compared to the train in case $B$. 

![Figure C.8: Point of time second reference point is passed in function of the point of time when the first reference point is passed for a signal system without any loop information and no reacceleration before the main signal is passed](image-url)
Figure C.9 illustrates the consequences of the four different cases A (green), B (orange), C (red) and D (blue) in the time-distance diagram.

![Time distance visualisation for the four specific cases for a signalling system with a fixed approach speed and no direct reacceleration](image)

**C.5 Other signalling systems**

In section C.3 and C.4 the influence is determined for two specific and principal signalling systems. More signalling types are available in railway operations. One example is the old, but widely used Signum system in Switzerland. Passing a closed distant signal causes a warning signal and drivers will slow down. Reacceleration of the train is possible as soon as the main signal can be observed and the signal has turned to green. Therefore, the transfer function for both running time and passing time C.1 and C.2 to determine the additional delay and the passing time at a second reference point can be determined by combining the two principal models. For early trains, the behaviour is similar to the signalling system with a fixed approach speed (having a non-continuous step in the function) whereas for late trains the behaviour is equal to the signalling system where direct reacceleration is possible. The boundary point between the two modes is when the train is at the position of the signal view point and the signal then turns to green.
Appendix D

Energy savings through integrated real-time rescheduling

D.1 Introduction

Reducing the number of avoidable train influences (not only train stops) may be used to save traction energy. The following section is used to present a short overview and provides some sample results.

D.2 Notation

- $E_{el,\text{total}}$: Total electrical energy used for a train movement
- $E_{el,\text{infl}}$: Electrical traction energy used for a train influenced by a signal
- $E_{el,\text{lost}}$: Lost energy due to an influenced train compared with train movement at a constant speed
- $F_{el}$: Electrical traction force
- $F_{el,\text{acc}}$: Accelerating traction force
- $F_{el,\text{brake}}$: Braking force using an electric brake
- $F_{el,\text{hold}}$: Traction force to hold a constant speed
- $\eta$: Energy efficiency factor
- $\eta_{\text{rec}}$: Efficiency factor of the electric regenerative brake
- $v(x)$: Actual speed
- $v_{\text{init}}$: Initial speed before a train is influenced by a signal (also the speed at the end of reacceleration)
- $v_{\text{red}}$: Minimum speed obtained when braking action is taken due to a closed signal
- $s_0$: Starting position
\[ s_e \] End position
\[ s_{\text{red}} \] Distance train runs at a minimum speed \( v_{\text{red}} \)
\[ a_{\text{acc}} \] Constant acceleration factor
\[ a_{\text{brake}} \] Constant braking factor

### D.3 Analysis for energy savings by reducing a train’s influences

As introduced in section 6.4.2.4.2, the energy usage \( E_{\text{el, total}} \) for a train movement between \( s_0 \) and \( s_e \) can be approximated by:

\[
E_{\text{el, total}} = \int_{s_0}^{s_e} \frac{1}{\eta} F_{\text{el}}(x)v(x)dx
\]

with the train’s actual speed \( v \), electrical tractive force \( F_{\text{el}} \) and the constant efficiency factor \( \eta \).

The energy used for an influenced train \( E_{\text{el, infl}} \) (with the phases braking, holding reduced speed and reacceleration) is defined as:

\[
E_{\text{el, infl}} = \int_{s_0}^{s_1} \eta_{\text{rec}} F_{\text{el, brake}}(x)v(x)dx + \int_{s_1}^{s_2} \frac{1}{\eta} F_{\text{el, hold}}(x)v_{\text{red}}dx + \int_{s_2}^{s_e} \frac{1}{\eta} F_{\text{el, acc}}(x)v(x)dx
\]

with the recuperation factor \( \eta_{\text{rec}} \) and the minimum speed \( v_{\text{red}} \) (which could be zero in the case of a halt).

The distance \( s_{\text{red}} = s_2 - s_1 \) is passed at a constant reduced minimum speed, and the initial speed (which is the target speed after reacceleration) is defined as \( v_{\text{init}} \). Assuming that initial and end speed are similar, the energy lost due to braking \( E_{\text{el, lost}} \) is determined as:

\[
E_{\text{el, lost}} = \int_{s_0}^{s_1} \eta_{\text{rec}} F_{\text{el, brake}}(x)v(x)dx + \int_{s_1}^{s_2} \frac{1}{\eta} F_{\text{el, hold}}(x)v_{\text{red}}dx + \int_{s_2}^{s_e} \frac{1}{\eta} F_{\text{el, acc}}(x)v(x)dx
\]

\[- \int_{s_0}^{s_e} \frac{1}{\eta} F_{\text{el, hold}}(x)v_{\text{init}}dx
\]

### D.4 Potential energy saving

The amount of energy saved by avoiding a halt or influence (slow-down) due to a closed signal is evaluated for selected trains. It is assumed that trains have to slow down with high rates (as when normally passing a closed distant signal, thus also pneumatic brakes will have to be used), trains are running on a horizontal line and acceleration follows braking directly (no period at a constant speed). The evaluations are based on a microscopic simulation with OpenTrack [Hür02] with discrete time steps of 1 second. Evaluations were done for an Intercity train, a
commuter train and a typical freight train. For all trains, an efficiency factor for both, traction and recuperation is assumed of $\eta = \eta_{rec} = 0.85$.

Of course, for trains running in hilly topographies, savings may even be significantly higher. Also for older vehicles and locomotives with permanently combined pneumatic and electric braking, the recuperation rate is smaller and energy savings would also be larger.

### D.4.1 Potential energy savings for an IC2000 train

IC2000 trains are a composition of double-decker Intercity coaches and a Re460. The coaches were developed by Schindler Waggon and have a permissible top speed of 200 km/h. The Re460 is a modern four-axle electric locomotive developed by ABB and SLM [Sch98]. The Re460 was originally used for both, freight and passenger trains. However, the SBB mainly uses the Re460 for passenger trains nowadays. Re460 locomotives have a maximum tractive effort of 300kN and 6100kW continuous power output. The top speed is 230 km/h. The locomotive has a weight of 84t and a normal load is around 700t. IC2000 trains are nowadays the backbone of the Swiss Intercity network.

For the simulation, a train load of 700t, a constant braking factor of $-0.5 \text{ m/s}^2$; and a constant maximum electrical braking force of 150kN were assumed.

<table>
<thead>
<tr>
<th>$v_{red}$ (km/h)</th>
<th>$v_{init} = 200\text{km/h}$</th>
<th>$v_{init} = 120\text{km/h}$</th>
<th>$v_{init} = 60\text{km/h}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>90.1 KWh</td>
<td>26.4 KWh</td>
<td>-</td>
</tr>
<tr>
<td>60</td>
<td>108.9 KWh</td>
<td>66.3 KWh</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>123.4 KWh</td>
<td>87.3 KWh</td>
<td>25.6 KWh</td>
</tr>
</tbody>
</table>

![Figure D.1: IC2000 Intercity train with a Re460 locomotive. (Source SBB AG)](image)

D.4.2 Potential energy savings for a FLIRT train

The FLIRT is a commuter train developed by Stadler Rail. The FLIRT’s typical characteristics are high acceleration, strong braking and low weight. FLIRT trains used in Switzerland by the SBB have a maximum tractive effort of 200 kN and 2,000 kW continuous power output. The train is 74 meters long, has a service weight of 120 tons and a top speed of 160 km/h [Sta04a].

For the simulation, a constant braking factor of $-0.8 \text{ m/s}^2$; and a constant maximum electrical braking force of 100kN (which is just sufficient to not need to use the pneumatic brakes) were assumed.

<table>
<thead>
<tr>
<th>$v_{\text{red}}$</th>
<th>$v_{\text{init}} = 120\text{km/h}$</th>
<th>$v_{\text{init}} = 80\text{km/h}$</th>
<th>$v_{\text{init}} = 60\text{km/h}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$60\text{km/h}$</td>
<td>$3.84\text{KWh}$</td>
<td>$0.45\text{KWh}$</td>
<td>-</td>
</tr>
<tr>
<td>$0\text{km/h}$</td>
<td>$4.48\text{KWh}$</td>
<td>$2.87\text{KWh}$</td>
<td>$1.56\text{KWh}$</td>
</tr>
</tbody>
</table>

Figure D.2: FLIRT commuter train. (Source Stadler Rail AG)
D.4. Potential energy saving

D.4.3 Potential energy savings for a typical freight train

The freight trains used in the North-South corridor with combined traffic are typical in Switzerland. These trains are often towed by one or multiple Re 482 (also known as a BR185 in Germany) locomotives developed by Bombardier with a continuous power output of 5600 kW and a tractive effort of 300 kN. Freight trains are operated in Switzerland at a maximum speed of 120 km/h.

For the simulation, one locomotive, a train load of 1400t, a constant braking factor of $-0.4\, m/s^2$, and a constant maximum electrical braking force of 150kN were assumed.

Table D.3: Overview of possible energy savings for a typical freight train

<table>
<thead>
<tr>
<th>$v_{\text{red}}$</th>
<th>$v_{\text{init}} = 120, km/h$</th>
<th>$v_{\text{init}} = 80, km/h$</th>
<th>$v_{\text{init}} = 60, km/h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 km/h</td>
<td>110.0 KWh</td>
<td>39.0 KWh</td>
<td>-</td>
</tr>
<tr>
<td>0 km/h</td>
<td>140.7 KWh</td>
<td>84.7 KWh</td>
<td>52.25 KWh</td>
</tr>
</tbody>
</table>

Figure D.3: Freight train with a Re482 locomotive. (Source SBB AG)
Appendix E

Rescheduling simulation scenarios

E.1 Overview of Lucerne area simulation scenarios for rescheduling point of time impacts

E.1.1 Planned schedule

Lucerne is a dead-end station connected with two lines heading to four main directions. Figure E.1 illustrates in detail the topology of the condensation area around Lucerne’s rail terminal. Due to the limited capacity in the condensation area of Lucerne, all train movements, routes and track occupation between the four switch areas (FMUE, GTS, HEIM, VBHF) are precisely planned. Initially planned timetable and routes, as basis for the rescheduling study, are listed below in Table E.1.

Figure E.1: Topology of the condensation area around Lucerne’s rail terminal
Table E.1: Planned routes for the original timetable

<table>
<thead>
<tr>
<th>Train ID</th>
<th>From</th>
<th>To</th>
<th>FMUE - GTS</th>
<th>GTS - HEIM</th>
<th>HEIM - VBHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>2518</td>
<td>LZ</td>
<td>EBR</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3320</td>
<td>LZ</td>
<td>LIT</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21933</td>
<td>EBR</td>
<td>LZ</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>21938</td>
<td>LZ</td>
<td>EBR</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3311</td>
<td>LIT</td>
<td>LZ</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2517</td>
<td>EBR</td>
<td>LZ</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3568</td>
<td>LZ</td>
<td>EBR</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>21135</td>
<td>ROTS</td>
<td>LZ</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2328</td>
<td>LZ</td>
<td>ROTS</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>21339</td>
<td>LZ</td>
<td>WZB</td>
<td>-</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>111</td>
<td>EBR</td>
<td>LZ</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21638</td>
<td>LZ</td>
<td>LIT</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2410</td>
<td>WZB</td>
<td>LZ</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21838</td>
<td>LZ</td>
<td>EBR</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>111</td>
<td>LZ</td>
<td>ROTS</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>90814</td>
<td>LZ</td>
<td>ROTS</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2321</td>
<td>ROTS</td>
<td>LZ</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

One of the main principles of improving traffic flow in condensation areas is that any slowing down of trains is avoided whenever possible. Nevertheless, to increase infrastructure utilisation, train 3320 is intentionally slowed down in the schedule between GTS and FMUE. The reason for slowing it down is the interdependence between trains 3311 and 3320 at HEIM and FMUE. A later departure of train 3320 would avoid the slowdown between GTS and FMUE. However, this would result in train 3311 applying its brakes and also for other subsequent trains in HEIM.

Speeding up or slowing down trains heading toward the bottleneck area are common rescheduling measures. However, speeding trains up is only possible if sufficient running time supplements are available between the last scheduled stop and the planned passing time of the condensation border in Lucerne. Running time simulations were executed with OpenTrack to identify the latest possible point of time when trains have to be speed up in order to catch an earlier slot. A unique distribution of the running time supplements along the line was also assumed (see also Figure E.2). The running time analysis showed, that trains 111 (1 minute), 2321 (1 minute), 2410 (2 minutes), 2517 (2 minutes), and 3311 (1 minute) can be speeded up and can catch an earlier slot.

Slowing down a train can be executed later in comparison to speeding up and is possible for all trains. However, slowing down a train also implies its reacceleration to the desired speed before the condensation area is entered in order to assure maximum traffic flow. Consequently, there is also a latest rescheduling point of time for slowing down trains.
E.1. Overview of Lucerne area simulation scenarios for rescheduling point of time impacts

Figure E.2: Identification of the latest rescheduling point of time to pass the condensation boarder at an earlier or later slot

E.1.2 Rescheduling measures and characteristics

The rescheduling measures as well as their consequences are listed below for each scenario. A sub-scenario corresponds to the latest possible point of time at which the described rescheduling measures can be implemented. After this point of time, the rescheduling performance decreases (e.g. larger knock-on delay). For each scenario, the following characteristics are listed:

1. Train which is speeded up;
2. Changes in train orders;
3. New train routes;
4. Trains intentionally slowed down within the condensation area;
5. Decisive rescheduling measure (causing the earliest rescheduling action);
6. Decisive rescheduling point of time; and
7. Total knock-on delay.
### Appendix E. Rescheduling simulation scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>1.4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Train speed up:</strong></td>
<td>3311: 1 minute</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Reorder of trains:</strong></td>
<td>-</td>
<td>-</td>
<td>3311 before 3320</td>
<td>-</td>
</tr>
<tr>
<td><strong>New train routes:</strong></td>
<td>2518: 3 2 - 2</td>
<td>2518: 3 2 1</td>
<td>3320: 3 2 2</td>
<td>3311: 3 2 2</td>
</tr>
<tr>
<td></td>
<td>3320: 2 - 2 - 2</td>
<td>3320: 2 - 2 - 2</td>
<td>21135: 2 - 1 - 2</td>
<td>21135: 2 - 1 - 2</td>
</tr>
<tr>
<td></td>
<td>3311: 2 - 1 - 1</td>
<td>3311: 2 - 1 - 1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>21135: 1 - 1 - 2</td>
<td>21135: 1 - 1 - 2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Trains slowed down:</strong></td>
<td>3311: GTS - HEIM</td>
<td>2518: GTS - FMUE</td>
<td>3320: VBHF - HEIM</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3320: GTS - FMUE</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Decisive rescheduling measure:</strong></td>
<td>3311: speed-up 1’</td>
<td>2518: Rerouting HEIM</td>
<td>3320: Rerouting VBHF</td>
<td>3311: Rerouting HEIM</td>
</tr>
<tr>
<td><strong>Decisive rescheduling point of time:</strong></td>
<td>9:54:34</td>
<td>9:54:35</td>
<td>9:58:41</td>
<td>9:59:33</td>
</tr>
<tr>
<td><strong>Total knock-on delay:</strong></td>
<td>1120 sec</td>
<td>300 sec</td>
<td>360 sec</td>
<td>540 sec</td>
</tr>
</tbody>
</table>

Figure E.3: Overview of rescheduling measures for Lucerne scenario 1

<table>
<thead>
<tr>
<th>Scenario</th>
<th>2.1</th>
<th>2.2</th>
<th>2.3</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Train speed up:</strong></td>
<td>2517: 2 minutes</td>
<td>2517: 1 minutes</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Reorder of trains:</strong></td>
<td>2517 before 3311</td>
<td>2517 before 3311</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>New train routes:</strong></td>
<td>3320: 3 2 - 1</td>
<td>3320: 3 2 - 1</td>
<td>3320: 3 2 - 1</td>
<td>21135: 1 - 1 - 2</td>
</tr>
<tr>
<td></td>
<td>21938: 2 - 2 - 2</td>
<td>3311: 2 - 1 - 1</td>
<td>21938: 2 - 2 - 2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3311: 2 - 1 - 1</td>
<td>21135: 2 - 1 - 2</td>
<td>3311: 2 - 1 - 1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>21135: 1 - 1 - 2</td>
<td>2517: 1 - 1 - 2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>21135: 1 - 1 - 2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Trains slowed down:</strong></td>
<td>3320: GTS - FMUE</td>
<td>3311: FMUE - GTS</td>
<td>3320: GTS - FMUE</td>
<td>3320: GTS - FMUE</td>
</tr>
<tr>
<td></td>
<td>3320: GTS - FMUE</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Decisive rescheduling measure:</strong></td>
<td>2517: speed up 2’</td>
<td>2517: speed up 1’</td>
<td>3320: Rerouting HEIM</td>
<td>21135: Rerouting FMUE</td>
</tr>
<tr>
<td><strong>Decisive rescheduling point of time:</strong></td>
<td>9:50:50</td>
<td>9:55:51</td>
<td>9:58:10</td>
<td>10:03:15</td>
</tr>
<tr>
<td><strong>Total knock-on delay:</strong></td>
<td>180 sec</td>
<td>360 sec</td>
<td>480 sec</td>
<td>540 sec</td>
</tr>
</tbody>
</table>

Figure E.4: Overview of rescheduling measures for Lucerne scenario 2

<table>
<thead>
<tr>
<th>Scenario</th>
<th>3.1</th>
<th>3.2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Train speed up:</strong></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Reorder of trains:</strong></td>
<td>21938 before 3320</td>
<td>2517 before 3311</td>
</tr>
<tr>
<td></td>
<td>2517 before 3311</td>
<td>-</td>
</tr>
<tr>
<td><strong>New train routes:</strong></td>
<td>3320: 2 - 2 - 1</td>
<td>21135: 1 - 1 - 2</td>
</tr>
<tr>
<td></td>
<td>21938: 2 - 1 - 1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3311: 3 2 - 2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3317: 3 2 - 2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3568: 2 - 1 - 1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>21135: 1 - 1 - 2</td>
<td>-</td>
</tr>
<tr>
<td><strong>Trains slowed down:</strong></td>
<td>3311: FMUE - GTS</td>
<td>3320: GTS - FMUE</td>
</tr>
<tr>
<td><strong>Decisive rescheduling measure:</strong></td>
<td>3320: reorder after 21135</td>
<td>21135: slow down</td>
</tr>
<tr>
<td><strong>Decisive rescheduling point of time:</strong></td>
<td>9:56:41</td>
<td>10:02:45</td>
</tr>
<tr>
<td><strong>Total knock-on delay:</strong></td>
<td>420 sec</td>
<td>660 sec</td>
</tr>
</tbody>
</table>

Figure E.5: Overview of rescheduling measures for Lucerne scenario 3
E.1. Overview of Lucerne area simulation scenarios for rescheduling point of time impacts

<table>
<thead>
<tr>
<th>Scenario</th>
<th>4.1</th>
<th>4.2</th>
<th>4.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Reorder of trains:</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>New train routes:</td>
<td>2517: 1 - 1 - 2</td>
<td>3558: 2 - 1 - 1</td>
<td>3320: GTS - FMUE</td>
</tr>
<tr>
<td>Trains slowed down:</td>
<td>-</td>
<td>-</td>
<td>21135: slow down</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>21135: slow down</td>
<td>21135: slow down</td>
<td>21135: slow down</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>10:02:45</td>
<td>10:02:45</td>
<td>10:02:45</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>60 sec</td>
<td>60 sec</td>
<td>60 sec</td>
</tr>
</tbody>
</table>

Figure E.6: Overview of rescheduling measures for Lucerne scenario 4

<table>
<thead>
<tr>
<th>Scenario</th>
<th>5.1</th>
<th>5.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Reorder of trains:</td>
<td>21135 before 2517</td>
<td>3558 before 21135</td>
</tr>
<tr>
<td>New train routes:</td>
<td>2517: 3 - 2 - 2</td>
<td>3558: 3 - 2 - 2</td>
</tr>
<tr>
<td>Trains slowed down:</td>
<td>3320: GTS - FMUE</td>
<td>3320: GTS - FMUE</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>2517: reorder after 21135</td>
<td>21135: slow down</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>10:03:15</td>
<td>10:03:45</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>60 sec</td>
<td>300 sec</td>
</tr>
</tbody>
</table>

Figure E.7: Overview of rescheduling measures for Lucerne scenario 5

<table>
<thead>
<tr>
<th>Scenario</th>
<th>6.1</th>
<th>6.2</th>
<th>6.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>2410: 2 minutes</td>
<td>2410: 1 minute</td>
<td>-</td>
</tr>
<tr>
<td>Reorder of trains:</td>
<td>2410 before 111</td>
<td>2410 before 111</td>
<td>2410 before 111</td>
</tr>
<tr>
<td>New train routes:</td>
<td>-</td>
<td>-</td>
<td>111: 3 - 2 - 1</td>
</tr>
<tr>
<td>Trains slowed down:</td>
<td>3320: GTS - FMUE</td>
<td>3320: GTS - FMUE</td>
<td>3320: GTS - FMUE</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>2410: speed up 2°</td>
<td>2410: speed up 1°</td>
<td>111: FMUE - GTS</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>10:01:49</td>
<td>10:07:11</td>
<td>10:09:45</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>0 sec</td>
<td>120 sec</td>
<td>180 sec</td>
</tr>
</tbody>
</table>

Figure E.8: Overview of rescheduling measures for Lucerne scenario 6

<table>
<thead>
<tr>
<th>Scenario</th>
<th>7.1</th>
<th>7.2</th>
<th>7.3</th>
<th>7.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>3311: 1 minute</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Reorder of trains:</td>
<td>3311 before 3320</td>
<td>-</td>
<td>2517 before 21938</td>
<td>-</td>
</tr>
<tr>
<td>New train routes:</td>
<td>3320: 3 - 2 - 2</td>
<td>3320: 3 - 2 - 2</td>
<td>3311: 3 - 2 - 2</td>
<td>3311: 3 - 2 - 2</td>
</tr>
<tr>
<td>Trains slowed down:</td>
<td>-</td>
<td>-</td>
<td>3320: GTS - FMUE</td>
<td>3311: GTS - HEIM</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>3311: speed up 1°</td>
<td>3311: Rerouting FMUE</td>
<td>3311: Rerouting HEIM</td>
<td>2517: slow down</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>9:54:34</td>
<td>9:58:15</td>
<td>9:59:53</td>
<td>10:00:45</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>60 sec</td>
<td>180 sec</td>
<td>300 sec</td>
<td>480 sec</td>
</tr>
</tbody>
</table>

Figure E.9: Overview of rescheduling measures for Lucerne scenario 7
E.2 Overview of Bern area simulation scenarios for rescheduling point of time impacts

E.2.1 Planned schedule

Around every full hour, the station of Bern is fully occupied (see Figure E.11). The limited amount of available station tracks causes, that platform 12 and 13 have to be used by two trains at the same time. The topology and timetable implicit that some Intercity trains arriving from the east have to reverse and depart direction east. This causes longer turnaround times of at least 5 minutes. Connections between Intercity trains are at least 3 minutes and had to be hold in the simulation studies.
E.2. Overview of Bern area simulation scenarios for rescheduling point of time impacts

Speeding up Intercity trains is possible for up to 2 minutes. However, train sequences and headways on feeder lines must be considered. The train RE 3064 from Neuchatel could be speed up such that 1 minute earlier arriving is possible. Commuter trains could not be speeded up. Trains stopping within the condensation area of Bern at Wandkorf, Ausserholigen or Stockacker have a dwell time of 30 - 60 seconds. It is in the simulation analysis implicitly assumed, that scheduled station dwell times are not exceeded, which would cause a further rescheduling.

E.2.2 Rescheduling measures and characteristics

The applied rescheduling measures for all scenarios in the Bern area are summarised in the following. Also the latest possible, decisive rescheduling point of time and the resulting knock-on delay for each scenario is listed.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>1.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>New train routes for arriving trains:</td>
<td>-</td>
<td>16466, 15166</td>
<td>16466</td>
<td>-</td>
</tr>
<tr>
<td>New train routes for departing trains:</td>
<td>2535, 982</td>
<td>835, 979, 982, 830, 15207</td>
<td>979, 982</td>
<td>982, 830</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>Rerouting 2535</td>
<td>Rerouting 835</td>
<td>Rerouting 16466</td>
<td>Rerouting 982</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>17:00:00</td>
<td>17:02:00</td>
<td>17:02:46</td>
<td>17:04:00</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>0 sec</td>
<td>360 sec</td>
<td>540 sec</td>
<td>930 sec</td>
</tr>
</tbody>
</table>

Figure E.12: Overview of rescheduling measures for Bern scenario 1

<table>
<thead>
<tr>
<th>Scenario</th>
<th>2.1</th>
<th>2.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>New train routes for arriving trains:</td>
<td>16466, 15166</td>
<td>-</td>
</tr>
<tr>
<td>New train routes for departing trains:</td>
<td>15065, 835, 979, 982, 830</td>
<td>982</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>Rerouting 15065 and 835</td>
<td>Rerouting 982</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>17:02:00</td>
<td>17:04:00</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>180 sec</td>
<td>210 sec</td>
</tr>
</tbody>
</table>

Figure E.13: Overview of rescheduling measures for Bern scenario 2
### Figure E.14: Overview of rescheduling measures for Bern scenario 3

<table>
<thead>
<tr>
<th>Scenario</th>
<th>3.1</th>
<th>3.2</th>
<th>3.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>982: 1 minute</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>New train routes for arriving trains:</td>
<td>1864, 982, 3230</td>
<td>1864</td>
<td>979</td>
</tr>
<tr>
<td>New train routes for departing trains:</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>Speed up 982</td>
<td>Rerouting 1864</td>
<td>Rerouting 979</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>16:41:12</td>
<td>16:41:51</td>
<td>16:47:10</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>0 sec</td>
<td>30 sec</td>
<td>120 sec</td>
</tr>
</tbody>
</table>

### Figure E.15: Overview of rescheduling measures for Bern scenario 4

<table>
<thead>
<tr>
<th>Scenario</th>
<th>4.1</th>
<th>4.2</th>
<th>4.3</th>
<th>4.4</th>
<th>4.5</th>
<th>4.6</th>
<th>4.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>-</td>
<td>-</td>
<td>2530 30 sec</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>New train routes for arriving trains:</td>
<td>16465, 3230</td>
<td>3230, 979, 2535</td>
<td>3230, 979, 2535</td>
<td>2535, 830, 2530</td>
<td>2535</td>
<td>2535</td>
<td>-</td>
</tr>
<tr>
<td>New train routes for departing trains:</td>
<td>-</td>
<td>16464</td>
<td>16464</td>
<td>16464</td>
<td>16464</td>
<td>2064</td>
<td>3064</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>Rerouting 16465</td>
<td>Rerouting 15164</td>
<td>Rerouting 3330</td>
<td>Speed up 2530</td>
<td>Rerouting 15264</td>
<td>Rerouting 2533</td>
<td>Rerouting 3064</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>17:43:30</td>
<td>17:46:00</td>
<td>17:47:02</td>
<td>17:59:00</td>
<td>17:51:28</td>
<td>17:54:00</td>
<td></td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>100 sec</td>
<td>336 sec</td>
<td>570 sec</td>
<td>630 sec</td>
<td>600 sec</td>
<td>700 sec</td>
<td>0:10 sec</td>
</tr>
</tbody>
</table>

### Figure E.16: Overview of rescheduling measures for Bern scenario 5

<table>
<thead>
<tr>
<th>Scenario</th>
<th>5.1</th>
<th>5.2</th>
<th>5.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>New train routes for arriving trains:</td>
<td>15166</td>
<td>15166</td>
<td>-</td>
</tr>
<tr>
<td>New train routes for departing trains:</td>
<td>979, 830</td>
<td>830</td>
<td>15267</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>Rerouting 979</td>
<td>Rerouting 15166</td>
<td>Rerouting 15267</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>17:04:00</td>
<td>17:06:52</td>
<td>17:12:00</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>0 sec</td>
<td>120 sec</td>
<td>360 sec</td>
</tr>
</tbody>
</table>

### Figure E.17: Overview of rescheduling measures for Bern scenario 6

<table>
<thead>
<tr>
<th>Scenario</th>
<th>6.1</th>
<th>6.2</th>
<th>6.3</th>
<th>6.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>982: 30 sec</td>
<td>2535: 30 sec</td>
<td>2535: 30 sec</td>
<td>2535: 30 sec</td>
</tr>
<tr>
<td>New train routes for arriving trains:</td>
<td>982</td>
<td>982, 3230, 979, 2530</td>
<td>982</td>
<td>835</td>
</tr>
<tr>
<td>New train routes for departing trains:</td>
<td>-</td>
<td>15165, 16465</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>Speed up 982</td>
<td>Rerouting 15165</td>
<td>Rerouting 982</td>
<td>Speed up 2535</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>210 sec</td>
<td>240 sec</td>
<td>270 sec</td>
<td>390 sec</td>
</tr>
</tbody>
</table>
E.3 Overview of Winterthur area simulation scenarios for rescheduling point of time impacts

E.3.1 Planned schedule

The schedule for Winterthur is part of the national integrated fixed-interval timetable. However, passengers normally are not transferring between Intercity trains. Also commuter trains from Zurich’s S-Bahn normally do not wait for late trains due to the high train frequency. For the simulation analysis a time window, starting with the Intercity trains arriving from Zurich, is chosen. The initial schedule with the tracks used by the trains is introduced in Table E.2.

<table>
<thead>
<tr>
<th>Train ID</th>
<th>Route toward station</th>
<th>Station track</th>
<th>Departing route</th>
</tr>
</thead>
<tbody>
<tr>
<td>831</td>
<td>KE - TOEM 924 - WRB 45</td>
<td>5</td>
<td>WNO 128 - OWT</td>
</tr>
<tr>
<td>531</td>
<td>KE - TOEM 124 - WRB 44</td>
<td>4</td>
<td>WNO 436 - WGR</td>
</tr>
<tr>
<td>8170</td>
<td>-</td>
<td>9</td>
<td>WRB 46 - WTOE</td>
</tr>
<tr>
<td>18773</td>
<td>-</td>
<td>6</td>
<td>WRB 43 - TOEM 224 - KE</td>
</tr>
<tr>
<td>19269</td>
<td>KE - TOEM 924 - WRB 45</td>
<td>5</td>
<td>WNO 228 - OWT</td>
</tr>
<tr>
<td>18873</td>
<td>OWT - WNO 228</td>
<td>8</td>
<td>WRB 43 - TOEM 224 - KE</td>
</tr>
<tr>
<td>20370</td>
<td>-</td>
<td>7</td>
<td>WNO 730 - HET</td>
</tr>
<tr>
<td>19673</td>
<td>HET - WNO 730</td>
<td>6</td>
<td>WRB 43 - TOEM 224 - KE</td>
</tr>
</tbody>
</table>

Dwell times are normally 120 seconds, thus delays are not reduced in the station. Connections are not held for delayed trains and also not penalised in this case. The Intercity trains 831 and 531 can be speed-up such that they arrive up to 2 minutes earlier in Winterthur than scheduled. Also, the commuter trains S19269 and S18873 as well as the freight train 30199 can arrive in maximum 1 minute earlier than originally planned. Slowing down a train to arrive later is for all trains possible.
E.3.2 Rescheduling measures and characteristics

The rescheduling measures including decisive time and total knock-on delay for each scenario in the Winterthur area is listed in the following.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>1.4</th>
<th>1.5</th>
<th>1.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reorder of trains:</td>
<td>18873 before 831</td>
<td>18873 before 831</td>
<td>15269 before 18873</td>
<td>19269 before 18873</td>
<td>30199 before 19673</td>
<td>-</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>17:36:50</td>
<td>17:37:03</td>
<td>17:39:00</td>
<td>17:43:00</td>
<td>17:44:22</td>
<td>17:46:30</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>630 sec</td>
<td>720 sec</td>
<td>1200 sec</td>
<td>1290 sec</td>
<td>1300 sec</td>
<td>1740 sec</td>
</tr>
</tbody>
</table>

Figure E.19: Overview of rescheduling measures for Winterthur scenario 1

<table>
<thead>
<tr>
<th>Scenario</th>
<th>2.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>-</td>
</tr>
<tr>
<td>Reorder of trains:</td>
<td>-</td>
</tr>
<tr>
<td>New train routes:</td>
<td>18773: W - TOEM</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>19269: rerouting</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>17:35:50</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>0 sec</td>
</tr>
</tbody>
</table>

Figure E.20: Overview of rescheduling measures for Winterthur scenario 2

<table>
<thead>
<tr>
<th>Scenario</th>
<th>3.1</th>
<th>3.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Reorder of trains:</td>
<td>-</td>
<td>18773: KE - W</td>
</tr>
<tr>
<td>New train routes:</td>
<td>-</td>
<td>18773: KE - TOEM</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>18773: reorder after 19269</td>
<td>30199: rerouting</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>17:37:39</td>
<td>17:40:38</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>300 sec</td>
<td>300 sec</td>
</tr>
</tbody>
</table>

Figure E.21: Overview of rescheduling measures for Winterthur scenario 3

<table>
<thead>
<tr>
<th>Scenario</th>
<th>4.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>-</td>
</tr>
<tr>
<td>Reorder of trains:</td>
<td>-</td>
</tr>
<tr>
<td>New train routes:</td>
<td>19673: W - WGB</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>19673: rerouting</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>17:45:00</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>0 sec</td>
</tr>
</tbody>
</table>

Figure E.22: Overview of rescheduling measures for Winterthur scenario 4
### E.3. Overview of Winterthur area simulation scenarios for rescheduling point of time impacts

<table>
<thead>
<tr>
<th>Scenario</th>
<th>S.1</th>
<th>S.2</th>
<th>S.3</th>
<th>S.4</th>
<th>S.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train speed up:</td>
<td>18873: 1 minute</td>
<td>18873: 30 sec</td>
<td>18873: 30 sec</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Reorder of trains:</td>
<td>18873 before S1</td>
<td>18873 before S1</td>
<td>18873 before S1</td>
<td>18873 before S1</td>
<td>-</td>
</tr>
<tr>
<td>Decisive rescheduling measure:</td>
<td>18873: speed up</td>
<td>19269: rerouting</td>
<td>18873: speed up</td>
<td>831: reorder after 18873</td>
<td>30198: rerouting</td>
</tr>
<tr>
<td>Decisive rescheduling point of time:</td>
<td>17:32:06</td>
<td>17:32:06</td>
<td>17:30:06</td>
<td>17:30:06</td>
<td>17:40:38</td>
</tr>
<tr>
<td>Total knock-on delay:</td>
<td>1240 sec</td>
<td>270 sec</td>
<td>330 sec</td>
<td>360 sec</td>
<td>570 sec</td>
</tr>
</tbody>
</table>

Figure E.23: Overview of rescheduling measures for Winterthur scenario 5
Bibliography


[RAM99] *European Standard EN 50126: Railway applications - The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS).* CENELEC European Committee for Electrotechnical Standardization, Brussels, September 1999.


Bibliography


