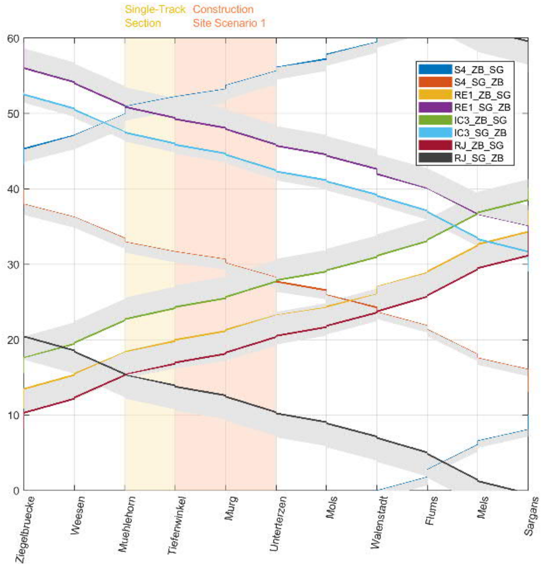
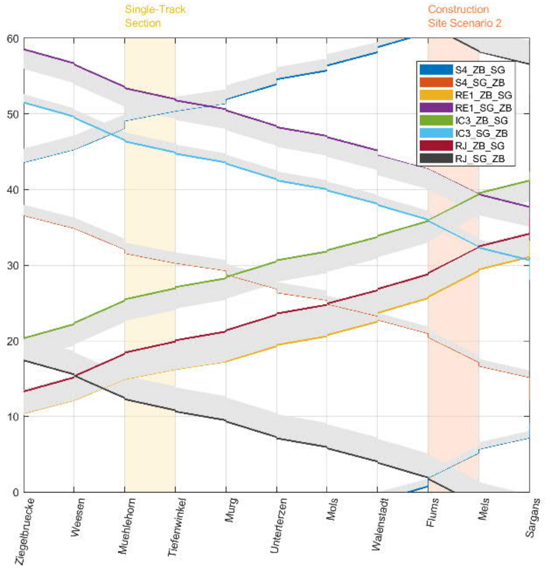
Zurich University
of Applied Sciences

School of
Engineering
IDP Institute of Data Analysis
and Process Design

Development of a prototype for the automated generation of timetable scenarios specified by the transport service intention

Final report





Raimond Wüst, Stephan Bütikofer, Severin Ess,
Claudio Gomez, Albert Steiner

Institute of Data Analysis and Process Design IDP
Zurich University of Applied Sciences ZHAW
8401 Winterthur, Switzerland

Winterthur, 30 September 2018 / updated: 28 June 2019

Imprint

Authors

ZHAW Zurich University of Applied Sciences
IDP Institute of Data Analysis and Process Design
Raimond Wüst*, Stephan Bütikofer, Severin Ess, Claudio Gomez, Albert Steiner
Rosenstrasse 3, PO Box
8401 Winterthur
Switzerland
*: Project leader, corresponding author: raimond.wuest@zhaw.ch

Project partner

Marc Reber
Swiss Federal Railways SBB
SmartRail 4.0, Programm TMS
Haslerstrasse 30
3000 Bern 65
Switzerland

Contracting authority

SBB Lab
Research Fund
Universität St.Gallen
Dufourstrasse 40a
CH-9000 St.Gallen
Switzerland

Project period

1 July 2017 to 28 September 2018

Date of submission of the final report

30 September 2018 / updated: 28 June 2019

Preferred citation style

Wüst, R.M., Bütikofer, St., Ess, S., Gomez, C., Steiner, A. (2019). Development of a prototype for the automated generation of timetable scenarios specified by the transport service intention. Final report (28 June 2019). Zurich University of Applied Sciences, Institute of Data Analysis and Process Design, Winterthur, Switzerland.

(All authors contributed equally to this report)

Management Summary

Within the next five to ten years, public transport in Switzerland as well as in other European countries will experience major technological and organisational changes. However, changes will also take place on the customer side, resulting in different mobility behaviour and demand patterns. These changes will lead to additional challenges for transport service providers in private as well as public domains. Time to market will be a key success factor, and it is unnecessary to mention that due to these factors the speed and flexibility of business processes in freight as well as in passenger transport industry have to be increased significantly.

Within the railway value chain (line planning, timetabling and vehicle scheduling, etc.) the coordination of the individual planning steps is a key success factor. SBB, as the leading service provider in public transport in Switzerland, has recognised this challenge and, together with various partners, initiated the strategic project Smart Rail 4.0. The ZHAW and especially the Institute for Data Analysis and Process Design (IDP) of the School of Engineering wants to be part of this transformation process and to contribute with research and educational activities. The IDP research, therefore, aims for the transformation of academic and scientific know-how to practical applicability. In a first step this concerns directly the current Smart Rail 4.0 TMS-PAS project activities, that concentrate on timetabling issues.

The IDP project team considers the integration of line planning and the timetabling process as crucial for practical applications. To address this in the current research project, we present an application concept that enables the integration of these two major process steps in the transport service value-chain. Although it turns out from our research, that the technical requirements for the integration of the process can be satisfied, rules and conditions for closer cooperation of the involved business units (the train operating companies and the infrastructure operating company), have to be improved and to be worked out in more detail.

In addition to a detailed application concept with use cases for the timetabling process, we propose a methodology for computer-aided timetable generation based on the central planning object known as ‘service intention’. The service intention can be used to iteratively develop the timetable relying on a ‘progressive feasibility assessment’, a feature requested in practice.

Our proposed model is based on the ‘track-choice’ and line rotation extension of the commonly known method for the generation of periodic event schedules ‘PESP’. The extension makes use of the track infrastructure representation, which is also used by the line planning and timetabling system Viriato. Public transport planners and operators widely use this system. With the help of Viriato, it is rather easy to configure the timetabling problem in sufficient detail. On the other side, the level of detail of the considered data is light enough to algorithmically solve practical timetabling problems of realistic sizes.

Taking into consideration the technical and operational constraints given by rolling stock, station and track topology data on the one hand, and the commercial requirements defined by a given line concept on the other, the method presented generates periodic timetables including train-track assignments. In the first step, the standardised data structure ‘service intention’ represents the line concept consisting of train paths and frequencies. Due to the utilisation of infrastructure-based track capacities, we are also able to assess the feasibility of the line concept given. Additionally, the method allows for handling temporary resource restrictions (e.g. caused by construction sites or operational disturbances). In order to assess the performance of the resulting timetable, we present a framework for performance measurement that addresses the customer convenience (in terms of start-to-end travel time) as well as operational stability requirements (in terms of delay sensitivity and critical relations). After the introduction of the methods, prototypes and use cases, we provide a practical proof of concept by successfully testing the framework in different test scenarios.

The methods presented in this report are part of a planning framework, which is currently developed together with the Smart Rail 4.0 project team and which covers significant parts of the railway value chain.

Acknowledgements

We highly acknowledge

- the financial support from the SBB Research Fund, c/o SBB Lab, Universität St.Gallen, Dufourstrasse 40a, 9000 St.Gallen,
- the SmartRail 4.0, Programm TMS Team AFO, consisting of Peter Howald, Thomas Künzi, Dr. Kaspar Schüpbach, Patrick Wild and Thomas Wieland for the very fruitful collaboration and their support of the project with personal resources, and finally
- our former project partners from IBM Research for their very helpful support with the track choice extension of the PESP-model: Dr. Marco Laumanns and Dr. Jacint Szabo.

Table of Contents

1 Introduction

1.1 Background of research

In recent years, the Swiss railway network has been significantly expanded, resulting in remarkable improvements of the public transport service offer. As a consequence, one can observe a strong increase in the load of the railway infrastructure. Operators of railway infrastructures in Switzerland (infrastructure operating company, IOC) are aware of the need for maintenance efforts. Regular intervals for rail infrastructure maintenance and replacement have to be planned at numerous locations in the Swiss railway network. Usually, these maintenance intervals are planned during night hours (around 0 to 5 a.m.) when passenger rail traffic is off-peak. On the other hand, night hours are strongly utilised by cargo traffic on many train lines. In addition, short maintenance intervals during night hours are very inefficient, prolonging the overall maintenance duration and hence are quite expensive.
These are good reasons to expand maintenance intervals and to start working already in the evening hours, e.g. after 10 p.m. instead of midnight. However, this would lead to reduced infrastructure availability for regular evening traffic. Furthermore, the duration of these intervals also has an impact on the timetable, such as increasing turnover times, delayed arrivals or broken connections. An additional effect is the reduced robustness of the corresponding timetable.

Because of this, planners of maintenance intervals and operations have a strong need for rapid development and assessment of comprehensive and reliable timetable scenarios that satisfy the requirements of the train operating company (TOC).

There are numerous approaches for algorithmic support of timetable generation known from academic literature (see section 2.1). Nevertheless, until today, knowledge from academic research still did not migrate into every day's planning procedures of TOC and other companies involved in public transport. This is mainly due to the fact, that the development of a timetable that includes national passenger and freight services on a local, regional, national as well as an international scale has to take political, user-focused, and operation focused aspects into account. Additionally, this highly iterative business process involves several organisations or business units of an organisation, each responsible for different tasks and parameters, that are hard to map down into standard data structures. Standard data structures, however, are required for implementing computer-aided workflows.

In an internal concept paper, a project group of SBB infrastructure comes up with the following additional findings:

- Timetable development is complex, complicated to plan, and hardly supported by the current scheduling system NeTS.
- NeTS was developed from 2007 and has become "old". It only supports manual planning processes.
- Processes, working methods and system support are not up to future requirements.
- The rapid development in the areas of digitisation and automation offer new opportunities.
- Further investments in the planning system can only be justified if new processes, together with further developed functionality enable a massive increase in efficiency.
- The planning philosophy, processes and data flows must be able to meet the requirements of a digitalised railway and customer world in the future.

For these reasons, SBB plans to simplify and coordinate its planning and operating processes and to support them with an integrated and automatable system landscape. In order to achieve this together with other partners from industry, SBB has started the SmartRail 4.0 project. The SmartRail 4.0 TMS-PAS subproject includes the development of a forward-looking, productive timetable development system that will replace the NeTS planning system in the near future. More specifically, as an initial step, a proposal for the “Automation of timetable creation and revision/restoration in the event of disruption" is requested. The resulting timetable is required to adhere to the desired regularity aspect as well to the flexibility aspect that is supposed to address alternating demand and robustness criteria.

1.2 Research Questions

In order to generate a research plan for the project, we formulated the following five research questions (RQ):

- RQ1 – State of the art: What is state of the art regarding automated train scheduling based on functional requirements like the service intention (SI) in literature and practice? What is known from literature about the operational feasibility of timetables generated in an automated or semi-automated way?
- RQ2 – Service Intention (SI) and underlying agreement between TOC and IOC: This research question addresses the aim and the deliverables of the planning process (interval planning, IP), which results mainly in the customer information required for the maintenance time interval and the necessary issues of the service level agreement (SLA) between TOC and IOC. The SmartRail 4.0 TMS-PAS subproject has identified several stakeholders who are affected by strategic process changes (see Appendix A). Since the timetable is the result of a sequential but iterative process involving several planning steps of different TOCs and the IOC, the question specifically concerns the process interface related to the SI: Which organisational units are affected by this process interface? What information must the functional requirement, i.e. the SI, contain in detail?
- RQ3 – Feasibility of algorithmically generated timetables: How can the data structure with the commercial information (included in the SI), and the operational information, which both result from answering RQ1, be merged with the mesoscopic topology in order to automatically generate timetable scenarios with verified operational feasibility? Moreover, how can it be improved in case of insufficient operational reliability? Which level of operational detail regarding rolling stock, turnaround times and safety restrictions like for instance headways have to be considered in use case 1 (UC1, see chapter 1.3.1) of the IP-process? What is the difference between the level of detail contained in timetable scenarios resulting from the mesoscopic model calculation and the microscopic operational production plan, which has to be implemented in the train control system? Are there reliable conditions that have to be met by a certain scenario resulting from mesoscopic timetabling to make sure that the result is feasible also with respect to microscopic operational detail? At what point in time before the actual start of a train trip do we need to add the missing information and how does this relate to the operational information of the timetable?
- RQ4 – Timetable Performance Measurement: Which are the relevant quantitative performance criteria regarding timetable stability, robustness and capacity utilisation in order to evaluate different timetable scenarios?

More specifically, with respect to UC1 and UC2 of IP concerning temporary timetables for maintenance intervals with reduced infrastructure availability, the following research question has to be addressed:

- RQ5 – Use case description for IP-process: Which changes in the operational constraints can be applied without violating the commercial requirements? Which restricted timetable scenarios can be considered as more performant than others and how shall the iterative process to figure out the best scenario look like? How are commercial requirements (e.g. connections or prioritisation between train runs) considered when looking for the best temporary timetable scenario for maintenance intervals?

1.3 Goals of this project

1.3.1 Concept for computer-aided interval planning

The main goal of this project is to develop a concept for computer-aided timetable development, which refers to research questions RQ1 to RQ5. The automated timetable creation and revision/restoration in the event of disruption is an important business requirement for railway companies like SBB. This holds especially for situations, where timetable scenarios have to be developed for maintenance time windows. In these cases, infrastructure components are temporarily unavailable during certain hours of the daily passenger train schedule (either during the late evening or early morning hours or even during several days), and these service restrictions have to be considered in the timetable. From the customer perspective, these restrictions should be as small as possible, but the resulting timetable must still guarantee sufficient operational stability. This means that it should be possible to quickly find and evaluate timetable alternatives with minimal impact on service quality in a fast and easy way. We call this planning process 'interval planning' (IP) and present a computerised method for automated IP-timetabling, which takes into account reduced resource availability, as well as constraints, resulting from operational and commercial requirements of the TOC and the published timetable. Our proposal for the IP business model is based on six use cases denoted as IP-UC0 to IP-UC5:

- IP-UC0 aims to make a selection of possibly (in certain situations of IP) reduced train lines, which has the smallest possible impact on the total travel time of customers concerned by the restrictions of the interval plan. If the resulting line concept is reduced with respect to the original one, we call it a relaxed service intention. In these cases, the customers have to be informed, that maybe they have to change their travel plans from an originally planned combination of lines to a new one, which is supposed to be optimal under the conditions of the IP time interval.
- IP-UC1 aims to verify the consistency of a transport service intention, based upon the functional specifications defined in the service level agreement (SLA) between the infrastructure IOC and the TOC. This SLA is assumed to address aspects of customer quality (like travel or transport time) as well as aspects of operational quality (like robustness against disruptions)
- IP-UC2 aims to find the best way to satisfy all or most of the given requirements under the condition of a given infrastructure availability. In the special case of IP, this infrastructure availability is very often reduced during the maintenance or construction time intervals.
- IP-UC3 and IP-UC4 aim to assess customer convenience (IP-UC3) and operational stability (IP-UC4) of the IP timetable in order to improve the timetable quality while iterating with IP-UC2.

Within the scope of this project, we will describe IP-UC1 to IP-UC4, make a proposal for the involved methods for algorithmic computer support and demonstrate the concept with two test cases (see chapter 4).

1.3.2 Prototype for computer-aided timetable generation

At Swiss Federal Railways (SBB), like in other European railway companies, there exist several timetable planning systems, each of which is used by different planning departments for specific purposes and for different time horizons of timetable planning. At SBB there are two timetable planning systems in use: (a) 'Viriato', which is used the by the IOC SBB-I for strategical and conceptual planning as well as by the TOC SBB-P for developing the service offer and (b) 'NeTS', which is primarily used by SBB-I for short term planning and preparing daily operations. While the Viriato system is based on a more abstract network topology (see Figure 5b, mesoscopic), NeTS has an interface to the micro-topological (Appendix B, Figure 27, microscopic) infrastructure database 'UNO'.

Before its implementation in NeTS, the quality of the manually and iteratively constructed timetable has to be assessed in terms of timetable realizability, stability and robustness. However, the timetable planning systems mentioned are not based on an operational model, which can associate individual trains to specific infrastructure elements in space and time. Especially timetable realizability can only be assessed if assumptions about train itineraries, safety and other operational restrictions between train runs are considered. On the other hand, most of these planning systems have flexible functional and data architectures, which enable users to work with those systems in different use cases. Although in both systems, timetable data can only manually be entered and visualised, in principle, it is possible to adapt the corresponding data models in such a way that they can be used for automated timetable creation.

As a consequence, in this project, we want to combine the data model for the timetable generation algorithm with explicit information about the timetable realizability. The major outcome will be a prototype for a timetable development application, especially well-suited for the rapid, computer-aided generation of comprehensive and reliable timetable scenarios, in order to support the planning process IP as mentioned. This planning application should be based on a standard timetabling system and should adhere to the concept of computer-aided interval planning.

Algorithmic models that can generate periodic timetables (see Wüst et al., 2013) should be based on the Service Intention (SI), which represents the set of transport services offered from one station to another at a certain time slot with a certain travel time needed. The SI serves as functional timetable requirement and is structured very similarly to the customer relevant input data of existing manual timetabling systems. These, on the other hand, provide several graphical output and reporting functions, which enable the planner to assess the automatically generated results and to modify the input data in a fast and easy way based on macroscopic timetable evaluation in order to generate alternative timetable scenarios with improved commercial and/or operational performance (see Figure 1.

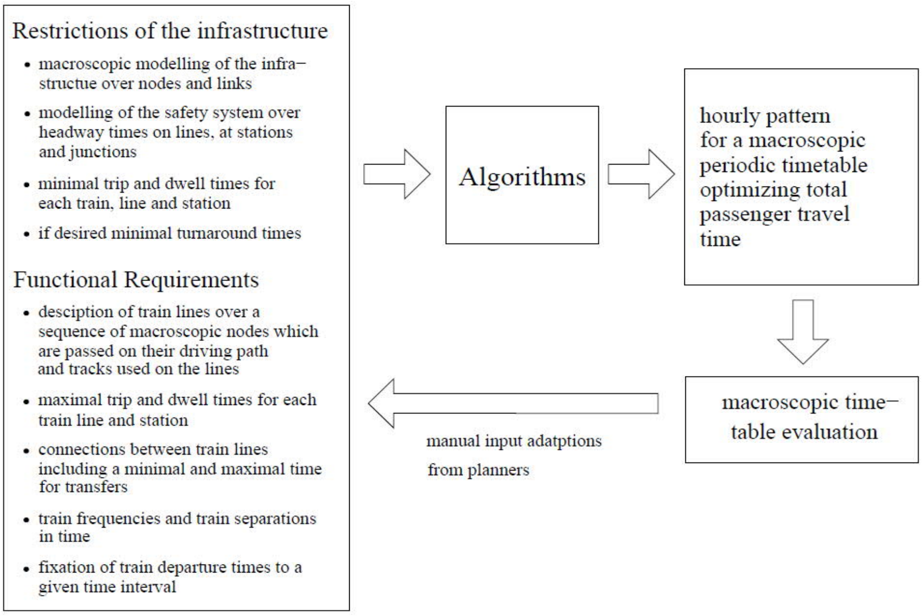


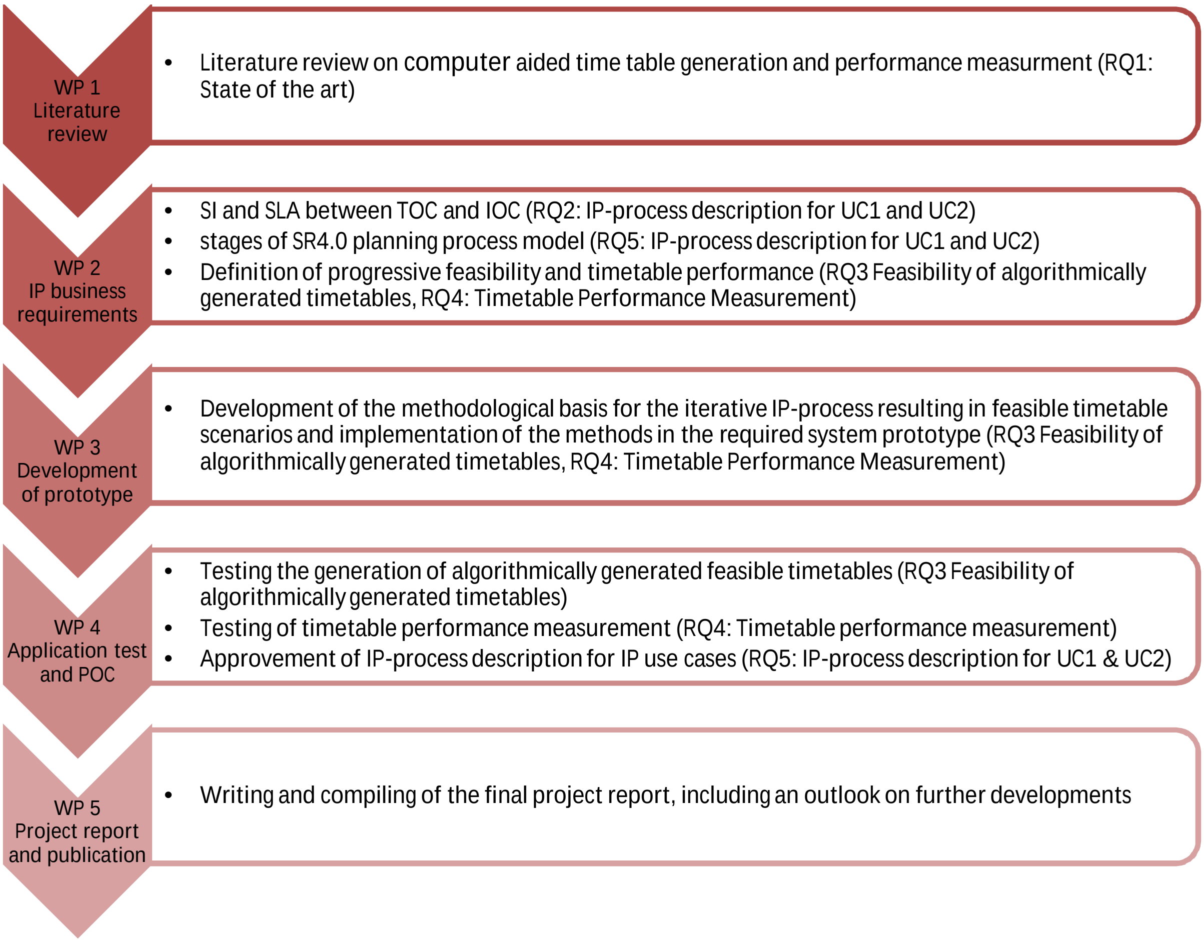
Figure 1: Overview of iterative timetabling process based on algorithmic decision support (from Herrigel (2015). The use of the SI as a functional specification for timetabling is in line with the future SBB planning and operation strategy, as described in Toletti and Weidmann (2016) and Sinner (2016). In this project, we want to demonstrate that if the macroscopic modelling of the infrastructure (see first bullet point of ‘Restrictions of the infrastructure in the left box) is extended by mesoscopic infrastructure information (see Figure 5b), the reliability of the resulting macroscopic periodic timetable can be improved.

1.3.3 Proof of concept based on a case study

The methodology of applying algorithmic decision support in real-world scheduling problems has been described in detail by Herrigel (2015). These timetabling algorithms use technical (e.g. run times between timetable points), operational (e.g. dwell times and headways) and commercial (e.g. passenger transfers) constraints between train runs. These constraints are input parameters and generate solutions for departure and arrival times of trains at stations, provided that these exist at all for the set of constraints given. However, in the IP process the aim is to find a certain timetable scenario which (a) still guarantees most of the existing train services including their published departure times (within a certain level of tolerance), and (b) takes into account reduced resource availability (e.g. track inventory) due to maintenance or construction work. In order to fulfil these requirements, the proposed algorithmic decision support has to be adapted, and specific performance measures have to be considered.

1.4 Research plan and work packages

An important aspect of the research plan for the proposed project is the intended close collaboration with SBB Infrastructure and more specifically with the expert team of the SBB project SmartRail 4.0 TMS-PAS (SR40), who also provided most of the project requirements and research questions. In order to answer the research questions mentioned and achieve the identified major outcomes, we structured the project into the following five work packages (WP 1 to WP 5):



- Literature review on computer aided time table generation and performance measurment (RQ1: State of the art)

- SI and SLA between TOC and IOC (RQ2: IP-process description for UC1 and UC2)
- stages of SR4.0 planning process model (RQ5: IP-process description for UC1 and UC2)
- Definition of progressive feasibility and timetable performance (RQ3 Feasibility of algorithmically generated timetables, RQ4: Timetable Performance Measurement)

- Development of the methodological basis for the iterative IP-process resulting in feasible timetable scenarios and implementation of the methods in the required system prototype (RQ3 Feasibility of algorithmically generated timetables, RQ4: Timetable Performance Measurement)

- Testing the generation of algorithmically generated feasible timetables (RQ3 Feasibility of algorithmically generated timetables)
- Testing of timetable performance measurement (RQ4: Timetable performance measurement)
- Approvement of IP-process description for IP use cases (RQ5: IP-process description for UC1 & UC2)

- Writing and compiling of the final project report, including an outlook on further developments

Figure 2: The project consists of five work packages (WP 1 to WP 5). The figure shows which research Questions (RQ) have been addressed by the work packages of the project proposal (ZHAW-IDP, 2017).

As mentioned, there was a close cooperation between the project teams of ZHAW and SBB during the entire project. Especially in work packages 2, 3 and 4, the collaboration was very intense. SBB project members define the inputs to the use cases IP-UC1 to IP-UC4 as well as the key performance indicators.

The transformation of evidence from scientific research into practical application is key to the successful co-operation between the institutions of ZHAW and SBB beyond the runtime of this project.

1.5 Overview of the structure of this report

This report is structured as follows: In Chapter 2, we describe the methodology for achieving the research goals described in section 1.3. This includes a short overview of the organisation of the Swiss public transport, the current approaches in computer-aided timetable generation and timetable performance measurement in section 2.1. We conclude this section with the design principles of our proposed methodology that is based on the 'service intention'. The service intention is the functional specification of the timetable, which persists during the successive process steps of railway timetable planning. This persistence over several iterative planning stages addresses the requirement of planning practitioners for an instrument for 'progressive feasibility assessment' of the timetable.

In section 2.2, we describe the special business requirements of interval planning. These are based on the Smart Rail 4.0 process model for timetable generation. This process model relies on a close cooperation between the different TOC and the IOC, including (to a certain degree) barrier-free data access for the infrastructure operating company. This improved data transparency is needed in order to ensure a defined service level, also in case of interval planning or operational disruptions. This is described in section 2.3. In Sections 2.4 (Development of Prototype), 2.5 (Network segmentation) and 2.6 (Computer-aided timetable generation based on the standard planning tool Viriato), we describe technical details of our prototype and test environment.

In chapter 3, we propose a set of use cases and an integrated application concept for computer-aided interval planning. These use cases refer to the Smart Rail 4.0 process steps. These use cases are integrated in the sense that they account for the respective input/output information of the successive process steps and thus enable the required assessment of 'progressive feasibility'.

In Chapter 4, we provide a detailed explanation of the application test based on the proposed use cases and our prototype implementation. In section 4.1, we first explain in a small test scenario how an iterative improvement of the traffic plan can be achieved using a performance measurement framework that perfectly fits our model for computer-aided timetable generation. In section 4.2, we present a case study for testing the methods introduced in chapter 2 and their interdependencies. We consider the encouraging results of the case study 'Kerenzerberg' as a proof of concept for our proposed framework.

Finally, in chapter 5, we conclude this report with a discussion of the main findings and an outlook on the expected results of the Smart Rail 4.0 project supplement. The supplement aims to develop a concept for the generation of the service intention, which is considered to be given as input for our elaborated use cases. In section 5.2, we provide an outlook on further research that we believe will be needed to successfully continue the strategic project Smart Rail 4.0 that aims at transferring the public transport sector into a new technological era, which has a tremendous impact on almost all business processes of the public transport value chain. In section 5.3, the publications resulting from this project are listed.

2 Methodology

In the operational management of railway networks, an important requirement is the fast adaptation of timetable scenarios, in which operational disruptions or time windows with temporarily unavailable infrastructure, e.g. during maintenance time windows, are taken into account. In those situations, easy and fast re-configuration of relevant input data for timetabling is of central importance. This local and temporal rescheduling results in shifted departure and arrival times and sometimes even in modified stop patterns at intermediate stations of train runs. This continuously updated information is required for operations control but also for customer information, as departure times, platforms, as well as important train connections, may change.

On the other hand, feasibility assumptions regarding these temporary changes of timetabling results have to be reliable in order to be used in operations. For obtaining reliable timetable data in terms of operational feasibility, it is prerequisite that train-track assignments, as well as operational and commercial dependencies, have been taken into consideration. Additionally, all these dependencies are supposed not to be conflicting with each other. Hence, finding the right level of detail for modelling track infrastructure and train dynamics is crucial for optimally supporting the planning process.

2.1 Survey of approaches to computer-aided timetable generation

One of the advantages of public transport in Switzerland is its outstanding usability compared to public transport in other European countries and, because of its high reliability in terms of travel time, even compared to individual transport modes. The extraordinary usability is confirmed by a high degree of public transport in the modal split (BFS, 2018). Two factors have a significant impact on the usability. The first one is the timetable regularity or periodicity, which allows travellers for easily remembering departure and arrival times and hence making travel planning much simpler, especially for regular travellers. The second one is the integrated chain of transport. No other country offers connectivity of different carriers in such a consequent manner as it can be experienced in Switzerland. Technically, this is made possible by a country-wide integrated fixed interval timetable (IFIT; see, for example, BAV (2011) and Herrigel (2015) for an explanation of the fundamental idea), which synchronises the service schedules of almost all carriers.

2.1.1 The Swiss public transport service

Whereas on the one hand, the high usability of public transport services originates in the IFIT, we observe some operational burdens of the resulting regularity on the other hand. Due to the high proportion of commuters in the number of public transport users, there is a huge difference in peak hour demand and off-peak hour demand. Because of its comparatively low flexibility, this makes it difficult to operate regular timetables all over the day. As a consequence, research groups started to develop methods in order to combine the strength of the memorability of regular timetables with the flexibility of non-regular timetables, see Caimi et al. (2011a, 2011b) and Robenek et al. (2016). Meanwhile, there are several methods known from the literature, that describe computer-aided procedures for calculating either periodic or aperiodic timetables or a combination of both. An overview is provided, for example, in Caimi et al. (2017).

2.1.2 Algorithmically generated timetables

In recent years, the requirement of finding the right level of detail for modelling track infrastructure and train dynamics in the planning process motivated several research groups to combine common timetabling procedures with constraints resulting from mesoscopic infrastructure information. Hansen and Pachl (2008) show (at critical route nodes and platform tracks), how running, dwell and headway times must be taken into account for train processing and present a deep timetable quality analysis depending on these parameters. De Fabris et al. (2014) calculate arrival and departure time, platform and the route in stations and junctions that the trains visit along their lines. Their solution is based on a multi-commodity flow approach in which they apply penalties for each train not inserted in the final timetable and for each train whose schedule differs from the desired one. Bešinović et al. (2016) present a micro-macro framework based on an integrated iterative approach for computing a microscopically conflict-free timetable that uses a macroscopic optimisation model with a post-processing robustness evaluation. Caimi et al. (2011b) extend PESP by proposing the flexible periodic event scheduling problem (FPESP), where intervals are generated instead of fixed event times. By applying FPESP, the output does not define a final timetable but an input for finding a feasible timetable on a microscopic level, see for example Caimi (2009) and Caimi et al. (2011b). For this reason, we consider the FPESP model to be the right choice for developing timetables with both desired purposes, flexibility in case of operational differences during the day and a medium level of available detail for early planning stages.

2.1.3 Timetable stability criteria

Software tools are also used to assess timetable stability. To these belong the systems, e.g. RailSys (Siefer and Radtke, 2005), OpenTrack (Hürlimann, 2001) and LUKS (Janecek et al., 2010). Common to all systems is that they use deterministic variables to model errors by scattering in individual primary delays. A large number of simulation runs are performed in order to aggregate the resulting deterministic result ("Monte Carlo simulation") for the individual process times. In this way, they receive statistically verified statements about the behaviour of different timetable variants (Büker and Seybold, 2012).

Similar approaches have been introduced with macroscopic data to extend the model domain in the systems Fasta (Noordeen, 1995) and SIMONE (Middelkoop and Bouwman, 2001)). In addition, the consideration at this aggregated level of detail has the advantage that it can also be used in earlier planning phases to evaluate strategic decisions. In order to accelerate the calculation times of the simulation, Büker directly uses case variables for delays and introduces suitable distribution functions in his dissertation (Büker, 2010). This idea makes it possible to evaluate timetables for large and complex networks.

Together with further elaborations (Büker and Seybold, 2012), it forms the core of a new evaluation software called OnTime (Franke et al., 2013), which has been available since summer 2011 and is used by SBB infrastructure for timetable planning. The data required for an OnTime evaluation is very flexible and therefore, suitable for macroscopic timetables in an early planning phase. As input parameters, at least one travel route based on a sequence of operating points with arrival and departure or transit time is necessary for each train line. Moreover, parameters like the duration and length of stay must be known, as well as primary delays for each train activity (entry, stop, take-off, departure) and their probability of occurrence. Having detailed knowledge of the track layout on open lines and stations increases the reliability of the stability forecast. It is possible to add dependencies between train times such as turnaround times and infrastructure information to specify route exclusions, minimum train times and train priorities.

From the input information, a so-called activity graph is defined to formalise the occurrence of primary delays and their propagation. Each node of this diagram represents a train activity for each train line and station as one of four types of activity: Entry signal reached (A*), arrived (A), ready to start (D*), exited (D). Connections between these nodes are defined for model trips, stops, connections between lines and route conflicts. Cumulative distribution functions, which are described in detail in (Büker, 2010), are used to model primary delays.

Conditional and unconditional convolutions of the distributions are finally used to propagate delays along the train activity graph. During this process, further inputs influence this delay propagation: travel and dwell times, for example, complement delays in driving and stopping activities, buffer times reduce secondary delays between different trains. Moreover, for the configuration of connection and conflict parameters, one has to take maximum waiting times and minimum travel times into account.

The Dutch timetable evaluation system PETER (Performance Evaluation of Timed Events in Railways) uses a special algebraic approach (see sections 2.1.3 and 2.4.4) based on Max-Plus Algebra. This approach has been elaborated in mathematical detail in Goverde (2007). The benefits of this algebraic approach for timetable stability analysis have been illustrated in Tabak (2008).

2.1.4 'Service Intention' based approach for timetable specification

A central element of the methodology investigated is the definition of the transport service. The intended transport service offer (Service intention: SI) can be derived from the demand volume for a given source-destination relationship at a particular annual, weekly or daily time (see also first process step 'demand assignment' in Figure 4). The service intention is a major component of the service level to which the infrastructure manager tries to adhere as much as possible in all situations.

While public transport timetables are known from everyday life with conventional timings, precisely planned to the minute and second, the SI is a suitable way of specifying the offer for scheduling. When creating the integrated fixed-interval timetable (IFIT) on the basis of SI's, usually system times are used (minimum travel times between node stations, see for example Herrigel (2015) and BAV (2011) for an overview of the system times for different expansion stages of the Swiss rail network). Here, the reference hour is divided into eighths. This results in time intervals with a time span of 7.5 minutes. While planning daily activities, travellers can easily memorise start and end of their journeys based on the respective eighth of an hour.

The SI is a time format with which train runs can be easily processed, stored and displayed in vector form. For this reason, our modelling approach is based on the service intention (SI). Technically spoken, the SI represents a data structure, which integrates commercial timetabling requirements given by the respective line concept on one side and technical constraints on the other. Technical constraints include operational dependencies on resource properties of track and rolling stock as well as route and safety conditions. The level of detail of the SI corresponds to the data types, which are typically used as input in the timetable development process. Functional timetabling requirements determine the length of dwell times at stations and stops, line frequencies and separations as well as transfers between lines at specific stations.

The SI data structure was first described in Caimi (2009) and Caimi et al. (2011a). Practitioners also call the functional part of this information 'line concept'. The SI results from a strategical planning step that takes into account assumptions regarding cumulative numbers of available resources such as the number of tracks per section and the dynamics and the circulation of rolling stock. Similar to, e.g. de Fabris et al. (2014), we call this level of abstraction of the available resources 'mesoscopic topology'. Together with the functional requirements of the SI, this mesoscopic infrastructure data model of a given scenario is entered into a standard timetable editor (see for instance section 2.4.1 and the description of Viriato in SMA (2018)). The SI is the functional specification and therefore, the input for the process of generating a timetable in detail. It is still flexible enough to allow different ways of operational planning and resource allocation.

2.2 IP business requirements

At the beginning of the SR40 project, the SBB project team has been executing a business analysis involving a review of all relevant business processes and a description of the future timetable generation process. The desired future business model is documented in Howald et al. (2017) and summarised in Figure 3. In Howald et al. (2017), the functional timetable specification is described as follows:

“The "Timetable" - hereafter referred to as the "Traffic Plan" - describes the requirements for the timetable (regardless of who the demand is for and at what time horizon) as a functional offer description. In the past, however, the timetable was already defined years in advance, and the train runs were planned exactly onto the second. In the future form, customers' wishes can be optimally mapped, and the effects of adjustments to a requirement are immediately recognisable. The infrastructure manager can thus obtain the necessary flexibility to adapt the operational capacity plan in the medium and short term in such a way that the agreed economic aspects can be met without the disadvantage of disrupting the customers' transport chain.

Figuratively speaking, these functional offer descriptions create capacity bands within which a train can move. The transport needs can theoretically be fully timed, which means that they are valid as long as the need exists. They conclude a commercial agreement between the bidder and the IOC. They thus also form the basis for informing customers in passenger and freight traffic.

Within the capacity ranges and resulting from the functional offer description, the individual capacity objects (trips, shutdowns and intervals) are displayed. These can be planned out daily and adapted to the respective situation (e.g. construction sites, weather). This creates a capacity plan that can be used as a production plan for the execution without manual post-processing. Adaptations in the traffic plan usually have a direct impact and trigger a new assessment in the capacity plan.

The final step implements an integrated, realistic, conflict-free production planning. The planning is geared to highly automated production. In this technical and operational consideration, capacity supply and train path are strictly separated from each other. In coordination with all co-production partners, production planning is done for day-specific train paths, routes, connections including waiting periods, personnel, rotations for rolling stock, manoeuvring procedures and services, as well as additional services (such as parking and shunting). The feasibility check is carried out consistently across all levels of the topology.

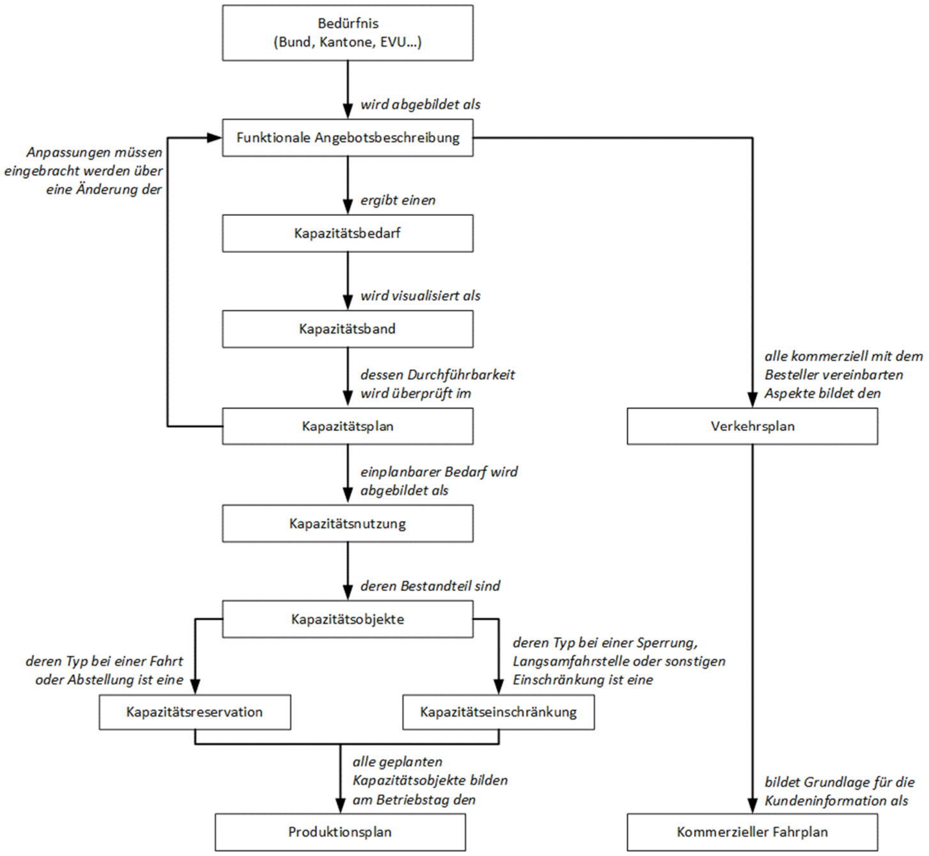


Figure 3: Overview of the timetable planning process from SBB document "Ergebnisdokument. Zielarchitektur «PPS2030» (Business & IT)". For details, see Projektteam «PPS2030» (2017, page 42).

In the case of IP or operational disruptions, the planner has to account for a planned (in case of IP) or instantaneous (in case of disruptions) reduction of the normal production capacity. This capacity reduction requires an adaption of the production plan (e.g. using different rolling stock or track routes or switches) which in some cases does not - or almost not - have an impact on the SI.

Building upon the business requirements expressed in the document Projektteam «PPS2030» (2017), the SR40 project team developed a timetabling process model that we used as input for designing the TOC to IOC interface and the IP-use cases.

2.2.1 SR40 process model for timetable generation

The input to the business model for timetable generation is represented by the functional description of the desired transport service. Figure 4 shows how this model for timetable generation can be split into an input part describing commercial requirements and an output part covering the operational view.

This desired transport service represents requests originating from the various people in power in public transport is consolidated by the TOCs like SBB-Passenger transport (SBB-P) or SBB-Freight transport (SBB-C). This functional description of the desired transport service is called Service Intention (SI). The functional requirements represent scenarios of transport chains, which have been consolidated before. In an initial step, the SI is translated into a capacity requirement, mapped onto railway lines and stations. It can be visualised in terms of capacity and in time-space diagrams. In a second step, the capacity requirement of the different train lines has to be verified for operational feasibility. This process step is called traffic or capacity planning, resulting in a validated version of the service intention, which accounts for capacity constraints defined by track, occupation, headway, transfer and line rotation time requirements. In addition, constraints resulting from maintenance and construction requirements are accounted for. All these aspects of capacity consumption are integrated into the capacity plan.

Our method attempts to operationalise these two steps in terms of use cases, prototypes for algorithmic data processing and timetable performance measurement.

As one can see on the right side of Figure 4, even in cases of reduced capacity (compared to resource conditions of the standard timetable), the IOC has the responsibility of providing the best service quality possible. That means that in case of interval planning or operational disruption, the IOC has to have access to demand and service specific data (managed by the TOCs) which determine the input for the process of generating a consolidated SI.

Figure 4 also indicates that the planning process is iterative (indicated by the grey backward loop arrows). This implies that the quality and the level of detail of the resulting plan is progressive. Therefore, the following question arises: Is "progressive planning" possible, and can the feasibility of the resulting plan be detected already at an early stage in the planning process?

Why progressive feasibility? A fundamental motivation for progressive feasibility is the fact that planning fundamentals and parameters, if known, often change over time. Therefore, in an early step of the process second-exact planning seems questionable. Progressive planning allows

- to avoid frequent rescheduling (→ planning instability)
- to deal better with planning uncertainties (location of new points, velocity-thresholds, revolutions, etc.) (→ robustness)
- to consciously include blurring (of routes and nodes, vehicle deployment, holding policy, etc.) in order to maintain flexibility for later planning stages, rather than second-by-second planning in the knowledge that the framework conditions are still changing frequently.
- To reduce overhead for topology maintenance.

For this reason, we made the following definitions for our study:

"Feasibility": Feasibility refers to a given traffic or capacity plan. A given transport or capacity plan is feasible if it can be implemented in production, as long as all parameters can be realised in the entered value ranges.

"Progressive feasibility": Progressive proof of feasibility means that the feasibility is ensured at a level of detail corresponding to the planning level. A change to a finer level of detail must always be possible. Figure 4 provides an overview of the process of production planning and timetable development. As can be seen from the grey arcs pointing backwards in the process chain, the development process is strongly iterative, meaning that it must always be possible to review a former input at a later process stage. Figure 4 also shows that the railway value chain is divided into two major process steps that are commonly known as 'line planning' and 'capacity planning'. Whereas the responsibility for the first process step is mainly with the TOC, because he is the owner of the transport business case, the responsibility for the second process step is with the IOC, who has the task to integrate the services of – in general – several different TOC.

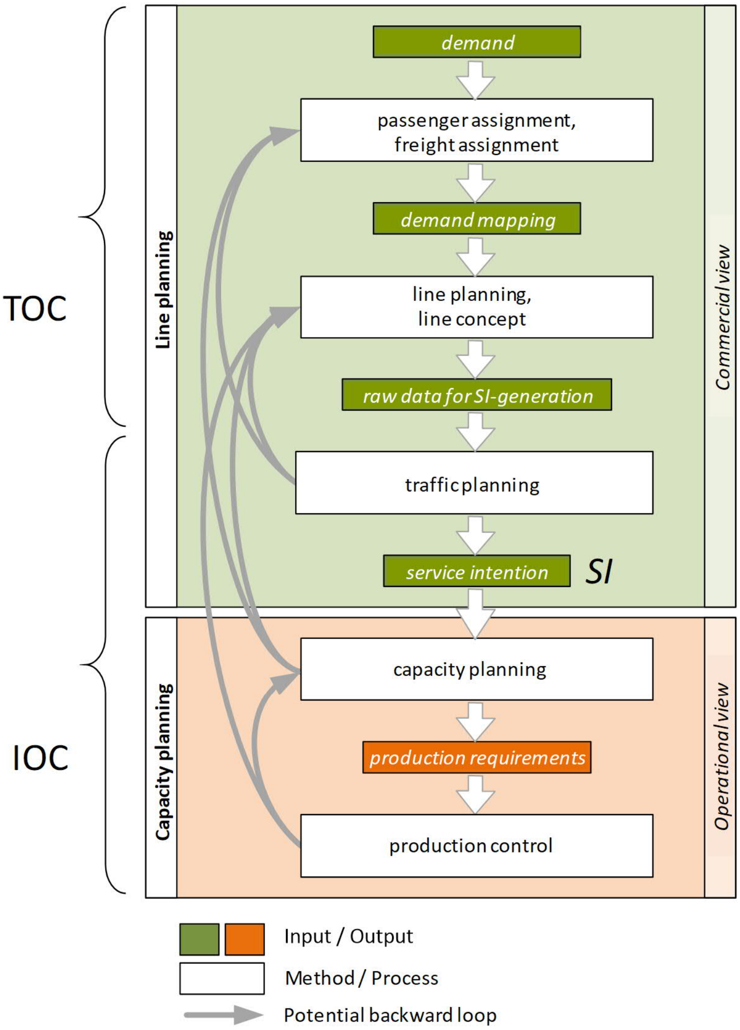


Figure 4: SBB SmartRail 4.0 timetable planning process: overview and integration of line planning into timetable planning model. Grey shaded area indicates the shared process and data access by TOC and IOC. These shared processes and shared data access should be ruled out in a service agreement that aims for a cooperative planning process. If there are several TOCs involved in the timetable scenario, there are good reasons for assigning the responsibility for the SI to the integrating IOC. If there is only one TOC involved, it might be more reasonable to leave the SI-responsibility with the TOC planner. Adaptation from Howald et al. (2017, figure 3 on page 8).

2.2.2 TOC to IOC-interface

The SI represents the functional timetable requirements and hence, according to Figure 4, the data interface between the different TOC and the IOC. As the operational timetable feasibility can be verified only after the integration of the service requests of the contributing TOC's this act of consolidation is an iterative process and the details of this process interface should be agreed on by the TOC and the IOC based on a formal document such as, e.g. a Service level agreement (SLA). In the case of operational irregularities, maintenance requirements and disruptions, it often happens that the originally proposed service appears to be not feasible in production. Therefore, the SLA should be designed in such a way, that it allows the IOC to make a proposal, which meets the SI as close as possible. Therefore, the IOC has to have access to the respective origin-to-destination (OD)-demand and cost generating data.

The SI defines a transport service for passengers or freight between an origin and a destination station including a weighting factor in terms of, e.g. the number of passengers or a value weight factor of freight transport and proposal for the service production in terms of involved train lines, line transfers and frequencies.

We propose the following six data interfaces (in Table 1 to Table 6) for the process steps of Figure 4. Except for the microscopic data interface, which is most likely an internal interface of the IOC, these interfaces have to be managed by both organisational units (TOC and IOC). Performance goals and detailed conditions are to be handled by appropriate service level agreements.

<table>
<tr><th colspan="2">Interface description “Demand assignment: Strategic dimensioning of transport performance”</th></tr>
<tr><td>Input
• Demand (origin-destination matrix, possibly time-stratified).
• Infrastructure as a network with nodes (access points) and edges (routes) with
o travel times per edge
o generic transfer times per node.
o possibly additional service requests per relation
o maximum travel times per relation</td><td>Output
• transport service concept
• transport performance to be provided per edge and time</td></tr>
<tr><td colspan="2">Remarks
• The transport service concept has to be understood as a kind of volume of transport services linking origins and destinations along the existing rail network section for a certain planning horizon. Without a corresponding line concept, it cannot be used for further planning steps.</td></tr>
<tr><td colspan="2">Literature
• Desaulniers and Hickman (2007)</td></tr>
</table>

Table 1: Interface description “Demand assignment: Strategic dimensioning of transport performance”.

<table>
<tr><th colspan="2">Interface description “Line planning”</th></tr>
<tr><td>Input
• transport service concept
o line pool (potential lines to operate on itineraries in the rail network)
o line description for each line in line pool (line category), served stations, implemented vehicle type
o implementation cost for each vehicle type per time (or distance) unit
o transport performance to be provided per edge and time
• Infrastructure as a network with nodes (access points) and edges (routes) with
o travel times per edge
o maximum frequency per edge
o generic transfer times per node</td><td>Output
• line concept
o line pool (to be implemented, in general, a subset of the input line pool)
o line plan with frequencies (periodicity, tact) and several required vehicles of each line</td></tr>
<tr><td colspan="2">Remarks
Line planning models usually work with a pool of possible lines (and their cost), from which the model selects, in order to provide desired transport performance. The goal is usually to minimise total travel time, travel time over all or the cost. Transfer times can only be estimated here.
The SI is almost a line concept plus other explicit requirements (desired time specifications for separation and transfer times between lines.
There is quite some evidence from literature, that the two process steps: “Demand assignment” and “line planning” have to be integrated into one process. This is the reason why we propose an integrated use case “line planning with demand assignment” (IP-UC0) in chapter 3.</td></tr>
<tr><td colspan="2">Literature
• Schöbel (2012)</td></tr>
</table>

Table 2: Interface description “Line planning”.

<table>
<tr><th colspan="2">Interface description “Traffic planning”</th></tr>
<tr><td>Input
• transport service concept
• simplified macroscopic infrastructure model
o Travel times on the route sections
o no headways</td><td>Output
• traffic plan (assignment of lines to time slots)</td></tr>
</table>

Remarks • Traffic planning is rarely reported as a separate planning step. • The modelling without headways means that no explicit train sequences can be set on the route segments. It can be shown that timetable feasibility in terms of infrastructure capacity and operational conditions cannot be detected at this level of detail. However, the model can be used to detect inconsistencies (sets of contradictory requirements) in the SI. • The model can take into account specifications for the (minimum and maximum) size of the time bands and can thus serve for the targeted distribution of flexibility and reserves.
Literature • Liebchen and Möhring (2007)

Table 3: Interface description “Traffic planning”.

<table>
<tr><th colspan="2">Interface description “Macroscopic capacity planning”</th></tr>
<tr><td>Input
• Service intention including time alignments (traffic plan)
• Macroscopic infrastructure model (including headways)</td><td>Output
• Macroscopic capacity plan (operational arrival and departure times)
• Transit times at all macroscopic operating points</td></tr>
<tr><td colspan="2">Remarks
In the macroscopic timetable generation, train sequences are determined based on the macroscopic infrastructure model. The methodology is based on the PESP model and is well understood and tested. In order to introduce flexibility time intervals, instead of time points can be planned as output (capacity bands instead of timelines).
Other topics such as partial periodicity are complex and so far, hardly described in the literature, but unavoidable from a practical point of view.</td></tr>
<tr><td colspan="2">Literature
• Caimi et al. (2017)</td></tr>
</table>

Table 4: Interface description “Macroscopic capacity planning”.

<table>
<tr><th colspan="2">Interface description “Mesoscopic capacity planning”</th></tr>
<tr><td>Input
• Traffic plan (specified by the SI, but taking capacity restrictions into account only at a high level of abstraction, e.g. number of IC-equivalent train slots per track, hour and direction)
• Line plan with frequencies (periodicity, tact) and several required vehicles of each line
• Mesoscopic infrastructure model
• Service level (balance of cost and customer convenience)</td><td>Output
• Mesoscopic capacity plan
• Itinerary on the mesoscopic topology</td></tr>
<tr><td colspan="2">Remarks
The mesoscopic timetable creation is the least well-investigated so far.
There are a few models in the literature, all of which operate with very different assumptions and infrastructure models, as no universally accepted mesoscopic infrastructure model has been established.</td></tr>
<tr><td colspan="2">Literature
• The description in de Fabris et al. (2014) is an example and served as an entry point for our research.</td></tr>
</table>

Table 5: Interface description “Mesoscopic capacity planning”.

<table>
<tr><th colspan="2">Interface description “Microscopic capacity planning”</th></tr>
<tr><td>Input
• Mesoscopic capacity plan
• Microscopic infrastructure model</td><td>Output
• Microscopic capacity plan (production specification)</td></tr>
<tr><td colspan="2">Remarks
There are various models for the microscopic timetable, which can be roughly divided into two different classes:
• time-continuous models, which map event times (for example occupancy and release times of infrastructure elements) as rational numbers and map conflicts into separation times
• discrete-time models, which select binary variables from a discrete set of driving alternatives (both in terms of time and route selection) and map conflicts over the alternatives. Both model classes do not (yet) scale enough and, from a certain problem size, lead to very large computation times. (This is very strongly dependent on details of the respective problem instance and can vary greatly). This class is currently more suited for the planning of small network sections (corridors, node planning).</td></tr>
<tr><td colspan="2">Literature
• Lusby et al. (2011)</td></tr>
</table>

Table 6: Interface description “Microscopic capacity planning”.

2.2.3 Process steps for timetable planning

Based on the business requirements described in the preceding chapters and evidence from our literature survey, we propose a business model with the following process steps, which partly integrate the functional steps shown in Figure 4.

Process step 1: "line planning with demand assignment". For the development of the transport service, which is the main product of the TOC and hence the core of his business model, there exist evidence from the literature to integrate the two process steps "demand assignment" and "line planning" into one use case. The main reason for this is that the demand has to be assigned to services, which are implemented by train lines. The macroscopic infrastructure defines the relevant network. In literature (see, e.g. Friedrich et al. 2017), an adequate data model is provided by the Public Transport Network (PTN). Customer demand (mostly commodity and time specific) is given from some forecasted origin-destination demand matrices. Corresponding services, provided by train lines include attributes like frequencies, routings and connectivity. The output of this process step can be used for the further development of the timetable but also for infrastructure capacity and utilisation planning. The elaboration of this process step is not subject of the current project but is requested by the SR40 project team in a supplement to the current project, with a different timeline. The documentation of the line planning process step will be included in the project report of the project supplement. For this project, we assume, that feasible line concepts exist for all timetabling scenarios of a test corridor. This implies that we do not have to relax the service intention for a certain scenario by executing this process step.

Process step 2: "generation of traffic plan". Based on the result of the line planning process step and the macroscopic infrastructure, the traffic plan for a given scenario is generated by assigning lines to time slots including interdependencies of the lines like total line trip time, time separations of lines and potential transfer times between lines at given stations. All line time slots are indicated by time intervals with specified extensions. No explicit train sequences can be set on the route segments. Timetable feasibility in terms of infrastructure capacity cannot be detected at this level of detail. However, the model can be used to detect inconsistencies (sets of contradictory requirements) in the SI.

Process step 3: "generation of traffic plan with capacity time band". Given the traffic plan and the mesoscopic infrastructure model, the mesoscopic capacity plan, including train itineraries on the mesoscopic topology is generated in this use case. The principles are described in an internal project report of SBB (Laumanns et al. 2017). The algorithmic generation of the resulting mesoscopic capacity plan (see 2.4.2) is the major outcome of this research and has recently been published in Wüst et al. (2018a and 2018b). The final traffic plan with capacity time band is the input to the elaboration of the traffic plan with micro-topology that is used for the implementation of the production plan. For this reason, its quality has to be iteratively assessed by additional process steps (see also section 2.3) and has to be improved appropriately.

Process step 4: "assessment of the stability of traffic plan". For the assessment of the stability of traffic plan we make use of an excellently well-described framework for railway timetable stability assessment (see section 2.1.3 and, e.g. Goverde, 2007). This framework makes use of a data model and functions that align perfectly with our proposed timetable generation algorithm (2.4.2). Several Performance indicators can directly be used to adjust timetabling parameters such as event flexibility for critical lines at identified stations in order to improve the robustness of the capacitated traffic plan against disruptions. More detailed explanations can be found in section 2.4.4.

Process step 5: "service quality assessment of a traffic plan". For the assessment of the service quality of the traffic plan with capacity time bands, we propose to determine the total travel time for a given timetabling scenario. The given timetabling scenario is therefore specified by a limited geographical perimeter and a limited time horizon. This is in line with the objective functions for the line planning and timetabling algorithms. For more details, see sections 2.4.2 and 2.4.4.

Process step 6: “generation of a traffic plan with microscopic topology”. For the final production plan with an exact train slot assignment to track sections on the level of the safety system for the configuration of the customer information system, the microscopic topology has to be taken into consideration. The elaboration of this process step is not subject of the current project. It is elaborated as a “proof of concept” within a dedicated part project (“FLUX”) by the SR40 project team.

More detailed descriptions of process steps 2 to 5 can be found in the use case description for use cases IP-UC1 to IP-UC4 in chapter 3.

2.3 Timetable performance requirements

The basic idea of the iterative improvement of the timetable performance as illustrated in Figure 1 and Herrigel (2015) has been described earlier in a case study, which had been requested by SBB (Wüst et al., 2017).

According to this idea, the required performance, which has to be met by a timetable, is determined by the agreed service level between TOC and IOC. The TOC has the responsibility to provide a certain convenience to the transport customer in terms of transport comfort and speed. In order to achieve this goal, the TOC has a certain budget for covering occurring production costs. The IOC has the responsibility to provide infrastructure and transport capacity to all TOC’s, which contribute to the overall transport service. For this reason, there are two independent performance criteria that the IOC timetable planner must meet when compiling the consolidated timetable. One performance measure refers to customer convenience (total travel time), the other to operational stability and cost efficiency. A planning scenario is defined by the (passenger and freight) transport requirements of the planning horizon and the means of production required to cover the demand. Both measures are explained in detail in the following sections.

2.3.1 Customers’ perspective

The timetable has a significant impact on customer convenience as it determines if intended train services can be executed reliably such that start-to-end transport services are as fast and as direct as promised in the published timetable. In case that a reduction of resources enforces a new planning of the service to be offered (e.g. in the interval plan), this new (probably only temporary) plan should enable transport services that are as close to the originally promised services. This does not necessarily mean that the timetable itself should be as similar to the original one as possible.

In order to reach this goal, the timetable planner has to have access to demand data and transport requirements that had led to the original timetable. If now the performance (in terms of customer convenience) of is assessed by a quantitative measure, we propose to use the overall travel time which is the sum of all (volume weighted) origin-to-destination trip times of freight and passengers.

That means that, if the overall travel time of the original timetable is known, any alternative timetable scenario can be compared quantitatively to the original scenario by calculating the ratio of the original overall travel time to the scenario overall travel time. We call this ratio service intention index (SII).

2.3.2 Operators’ perspective

The operators’ requirement for the quality of a timetable is mainly determined by the cost, which he has to invest in achieving the agreed level of service and the available resources. The schedule has a direct impact on operating costs due to its stability. If disruptions happen or fluctuations in operational process times force lines to be changed or cancelled and connections to be broken, the service level cannot be guaranteed anymore, and efforts have to be made to reduce negative impacts for the customer. In section 2.4.4, we introduce quantitative measures for stability and capacity utilisation and a formal framework based on event activity models. This framework is called Max-Plus-Algebra and was first described in all detail by Goverde (2007). Similar to the SII, it has the advantage that the planner can compare to different timetable scenarios based on quantitative measures. The information that one can extract from the different stability indicators (e.g. critical circle times, delay sensitivity and delay impact, capacity utilisation) can directly be used to improve the parameters of the computer-aided timetabling (e.g. local and global flexibility requirements).

2.4 Development of Prototype

The system prototype that has been developed in WP3 should be well suited to demonstrate the main ideas of our proposed concept for computer-aided IP. That means it should fit well together with practical IP cases and built upon available tools and concepts as far as possible. In addition, it should be possible to answer the research questions described in section 1.2. Based on our literature survey, we identified the several modules that are required for the system prototypes. These modules are discussed in the five following subsections.

2.4.1 Mesoscopic infrastructure modelling

The generation and investigation of feasible event times for individual train runs and corresponding resource allocations fitting into the structure of an IFIT are usually done manually. For this reason, timetabling is considered a time consuming and challenging task, even for experienced planners. On the other side, algorithmic approaches for solving this task computationally require models based on microscopic information about track capacity, like for instance in Bešinović et al. (2016) or - in an intermediary step – define possible train routes like in de Fabris et al. (2014), from which headway constraints for trains can be derived. Headway constraints can also be used for solving standard periodic timetable problems. In order to facilitate this data preparation step, we present a generic approach, which makes use of the mesoscopic infrastructure, a data structure which is implemented and managed in a standard timetable planning system like Viriato (refer to, e.g. SMA, Viriato - software for railways. Info Folder 2018).

In order to illustrate the level of detail of the respective infrastructure mapped onto a mesoscopic topology as opposed to macroscopic or microscopic topology, we refer to Figure 27 in Appendix B and the corresponding explanation provided by Howald et al. (2017). From this mesoscopic topology, we transfer the information regarding the node sequence as well as the capacity of each node. In our topology graph, we map operation points and sections that link two operation points together onto nodes of a mathematical graph. For our case study, we assume that one can change tracks at any transition between nodes in all possible combinations. Creating timetables considering mesoscopic infrastructure enables a much better feasibility assessment of the result compared to considering only macroscopic infrastructure. On the other hand, the difference between microscopic infrastructure in terms of a feasibility assessment is negligible. On the one hand, implementing the mesoscopic topology together with the event flexibility according to the FPESP model, introduced in section 2.1.2 allows to generate periodic timetables with a reasonably good assessment of feasibility. On the other hand, this method generates results of sufficient flexibility to find a conflict-free resource allocation taking a micro-topological level of detail into consideration or if planning has to account for slightly different individual conditions (e.g. during the course of a day or considering operational variability).

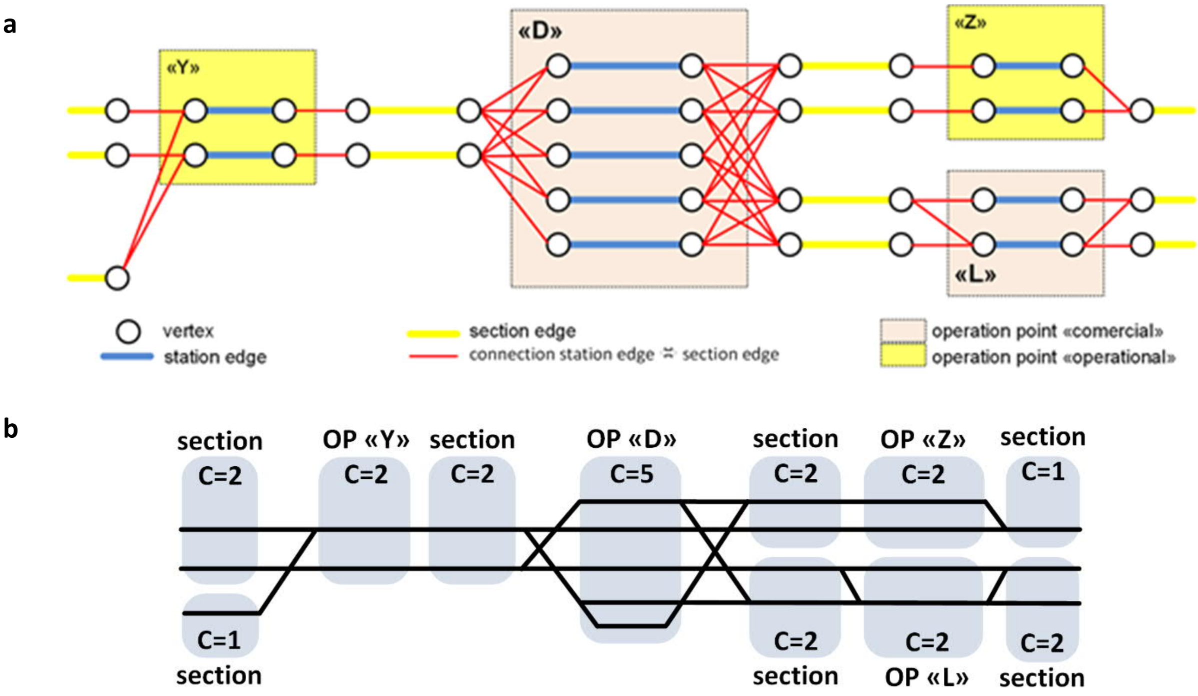


Figure 5: a) Mesoscopic infrastructure example of Figure 27 mapped into our graph representation b). Each operation point and each linking track segment is mapped into a graph node, represented by a grey shaded box. **C** indicates the track capacity of each node. Switches between node tracks allow changing tracks when moving from one node to the other.

2.4.2 Event activity network and track-choice PESP model

The modelling approach that we propose for the SR40 project combines the idea of the event flexibility of the train runs to be scheduled in the timetable (FPESP) with the idea of the mesoscopic topology in so far as to assign to each of the train runs a sequence of mesoscopic track nodes.
In the following, we will call this approach Track-Choice PESP (TCPESP) as it can be considered as an extension of PESP, the commonly known timetable planning model described for instance in Liebchen and Möhring (2007). The model includes event flexibility constraints (see Caimi et al., 2011b) and can automatically select the relevant headway constraints resulting from the mesoscopic track node assignment into the optimisation problem.

The objective of the computer-aided timetable generation is to find either an individual time stamp (PESP model) or a small temporal fixed time interval (FPESP model) for each timetable event representing either an arrival or departure of a train run at a station (customer timetable) or at an operation point (operational timetable). In our operational timetabling model, an operation point can be either a station, a junction or a section connecting two neighbouring operation points. In our case, each train run to be scheduled needs such a pair of event time intervals for each of the operation points traversed by the corresponding route of the train. If for instance as shown in Figure 7a, we have two train lines, line 1 connecting "z" via "D" to "L" and line 2 connecting "y" via "D" to "L", both ending in operation point "L", then one can derive from this functional requirement for the two lines, that they need to have a turnaround constraint between the two corresponding train runs, one ending in operation point "L" (e.g. train run 11 in Figure 6) and one departing from operation point "L" (e.g. train run 12 in Figure 6). If we additionally request that the two lines offer a transfer of passengers between the two lines in "D", then we derive from this connection requirement a so-called "event activity network" (EAN). Each of the numbered vertices in the example graph represents an arrival (or start) event of a train run at an operation point or a departure (or end) event of a train run at an operation point. The connecting arrows represent activities (run, dwell, or transition) or dependencies (headway, connection, turnaround) between two events. For more details of the mathematical model, we refer to Wüst et al. (2018). Based on the proposed track TCPESP-model, it is possible to calculate timetable events for each train run at all traversed operation points, that respect all mentioned constraints.

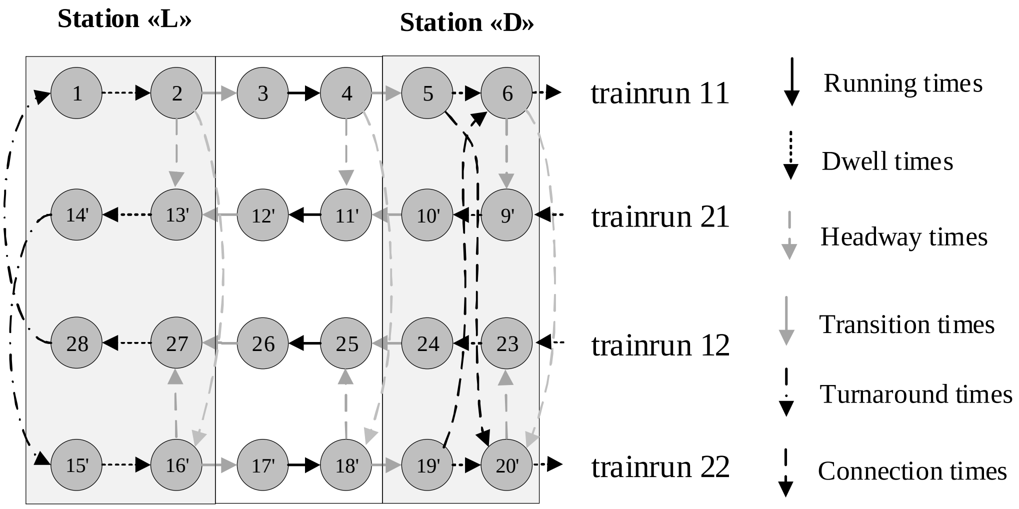
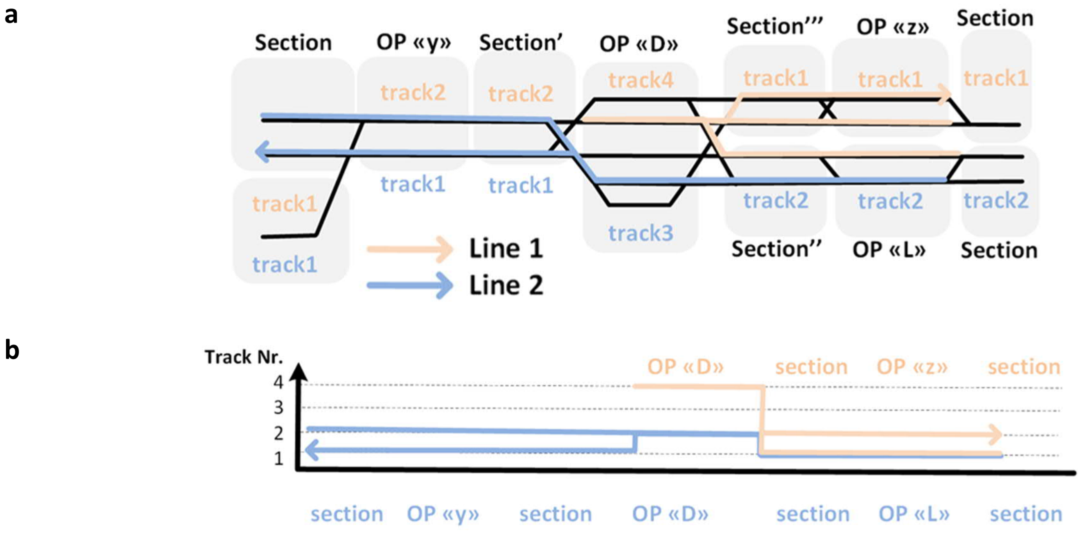


Figure 6: Sample of an event activity network (EAN). Nodes belonging to grey shaded boxes indicate events at an operation point of type ‘working points’. Other nodes indicate track type arrival and departure events. Arrows indicate different types of time dependencies.

A feasible track allocation for this situation is illustrated in Figure 7. There are several options to plot this track allocation in a diagram. One option might be, to enumerate track numbers along with a projection of geographical coordinates. This is typical for graphical timetable representations in standard timetabling tools. Figure 7b shows such an example of the situation of Figure 7a. As one can see, such a projection is not very clear regarding the track assignment. Also, a track enumeration across different operation points does not exist in general. Therefore, we propose the track assignment diagram for this purpose. For the example of Figure 7a, such a track assignment diagram is illustrated in Figure 7c. It is a generalisation of the well-known track assignment diagram for stations (in German: Gleisbelegungsplan).



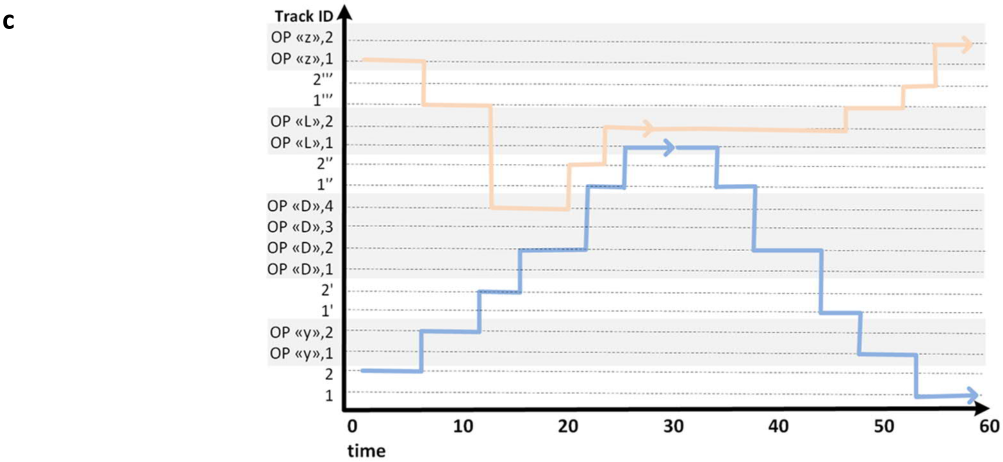


Figure 7: a) Potential track allocation for the situation of Figure 6 with line 1 connecting “z” via “D” to “L” and line 2 connecting “y” via “D” to “L”. Both lines have turnarounds in “L” and offer a transfer service on-to each other in “D”. b) shows a track assignment as a projection to geographical coordinates. More recommendable is a general track assignment diagram with track IDs plotted against time, as illustrated in c).

2.4.3 Periodic Timetabling with Event Flexibility

In order to avoid tedious iterations between the process steps “microscopic capacity planning” and “mesoscopic capacity planning” in case of the infeasibility of the micro-level problem, one can improve the chance of finding a feasible solution by enlarging the solution space in the micro-level. This approach has been described in detail in Caimi et al. (2011b). We also implement this event flexibility method by adding some flexibility for the events of the EAN by introducing lower and upper bounds to the event times of the arrival and departure nodes in Figure 6. The final choice of the event times in the range between the lower and upper bound shall be independent for each event such that each value of the end of an activity arc should be reachable from each time value at the beginning of that activity arc (see Figure 8).

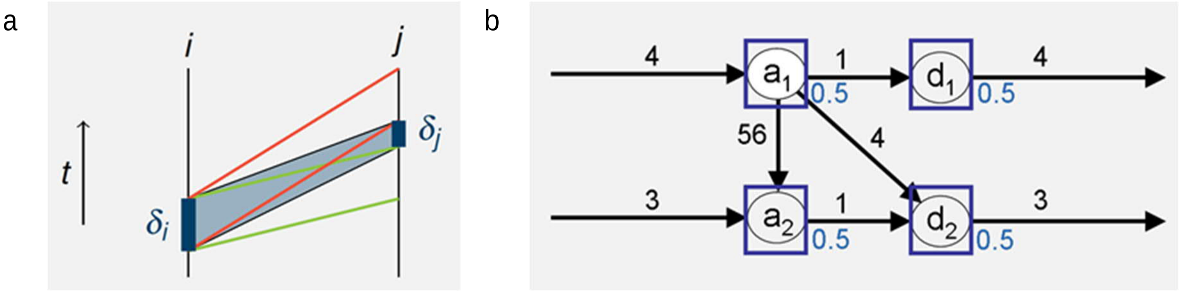


Figure 8: Target oriented placement of time reserves: a) Time frames $\left[t_i, t_i+\delta_i\right]$ in place of time points t_i. By implementing this method, the normal PESP constraints $l_{i,j} \le \left[t_j - t_i\right]_T \le u_{i,j}$ become $l_{i,j} + \delta_i \le \left[t_j - t_i\right]_T \le u_{i,j} - \delta_j$. All events may occur independently from each other within the respective time frame. In the EAN example of b) this means that instead of planning time points (a_1, d_1, a_2, d_2) we plan time frames with $\delta_1 = \delta_2 = 0.5$.

We are not forced to add this flexibility to all the events, but we can select the nodes where we want to add it, for instance only nodes corresponding to events within a main station area with high traffic density, where it is more difficult to schedule trains on the microscopic level. In general, one can say that this placement of flexibility is the timetable configuration feature that has the highest level of influence on improving operational stability. This is where the information provided by the Max-Plus measures of delay sensitivity and delay impact (see 2.4.4 and 4.1.3.3) can be utilised in order to achieve timetable robustness. For our proposed timetabling model, we integrate the TCPESP method with the "flexible PESP" (FPESP) method in order to generate event slot timetables on a mesoscopic level. For more details regarding the FPESP method, we refer to the article by Caimi et al. (2011b).

2.4.4 Timetable Performance Measurement

The Timetable Performance Measurement is based on the Max-Plus-Framework (MPF) and helps to analyse the capacity and stability characteristics of time-planned discrete event systems (DES). Our proposal to use the timetable-based Max-Plus framework to determine the schedule's quantitative performance indicators (KPIs) is extraordinary well suited for use in conjunction with the Swiss approach of designing consistent integral timetables.

The Max-Plus-Framework is based on a deterministic process model. This is in contrast to the On-Time approach mentioned, which considers the typical stochastic features of train traffic in the robustness analysis. However, the various capacity-based key figures of the MPF are well suited for a systematic analysis of periodic timetables. Moreover, the concept of critical cycles provides an excellent tool to identify measures to increase the stability of the system under investigation.

Goverde et al. (2011) show how the deterministic Max-Plus framework can be extended to stochastic processes, and thus, this approach belongs to the same model family as the SBB system OnTime. The Max-Plus-Framework complements the existing SBB system approaches. The timetable-based Max-Plus framework for quantitative performance metrics (key performance indicators or KPI's) takes very well into account the Swiss approach to planning integral timetables.

The timetable performance measurement consists of two perspectives:

- A customer-oriented performance perspective, measured by the total travel time, the service intention index (SII) of a timetable scenario (see also section 2.3.1 and section 4.1.3.3 for an application example and Appendix E for a detailed model description).
- An operation stability performance perspective, measured by the critical circuit times, and delay sensitivity of the timetable scenario (see also section 4.1.3.2 and Appendix C)

Both perspectives are implemented in a MATLAB computation tool. We call it Max-Plus Performance Analyzer (MPPA), and it can be invoked directly by the planner after generating a traffic plan in IP-UC2. The input is the EAN, defining all timetable event dependencies and a list of all resulting timetable event times. The Performance Analyzer calculates and displays all quantitative figures of timetable performance together with diagrams for a suitable graphical representation.
In total, the output contains the following performance indicators:

Indicator	Type of information [Units]	Performance aspect	Purpose of information in iterative timetable improvement
Eigenvalue λ of critical circuit	Numerical [min]	operational perspective	The longest cumulative time of coupled process times. This time is divided by the length of the timetable period represents capacity utilisation. According to UIC 406 capacity method and depending on the type of infrastructure, a value between 0.7 and 0.9 should not be exceeded for assuring timetable stability.

Stability	Label [stable/critical/unstable]	operational perspective	Timetable has to be iterated or not
Buffer time	Numerical [min]	operational perspective	Relative measure used for scenario comparison
Recovery value (local)	Numerical [min]	operational perspective	Available time duration for recovery in case of process delay
Delay sensitivity value (local)	Numerical [min]	operational perspective	duration of delay of any other timetable event without having an impact on the respective timetable event
Delay impact value (local)	Numerical [min]	operational perspective	duration of delay of the respective timetable event without having an impact on any other timetable event
Sum of travel time deviations	Numerical [min]	customer perspective	Relative measure used for scenario comparison
Overall travel time	Numerical [min]	customer perspective	Absolute measure used for scenario comparison
SII	Numerical [0,...,1]	customer perspective	Relative measure used for scenario comparison

Table 7: Overview of the performance indicators contained in the output of the framework presented.

2.5 Network segmentation

In order to avoid putting too much effort into entering information that is not needed or focus on the relevant perimeter for the IP timetabling scenario, one has to identify which part of the entire railway network has to be accounted for. The relevant lines and services operating on the subnetwork, which will be affected by the construction sites have to be identified in a first step. In a second step, those lines, which are coupled (e.g. by transfers or technical dependencies) to these affected lines, have to be found. For this line filtering task, the Max-Plus framework (see section 2.3.2) provides well-suited functions which are based on graph connectedness. In the second step, one has to identify the subnetwork nodes which isolate the relevant infrastructure segments from the irrelevant periphery. In this way one obtains a disaggregated subnetwork containing the relevant infrastructure segments and an aggregated subnetwork, representing infrastructure on the macroscopic level (see the dashed square area on the top of Figure 9a).

The disaggregated subnetwork is configured with all mesoscopic details. On this disaggregated subnetwork all train movements are planned in detail for every single IP-scenario. For each line coming from or going beyond the boundary nodes of the disaggregated subnetwork, we create a virtual end station node which is connected by a single section to the corresponding boundary node. The section lengths with the appropriate trip times, the turnaround times of the line outside the disaggregated subnetwork together with the run- and dwell times within the disaggregated subnetwork have to add up to the proper roundtrip time. This segmentation of disaggregated subnetwork and aggregated subnetwork into a new mesoscopic infrastructure model is illustrated in Figure 9b.

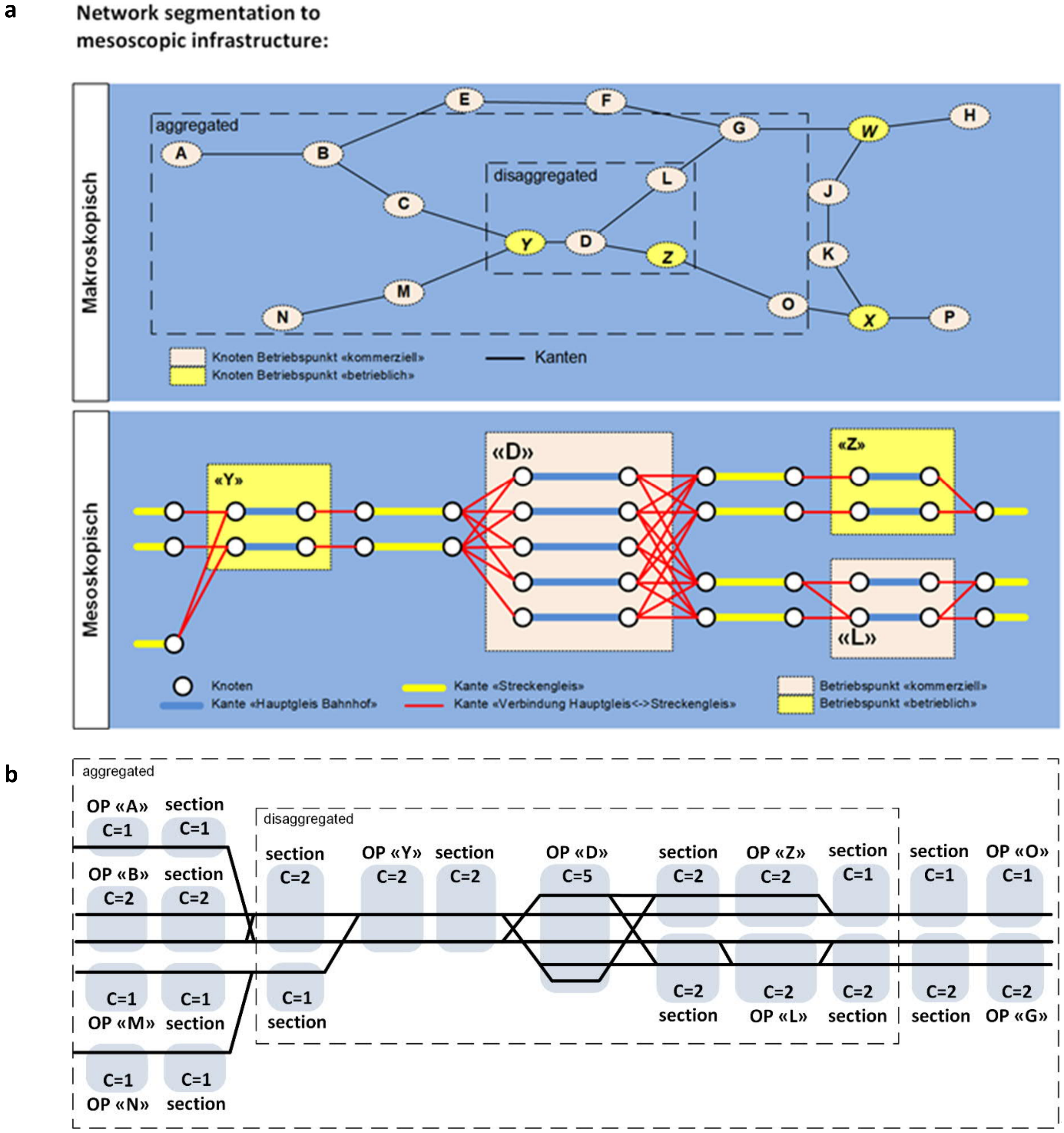
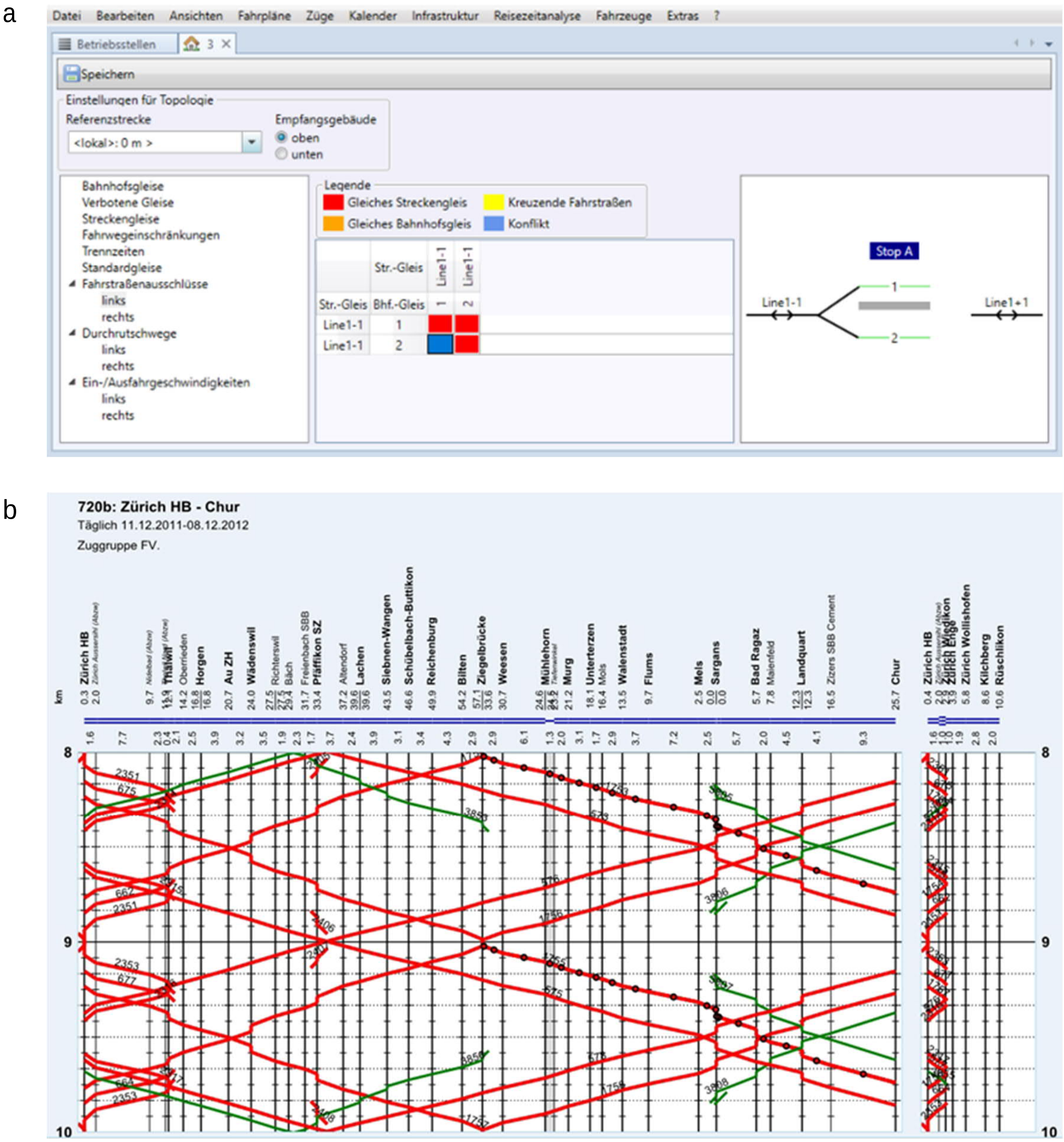


Figure 9: In order to divide the relevant infrastructure for the IP timetabling scenario into a segment with the relevant level of detail and a peripheral part with a more coarse level of information, the railway network is divided into subnetworks. For details, see explanations in the text above.

2.6 Computer-aided timetable generation based on standard planning tool Viriato

One of the main goals of the project was to make the algorithmic timetable generation based on the proposed TCPESP-method available to practitioners. Therefore, the generic configuration of whatsoever timetabling scenario should be possible, using a standard timetabling system such as “Viriato”, which is in use at SBB for service planning (see Viriato Info Folder, 2018). All kinds of relevant timetabling information like line and infrastructure data attributes can be entered easily in the appropriate masks (e.g. track connectivity data such as route exclusions between section and station tracks. Figure 10a shows an example. Moreover, results of computer-aided timetabling can be displayed in standard diagrams such as graphical timetables (see Figure 10b or net graph diagrams. For information that is more detailed we refer to the Viriato User Manual, 2016).



Einstellungen für Topologie

Referenzstrecke
<lokal>: 0 m >

Empfangsgebäude
● oben
○ unten

Legende

🟥 Gleiches Streckengleis
🟧 Gleiches Bahnhofsgleis
🟨 Kreuzende Fahrstraßen
🟦 Konflikt

	Str.-Gleis	Line1-1	Line1-1
Str.-Gleis	Bhf.-Gleis	1	2
Line1-1	1	🟥	🟥
Line1-1	2	🟦	🟥

720b: Zürich HB - Chur
Täglich 11.12.2011-08.12.2012
Zuggruppe FV.

Figure 10: Viriato editor mask for entering timetabling configuration data attributes. a) shows an example of track connectivity on one side of an operation point. Connectivity of station tracks and neighbouring section tracks, as well as potential conflicts, can be entered and configured using appropriate data masks. b) Timetabling result displayed in Viriato graphic timetable diagram. For more examples, see Viriato User Manual (2016).

The standard functionality of Viriato already fulfils most of the data configuration and reporting options that are required for working with the proposed timetabling framework. For our modelling, a very important feature of Viriato is the support of the mesoscopic infrastructure data model.

3 Use cases for computer-aided interval planning

In order to cover most of the cases that can occur during the task of IP, we have to take into account all sequential process steps of timetable planning. The aim is to generate an IP application concept that can be evaluated in practice, e.g. being part of an SR40 “proof of concept”, we propose a set of IP use cases, which correspond to the timetabling process steps introduced in section 2.2.3 (see Figure 4). Depending on the strength of the restrictions resulting from the planned construction or maintenance work, there are two potential situations to consider. Either, an existing and published timetable can be modified in such a way, that the intended service can still be offered. Alternatively, there is no possibility of maintaining the service originally offered, at least concerning the geographical area concerned and a corresponding time window. These two situations have to be reflected by the general process steps of timetable planning (see section 2.2.3) as well as in the use cases of IP. In the general case of timetable development, it frequently happens, that the intended services for a planning horizon cannot be realized without an adaptation or extension of the corresponding infrastructure that will be installed at the time of implementation. In the IP case, because an important piece of rail infrastructure might temporarily be out of service (e.g. track obstruction) or with reduced functionality (e.g. track speed reduction), the published services might not be realisable in the given scenario. In both cases, the intended service is either inconsistent or a feasibility check, taking required resources (e.g. the number of vehicles or available track infrastructure) into account shows, that a relaxation of the offered service cannot be avoided. The use cases that we propose address the two different scenarios.

3.1 Overview of IP use cases

IP-UC0 “SI-relaxation”: The requirement, that an SI relaxation is needed is the relevant trigger to reinitiate IP-use case IP-UC0 “SI-relaxation”, which corresponds to process step 1 of section 2.2.3 “line planning with demand assignment”. IP-UC0 is based on a macroscopic level of infrastructure detail. At this level of detail, the capacity of each linking edge between neighbouring operation points is given as a maximum frequency of train runs in the input to IP-UC0 (see process interfaces in section 2.2.2). In this way, one can make sure, that a restriction of the rail infrastructure (e.g. resulting from speed reductions or tracks, being temporarily out of operation) will result in a relaxed release of the SI compared to the original one.

IP-UC1 “Consistency check of SI”: In order to make a reliable assessment whether or not the intended service is consistent, a corresponding test is executed in IP-UC1 “Consistency check of SI”. If the SI contains conditions that are conflicting with each other, IP-UC1 detects the inconsistency and has to identify those conditions that contribute to the fact that the problem is overdetermined. In case the SI is feasible, IP-UC1 should indicate the remaining capacity range (per operation point and time) of a potential traffic plan. Its feasibility, however, cannot be assessed by IP-UC1, as this use case includes no detailed capacity assignment of the intended line services.

IP-UC2 “Traffic plan with capacity time band”: In this use case, a traffic plan including track assignment is generated, if possible. Also, IP-UC2 may result in the observation, that concerning the available track and vehicle infrastructure, there is no feasible solution, and a relaxation of the SI is required.
If a solution can be generated in IP-UC2 and additionally taking into account the variability of the process duration and the corresponding event times, it might happen that (under operational conditions) the resulting traffic plan is not feasible.
In order to assess a traffic plan concerning the operational feasibility, quantitative measures for timetable stability and capacity utilisation are required. In case, the required operational feasibility is not given, appropriate suggestions for adaptations of the SI and the corresponding traffic plan must be generated.

IP-UC3 "Stability of traffic plan": Quantitative measures, as well as proposals for improvement measures, are elaborated in IP-UC3. The stability performance framework proposed in IP-UC3 also allows providing information about sensitive interdependencies of technical and commercial constraints, which are part of the respective SI. The operational stability is assessed with the help of the Max-Plus framework described in section 4.1.3.3. There we propose the event-based indication of delay impact and delay sensitivity, both being part of the Max-Plus framework, to be used for iterative assignment of event flexibility while generating a "traffic plan with capacity time band" in IP-UC2. Methodological aspects regarding the Max-Plus framework are described in detail in appendices D and E.

IP-UC4 "SII of traffic plan": IP-UC4 results in an index for the customer level of service in order to compare two potential scenarios with each other and in order to find a user (or SLA) defined balance between customer convenience and operational stability measures. We call this measure Service Intention Index (SII). Only if the desired quality of the timetable is achieved concerning the intended service goals for customer convenience and operational reliability, the traffic plan is further used for production planning and customer information. Methodological aspects regarding the Max-Plus framework are described in detail in Appendix F.

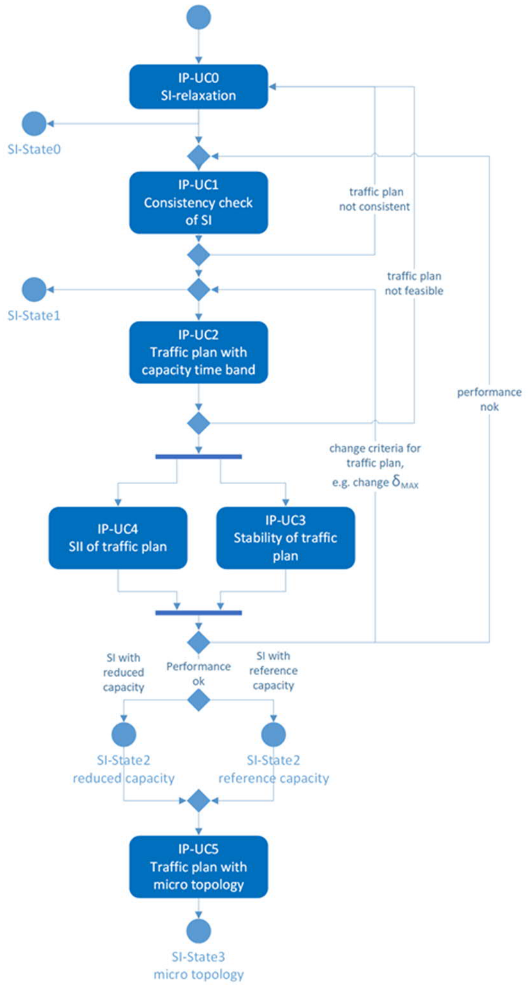


Figure 11: Overview of use cases. The main actions executed within each use case are explained in the text. In this project, we focus on use cases IP-UC1 to IP-UC4. The elaboration of use case IP-UC0 is not part of this project. The elaboration of IP-UC5 is subject to another SR40 part project.

In our application test described in chapter 4, we concentrate on use cases IP-UC2 to IP-UC4 in order to demonstrate the core mechanisms of the computer-aided IP. Figure 11 gives an overview of all relevant IP use cases together with the indicated flow conditions and the corresponding states of the SI at the end of a successful use case execution. One can also see at the lower part of Figure 11, that after the positive assessment of the timetable performance in IP-UC3 and IP-UC4, two or more versions of the SI are existing and can be handled separately. One version for the reference traffic plan (the normal plan for the planning horizon) and the temporary valid and eventually restricted traffic plan (the interval plan).

3.2 Description of use case IP-UC1 to IP-UC4

In this section, we describe the interaction of the different actors in use case IP-UC1 to IP-UC4. Our prototype environment is based on five actors, one of whom is the human planner (see sequence diagram in Figure 12). The remaining actors represent system components. Provided, that the SI is given in SI-state 0, the planner initiates IP-UC1 by entering the SI-data into Viriato (1.). The data entry mainly consists of assigning train runs to infrastructure by indicating the sequence of operation points of its route with and without stops, minimum trip and dwell times. If attributes of station or segment tracks have changed, these amendments also have to be entered.

In the second step (2.), additional SI data, which currently cannot be entered and stored into the Viriato database have to be entered into an R data frame. This additional data concerns connection times between train runs and turnaround times and other time dependencies (e.g. fixed time windows for lines or time separations between lines) as well as event flexibility configurations.

In step (3.), the planner chooses the mode of the computer-aided timetable algorithm. This can be either to make a consistency check of the SI (IP-UC1: no mesoscopic infrastructure information from Viriato is required for this function) or to calculate a traffic plan with capacity time band (IP-UC2). In mode IP-UC2, the planner can choose a suitable version of the objective function (e.g. MinTravel, MaxMinFlex or ConTravel, see also section 2.1.2 for a case-specific choice). After finishing the configuration, the Planning Tool ZHAW launches the GAMS-component, which translates the model configuration into a standard MIP-solver Problem (see GAMS, 2018) and tries to calculate a solution. If no solution can be found (possibly within a pre-defined time window) and does not exist, the GAMS component finishes with an alert informing about the infeasibility of the given problem. In the latter case, a new attempt has to be initialised with step (1.), using an alternative SI configuration.

Otherwise, the resulting traffic plan has to be tested in IP-UC3 (5.) and IP-UC4 (6.) for adherence with the timetable performance requirements (see section 2.3). The planner, using a function of the Planning Tool ZHAW, also initiates these tests. If the timetable performance is sufficient, the state of the SI is set to SI-state 2, and the use case ends. Otherwise, the SI remains in SI-state 1 and a new IP-UC2 iteration is initialised with adjusted configuration parameters.

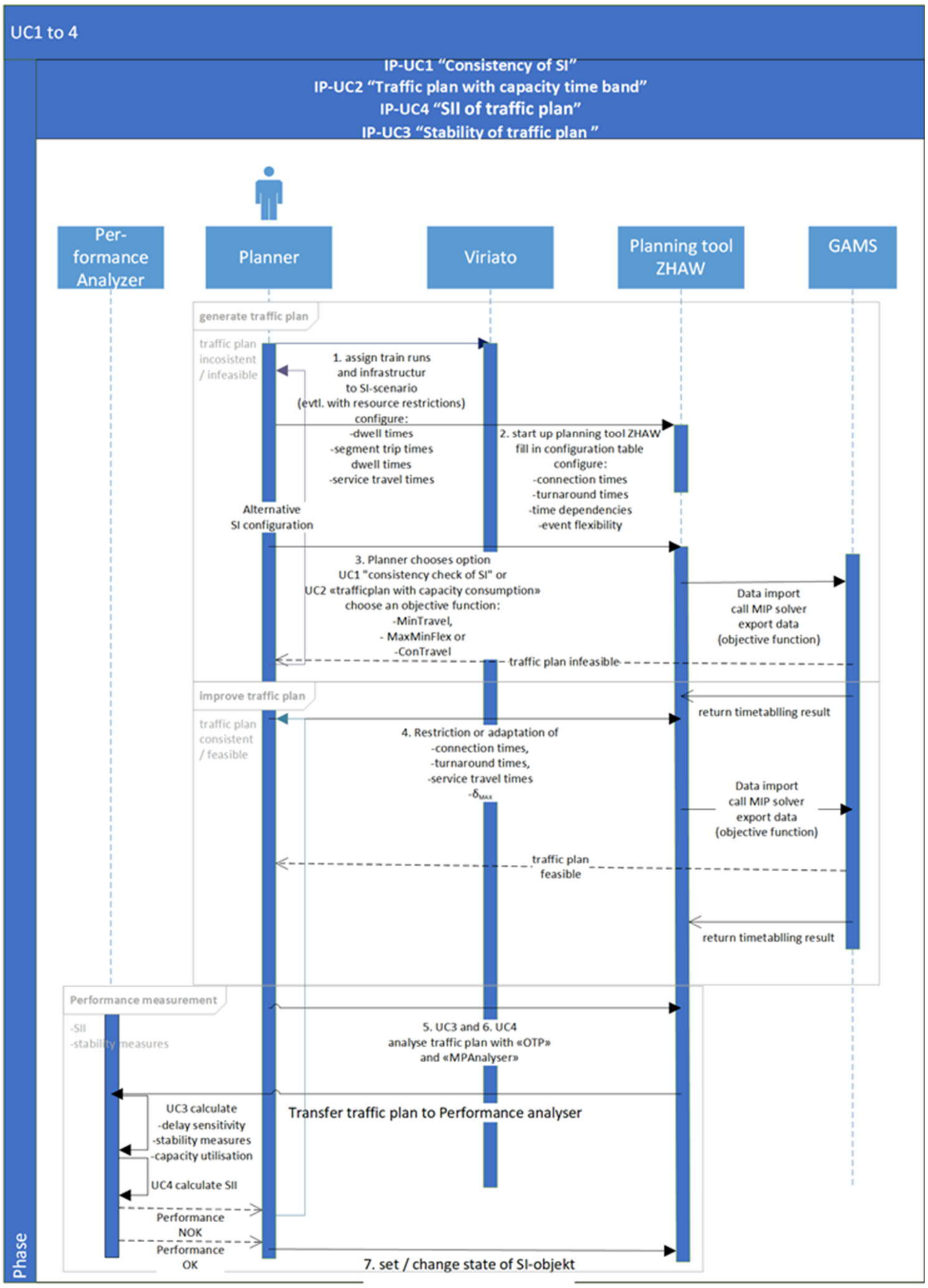


Figure 12: Sequence diagram illustrating the use cases IP-UC1 to IP-UC4 and the tasks and function of the involved planner and system components. For further descriptions see text.

3.3 Application Test Environment

In this section, we describe our system environment for the application test with a detailed overview. Figure 13 shows the information flow for system components used in use cases IP-UC1 to IP-UC4. The planner (“Verkehrsplaner” in the box “Verkehrsplanung mit Viriato”) configures the SI by manipulating SI-parameters in the Viriato - database using the Viriato-UI and the R – data frames and considering the required timetable performance measures. The Planning Tool ZHAW reads the relevant data from the Viriato database and writes them into a table with an adequate input format for GAMS. In case IP-UC1 is executed, the GAMS-component is configured with a PESP configuration data set. In case IP-UC2 is executed, the GAMS-component is configured with a TCFPESP configuration data set.

If the traffic plan generated by GAMS feasible, it will be analysed by the MPPA (Max-Plus Performance Analyser) (IP-UC3). Its output is a set of operational performance indicators that are stored in a database. These performance indicators as well as the SII (i.e. the output of the SI-assessment (see the box "SI-Bewertung (Verkehrsplan)" in Figure 13) is used by the planner in order to iteratively improve the traffic plan.

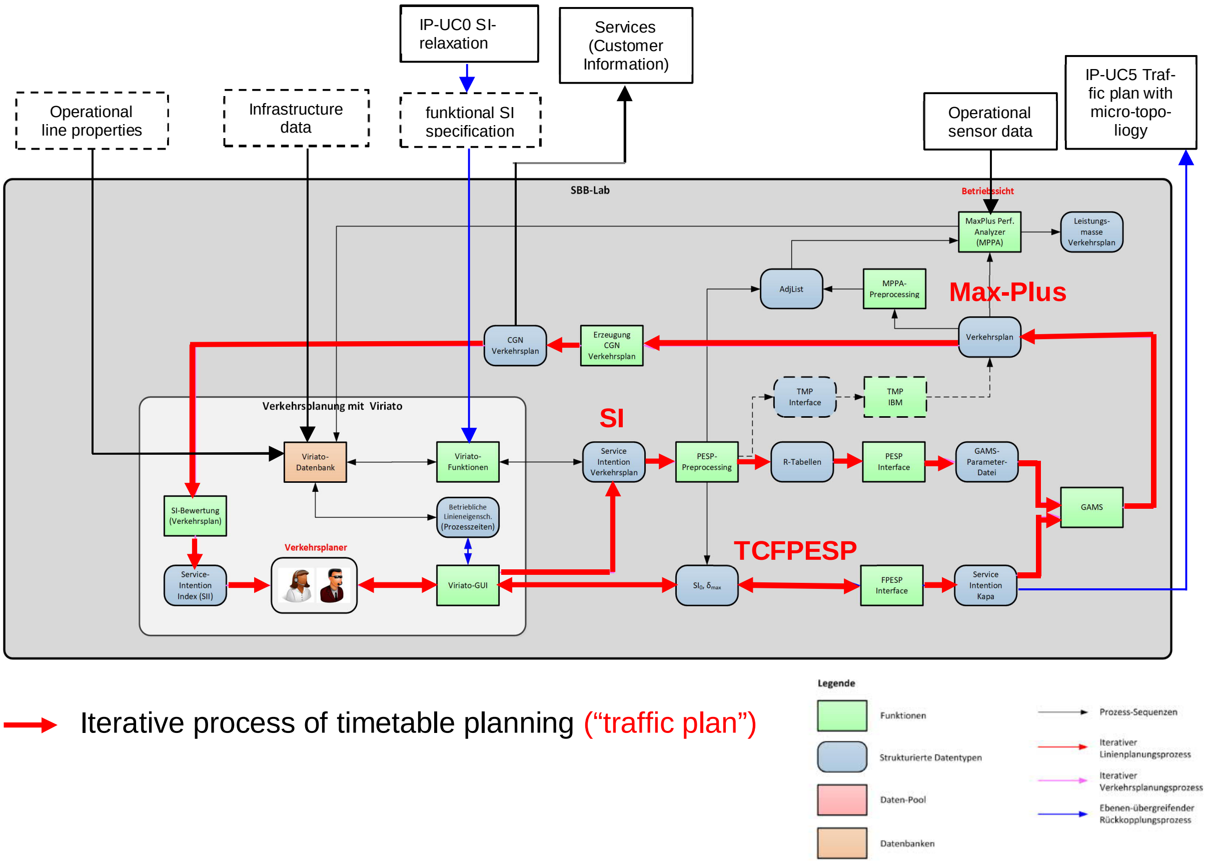


Figure 13: Activity diagram showing activities used in use cases IP-UC1 to IP-UC4. See text for further explanations regarding the flow of information. The entire system environment is shown in Appendix C.

The above activity diagram in Figure 13 is part of a complete overview of the system environment, which also includes the components involved in the SI-relaxation of IP-UC0. The elaboration of the additional components is part of the SR40-supplement to the current project. The complete system overview is shown in Appendix C.

4 Application and validation of the framework

This chapter aims to explain and apply some important methodological steps already mentioned in chapter 0 and to validate the overall framework developed as a proof of concept (POC). We do this by means of two examples and some corresponding explanations. In section 4.1 we start with a small-scale network, which was inspired by a simple but illustrative timetable scenario presented in Goverde (2010) and substantially extended for our purpose, where the main focus is on testing IP-UC3 and IP-UC4 as described in chapter 3. We do not explain the methodological detail here again, but instead refer to Appendix D to G, where the Max-Plus framework (Appendix D and E), the computation of the so-called Service Intention Index (Appendix F), and the computation of the so-called Cumulative Delay Impact and Cumulative Delay Sensitivity measures (Appendix G) are explained. In section 4.2, a real-world example shows in great detail all steps required to compute a timetable. In this section, an important part is on testing IP-UC1 and IP-UC2.

4.1 Small-scale test network

4.1.1 Introduction

The test network was inspired by a timetable scenario with two stations and three lines, presented in Goverde (2007) and is shown In Figure 14. The network consists of two stations (Station A and Station B) and three lines. The eight nodes represent departure events (1, 2, 3, 4) and arrival events (5, 6, 7, 8). In station A, line 1 has a connection of two minutes with line 2 (in both directions), and in station B, line 3 has a connection of again two minutes (in both directions) with line 2. The numbers of the events are indicated within the nodes, and the scheduled event times are shown as italic numbers above (node 5, 7, 3, 4) and below (node 1, 2, 6, 8) the corresponding event.

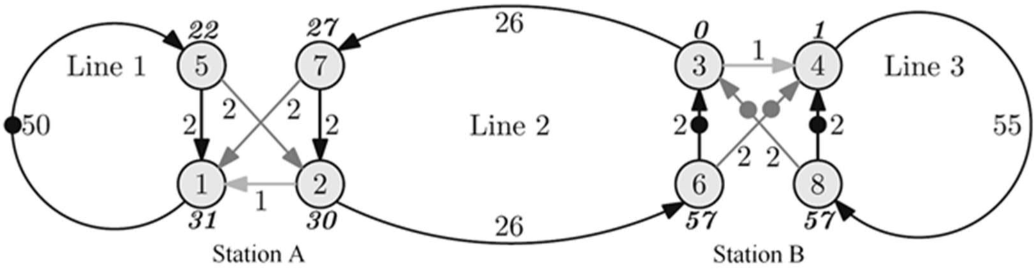


Figure 14: Network example from Goverde (2010, p. 272) that forms the basis for our small-scale test network.

Each line has a dwell time of two minutes as indicated by the arcs connecting the arrival and departure events of a line at a given station. The running time of line 2 from B to A and vice versa is 26 minutes and is indicated at the corresponding arcs. The total roundtrip of line 1 takes 52 minutes, the one of line 3 takes 57 minutes. Finally, there is a headway of two minutes separating the departures of line 2 and 1 in station A and the departures of line 2 and 3 in station B. The dots on the activity arcs indicate that there is a jump in the hourly period between the two corresponding events.

In order to demonstrate the principles of use cases IP-UC3 and IP-UC4, we adapted and extended the network shown above as an explicit mesoscopic topology and specified an SI.

4.1.2 Network representation

As mentioned in the previous section, the network shown in Figure 14 is limited to the details of the two station nodes (Station A and Station B) connected by line 2, and three stops, Stop A and Stop AT (served by line 1) and Stop BT (served by line 3). The synchronisation conditions (line connections) for the test case that are relevant to planning are limited to stations A and B. The graph of the network for our example scenario is illustrated in Figure 15. The corresponding event activity graph is shown in Figure 16a. There are still transfers of passengers from line 2 to line 1 and vice versa in station A on from line 2 to line 3 and vice versa in station B. However, as an extension of the network of Figure 14, line 1 now has two stops: Stop A and Stop AT. Furthermore, the outbound service of line 1 is connected to an inbound service via a turnaround constraint in Stop AT. Line 3 has one stop only, i.e. Stop BT, where again the turnaround activity from the outbound to the inbound service is supposed to happen.

The infrastructure on which we implemented this example is shown in Figure 9c. While line 2 is running on a double track section, line 1 and line 2 are running on single-track sections. In order to illustrate the data configuration of the timetabling problem, according to IP-UC2, we implement the original line scenario of Goverde with a slightly modified travel-time configuration. The period of each train run (service with service ID) is indicated in Table 8.

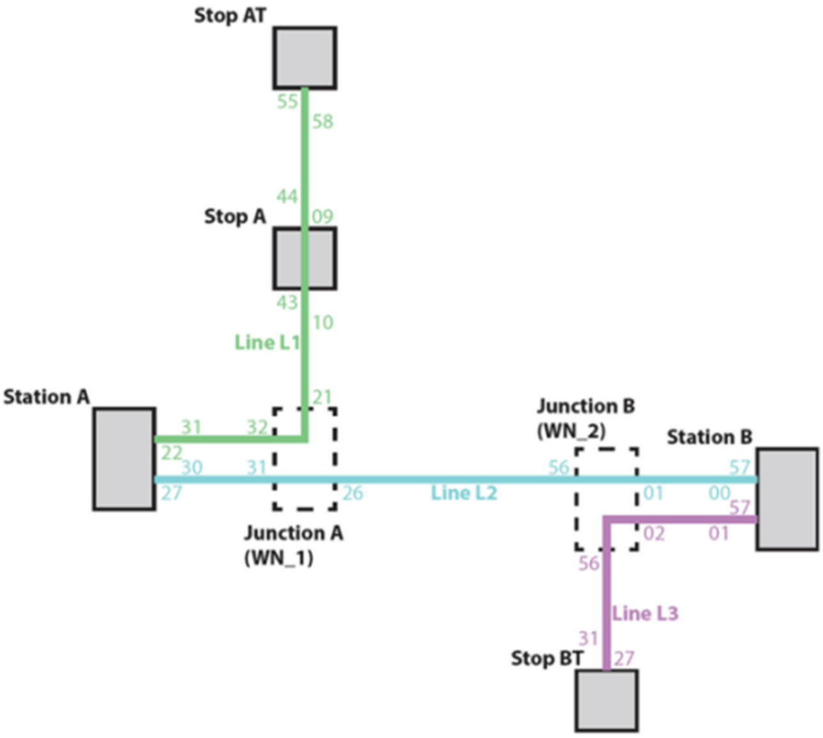


Figure 15: Network graph indicating line departure and arrival times at stations (continuous boundary line) and junctions (broken boundary line).

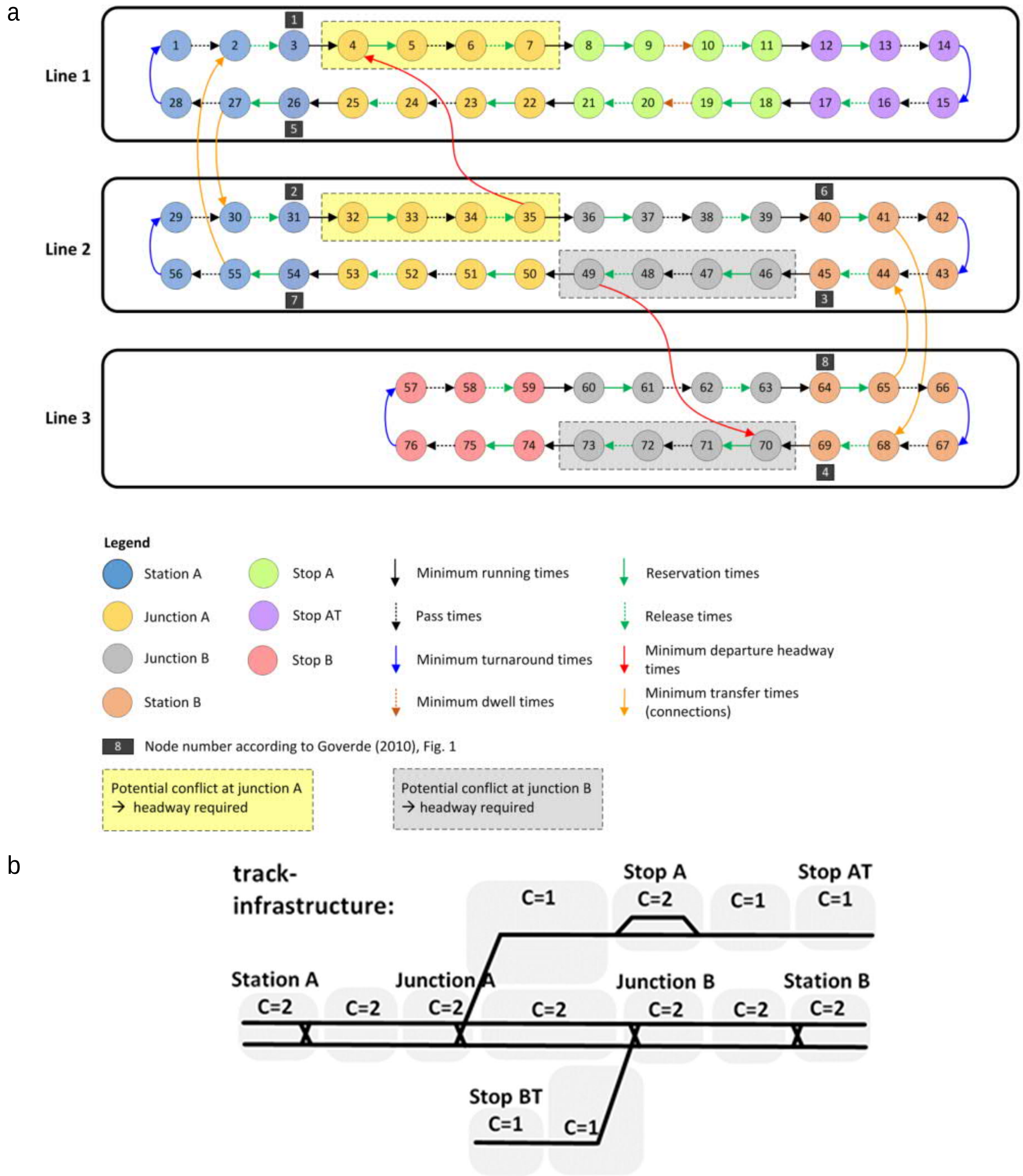


Figure 16: a) Expanded event activity network with operational (headways) and commercial (line connections) dependencies based on Goverde (2007), with extensions. Line 1: serving Station A, Stop A and Stop AT; Station A: connecting Line 1 and Line 2; Line 2: connecting Stations A and B; Station B: connecting Line 2 and Line 3; Line 3: serving Station B and Stop BT. b) Track infrastructure with the mesoscopic section topology and an indication of track capacities (indicated by the corresponding number of horizontal lines) for each operating point (indicated as shaded boxes).

The SI of our network here offers an hourly service of line 2 between major stations A and B with connections to and from line 1 in station A and to and from line 3 in station B. A complete rotation of Line 1 and 2 lasts 120 minutes, one of Line 3 lasts 60 minutes. Therefore, two vehicles are needed for rotations of line 1 and 2 and only one vehicle is needed for line 3. Line services with train runs and corresponding periodicity and minimum circulation times, as indicated in Table 8.

Line ID	Service ID	Min. trip time	Period
1	11	50	60
1	12	50	60
1	13	50	60
1	14	50	60
2	21	50	60

Line ID	Service ID	Min. trip time	Period
2	22	50	60
2	23	50	60
2	24	50	60
3	31	20	60
3	32	20	60

Table 8: Line services (train runs) with respective minimal trip times and periods. Odd numbers of Service IDs indicate train runs in one direction; even numbers indicate train runs of the same line in the opposite direction.

4.1.3 Test cases and results

To investigate the framework, in particular concerning IP-UC 3 and IP-UC 4, we defined four test cases outlined in Table 9. In section 4.1.3.1 an example of a resulting is discussed, in section 4.1.3.2 we assess the system from an operator's point of view (IP-UC3), and in section 4.1.3.3 we show the assessment from a customer's point of view (IP-UC4).

Case 1	Description	Relaxation action
1	Services are provided according to the reference timetable	none
2	Services are provided according to the reference timetable. However, due to a construction site, the duration of trips between Stop A and Stop AT (in both directions) takes five more minutes and hence the system is no longer stable.	none
3	Same services and restrictions as for case 2. Although the applied relaxation action leads to a decoupling of line 1 from the rest of the system, line 1 is still critical, i.e. the sum of the process times are identical or larger than the period of the system.	To limit the impact of the unstable line 1 on the rest of the system, connections between line 1 and 2 in Station A are no longer considered for planning.
4	Same services and restrictions as for case 2. Compared to case 3, the relaxation applied substantially increases the stability of the system. Furthermore, its overall travel time decreases.	To increase the stability of the system, line 1 is no longer serving Stop AT but performs a turnaround at Stop A instead. The connection between Stop A and Stop AT is maintained by some bus service, where the travel time between Stop A and Stop AT (in both directions) is 20 minutes longer than for the regular timetable. The bus service is not part of the timetabling but runs independently.

Table 9: Description of the four test cases together with the corresponding relaxation actions and their expected effects on the system stability.

4.1.3.1 Timetable generation

Figure 17 illustrates the results of the TCPESP algorithm for test case 1 (as specified in Table 9). In addition to the output of the conventional PESP algorithm given by arrival and departure event times, the result that we obtain from our TCPESP model includes track assignment information for each train run. The rail infrastructure of our test scenario consisted mainly of two single-track lines (Line 1 and 3) and one double track line (Line 2). At Stop A we have to tracks admitting crossings of Line 1 in opposite directions. We indicate the resulting track assignment by assigning track numbers (T1 and T2) to each train run during their presence on a given track section, as shown in the upper part of each line diagram. If there is only one track available at a working point, there is only one horizontal grey line (T1), if there are two tracks, then there are two lines (T1 and T2). Each arrow indicates the direction of an individual train run. The line types in the track diagrams correspond to line types in the time diagrams.

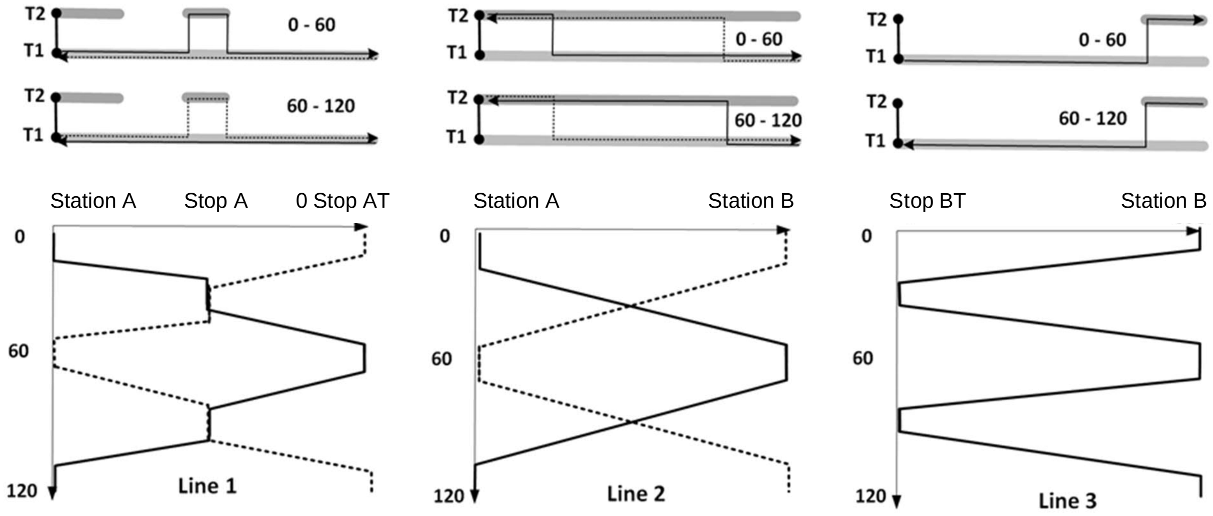


Figure 17: Scheduling results obtained from our TCPESP model. A train diagram with the arrival and departure event times is plotted together with the track assignment Vertical axis: the time between 0 and 120 minutes, horizontal axis: sequential location. T1 and T2 with grey shaded horizontal lines on top of each location-time diagram indicates track assignment for each vehicle circulation of the three given lines (arrows in both directions).

From Figure 17 we see, that the TCPESP algorithm only permits counter-rotating train runs to meet in double track sections (Line 1) and the connecting train runs to meet in a station on neighbouring tracks (platforms; Station A: line 1 and 2, Station B: line 2 and 3).

4.1.3.2 System assessment from an operator’s point of view (UC3)

One of the most important factors for operators is the stability of their system. A public transport system with a periodic timetable can be regarded as a Discrete Event System (DES). Over the last around two decades, the application of the so-called Max-Plus algebra (MPA) has become a well-established method to determine, amongst other factors, the stability of the system. A very good introduction of the fundamental concept of the MPA can be found, for example, in Goverde (2005) or Heidergott et al. (2006).

Some of the key benefits of the MPA is that various important characteristics of a DES, or to be more precise, of DES with deterministic process times) can be calculated analytically. We can distinguish the following three cases to assess the stability of a system, depending on the eigenvalue λ_0 of the corresponding critical cycle (see also Appendix E, section E2 ff.) compared to the cycle length T of the system:

$\lambda_0 < T$ → the system is stable,

$\lambda_0 = T$ → the system is critical,

$\lambda_0 > T$ → the system is unstable.

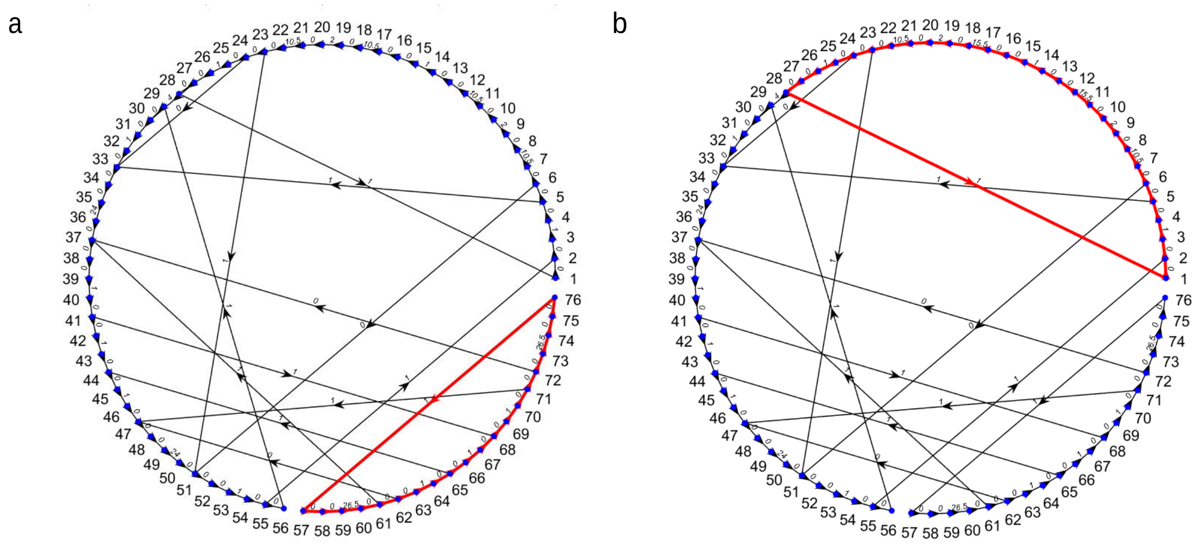
Each DES considered here has ≥ 1 circuits. A circuit is a closed sequence of processes (arcs) connecting the nodes of the system (i.e. arrival and departure events). The critical circuit is the one, which is the most critical regarding disruptions of events.

Besides knowing the stability of the system under investigation, it is equally important to have reliable information on its buffer times available in case of disruptions. The buffer time specifies the amount of time left at a node (event) given some specified disturbance of any another node (arrival or departure event). As long as the buffer is > 0 for all pairs of nodes, the state of the system is stable. This information can be used in IP-UC2 to control the flexibility in the TCFPESP model, i.e. to force the flexibility to be at the right place.

Below we highlight some of these aspects for the test cases specified in Table 9 by applying our MPPA framework (outlined in detail in appendices D and E):

- The graph of the network with its corresponding critical circuit for the four test cases (Figure 18),
- The recovery matrix for test case 1 (Figure 19),
- The cumulative delay impact (CDI) and the cumulative delay sensitivity (CDS) for test case 1 (Figure 21).

Based on the timetables calculated, we first show the resulting network graphs together with their corresponding critical circuits (red edges in Figure 18).



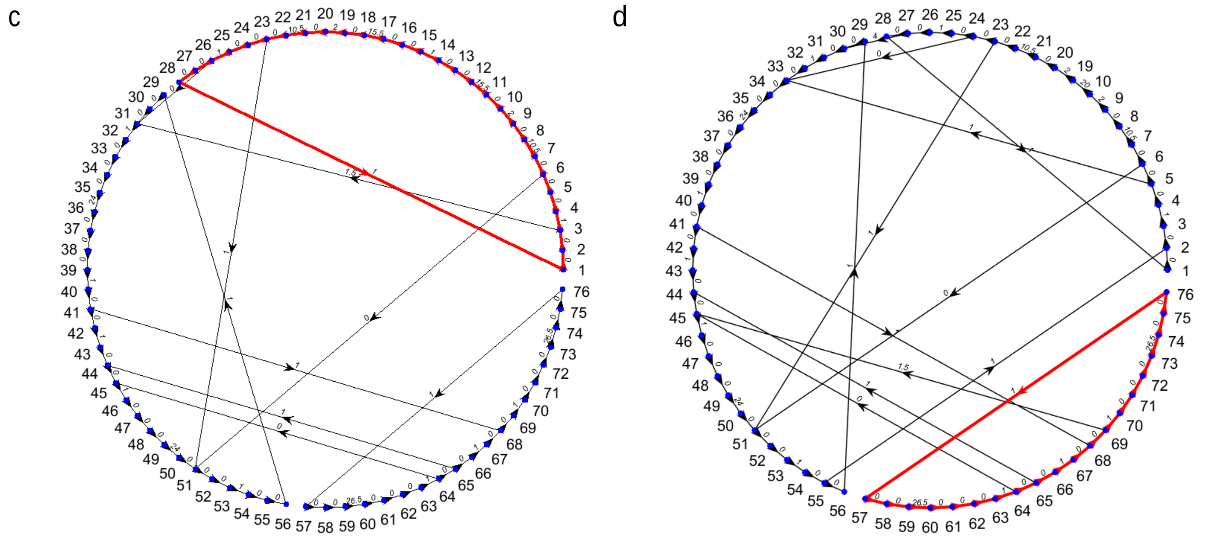


Figure 18: Representation of the network graphs of test cases 1 to 4, including the specific critical circuits (red lines). a: case 1, b: case 2, c: case 3, and d: case 4. The nodes numbers are in accordance with those shown in Figure 16a, whereas for case 4 node numbers 11 to 18 are missing, as Stop AT is not served. We see that arcs of the graphs differ much between the four figures. This is due to the case-specific generation of the timetable (see section 4.2 for details) and its resulting process arcs (e.g. headways or connections; see types specified in Appendix D, section D4).

The sum of the critical process times for the four cases, i.e. their eigenvalues, are as follows: cases 1 and 4: 57 minutes (system is stable with an overall buffer of at least 3 minutes), cases 2 and 3: 60 minutes (system is in a critical state, i.e. even a small disturbance of the events along the critical leads to an instability. These values are also shown in the overall assessment of the test cases in Table 18. The critical circuits can be used to derive the relaxation actions from Table 9 since actions breaking the critical circuits leads to a greater buffer in the system.

Having a critical or even unstable system means that there is not enough buffer time available to handle disruptions. Hence, it is a good starting point to have a closer look at the so-called recovery matrix. The recovery matrix tells us how much buffer time is available between any of the events of the system. In Figure 19, the recovery matrix is depicted for test case 1. The size of the matrix is 76 times 76, as the system of test case 1 consists of 76 events. Along the vertical axis, we see the number of the so-called incoming nodes, i.e. the node entering a process arc, whereas along the horizontal axis we see the number of the so-called outgoing nodes, i.e. the node leaving a process arc. The colour of the cells represents the actual buffer time between an incoming and outgoing node, i.e. the time of a disturbance of an incoming node such that there is a delay of the outgoing node. It can easily be seen that for a stable system, all entries of the matrix need to be ≥ 0.

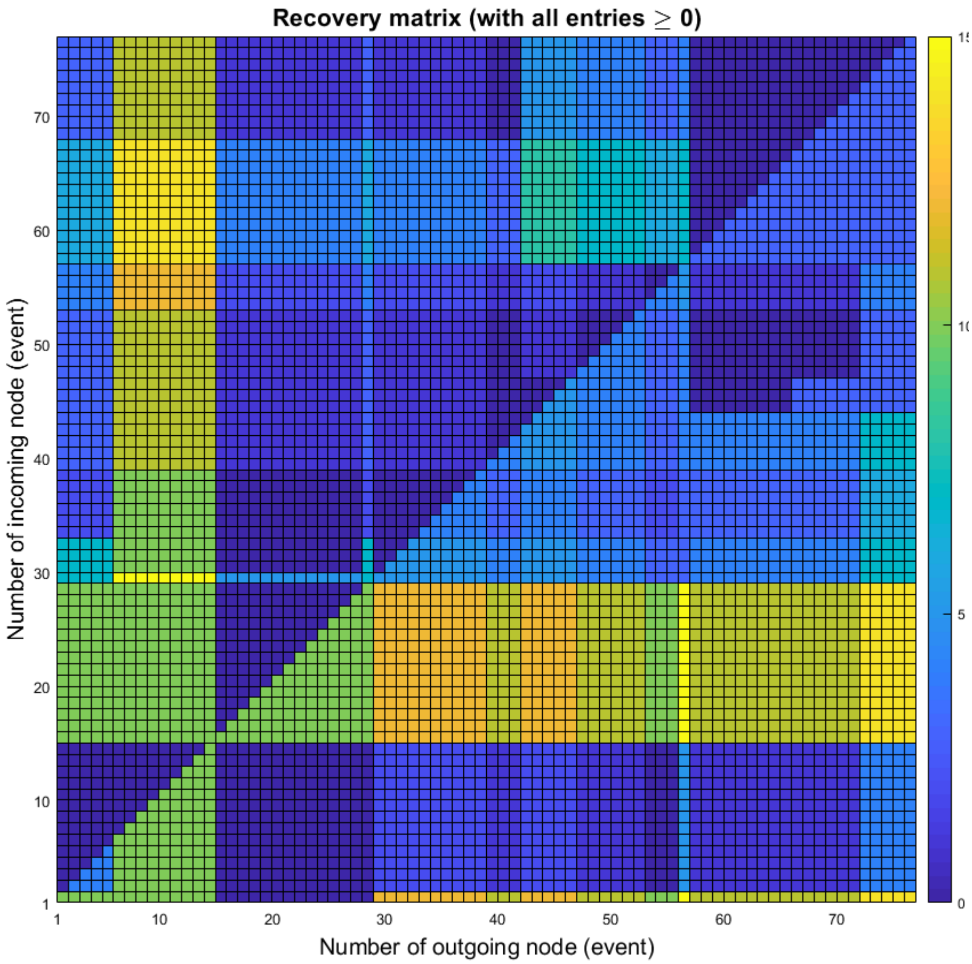


Figure 19: Recovery matrix for test case 1.

Based on the recovery matrix we next determine (i) the overall impact of an (incoming) event on the system and (ii) the overall delay sensitivity of an (outgoing) depending on the system. To achieve this, we introduce the so-called cumulative delay impact (CDI) and the cumulative delay sensitivity (CDS), respectively.

The cumulative delay impact captures the overall impact of a delay at an (incoming) event to all other events of the system. To illustrate this, we consider a row of the recovery matrix. Let us apply some delay to the incoming node of, e.g. three minutes; we then add up for all events the positive difference of this delay and the entry of the recovery matrix. In case the entry is larger or equal than the delay, the delay has no negative effect on the outgoing event. Otherwise, we capture the resulting difference in the value of the delay and the entry in the matrix. For CDI, we finally add up the positive differences over all outgoing events. For a detailed mathematical description, please check Appendix G.

The information gathered in the recovery matrix, the cumulative delay impact and the cumulative delay sensitivity can be used to control the loop between IP-UC3 and IP-UC4 (see chapter 3.2). This aspect is not further elaborated in this report but will be the subject of research in the future (see chapter 5.2).

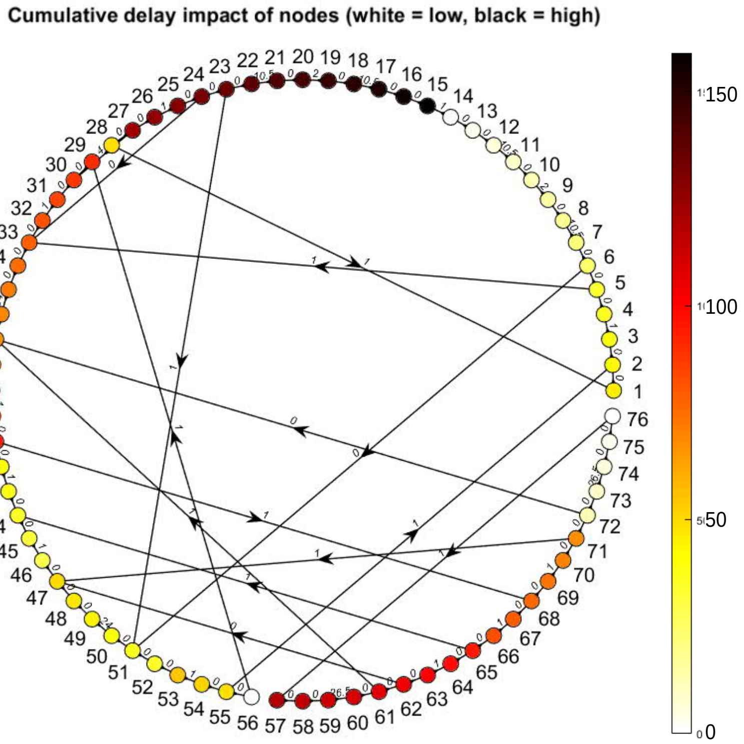


Figure 20: Cumulative delay impact for all nodes of the system for test case 1. The colour represents the value of the cumulative delay impact in minutes.

The CDI captures how much a delay at some (incoming) event affects all other network events. The computation of the cumulative delay sensitivity is very similar. It captures how much a delay applied to all (incoming) events affects a selected (outgoing) timetable event. To illustrate this, we consider a column vector of the recovery matrix. Let us apply a delay to each incoming node of again three minutes. We next add up the positive difference of the delay compared to the corresponding entry of the recovery matrix. In case the entry is larger or equal than the delay, the delay has no negative effect on the outgoing event. Otherwise, we add the resulting difference.

To get the CDS for a specific outgoing event, we add up the positive differences over all incoming events along the corresponding row in the recovery matrix. Again, in Appendix G, we provide a detailed description of both the CDI and the CDS.

Based on the above explanations, we can see that assigning more flexibility to events with high delay impact and at the same time assigning less flexibility to events with low delay impact, are supposed to increase the robustness of the timetable.

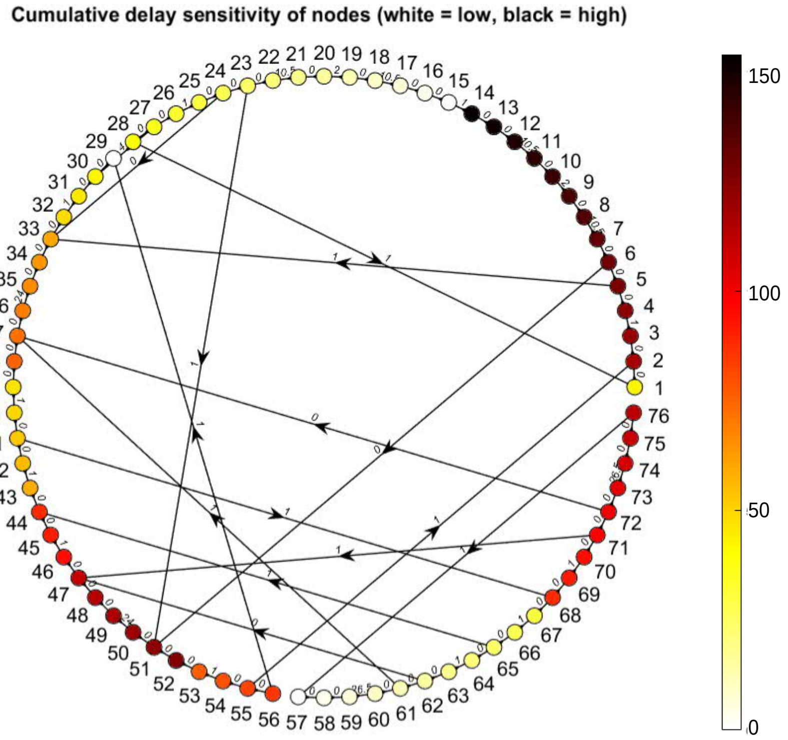


Figure 21: Cumulative delay sensitivity for all nodes of the system for test case 1. The colour represents the value of the cumulative delay impact in minutes.

4.1.3.3 System assessment from a customer's point of view (IP-UC4)

To assess the perceived quality of a journey, various factors like, for example, the travel time from origin to destination, the overall number of transfers, the waiting time or the comfort in the transport vehicles might be considered. In this project, we focus on travel time only, as it is the most useful measure in the context presented here. The travel times required, and the Service Intention Index are computed according to the detailed description in Appendix F.

The assessment of the three disponed cases (2, 3 and 4) with the planned one (1) will be described in the next three sections as follows:

- Comparison of cases 1 and 2 in detail (section a); In addition to the travel time differences and the SII for each origin-destination combination shown (see subsections b and c, below), we additionally present travel time matrices for cases 1 and 2. This shall help to understand the system better and to be able to compare the results with those presented in other sections of chapter 4,
- Comparison of cases 3 and 1 (section b),
- Comparison of cases 4 and 1 (section c).

a. Comparison of cases 2 and 1 in detail

		Line 1			Line 2		Line 3	
		Station A, platform 1	Stop A	Stop AT	Station A, platform 2	Station B, platform 1	Stop BT	Station B, platform 2
Line 1	Station A, platform 1	0	692	1442	30	1625	3335	1626
	Stop A	692	0	632	695	2552	4262	2553
	Stop AT	1442	632	0	1445	3302	5012	3303
Line 2	Station A, platform 2	30	695	1445	0	1622	3332	1623
	Station B, platform 1	1685	2432	3182	1682	0	1653	30
Line 3	Stop BT	3395	4142	4892	3392	1653	0	1652
	Station B, platform 2	1686	2433	3183	1683	30	1652	0

Table 10: Travel time between all stations/stops for test case 1.

Origin-destination travel times [in seconds]		Line 1			Line 2		Line 3	
		Station A, platform 1	Stop A	Stop AT	Station A, platform 2	Station B, platform 1	Stop BT	Station B, platform 2
Line 1	Station A, platform 1	0	692	1742	30	1565	3335	1566
	Stop A	692	0	932	695	2492	4262	2493
	Stop AT	1742	932	0	1745	3542	5312	3543
Line 2	Station A, platform 2	30	695	1745	0	1562	3332	1563
	Station B, platform 1	1565	2372	3422	1562	0	1653	30
Line 3	Stop BT	3335	4142	5192	3332	1653	0	1652
	Station B, platform 2	1566	2373	3483	1583	30	1652	0

Table 11: Travel time between all stations/stops for test case 2.

Origin-destination travel times [in seconds]		Line 1			Line 2		Line 3	
		Station A, platform 1	Stop A	Stop AT	Station A, platform 2	Station B, platform 1	Stop BT	Station B, platform 2
Line 1	Station A, platform 1	0	0	300	0	-60	0	-60
	Stop A	0	0	300	0	-60	0	-60
	Stop AT	300	300	0	300	240	300	240
Line 2	Station A, platform 2	0	0	300	0	-60	0	-60
	Station B, platform 1	-120	-60	240	-120	0	0	0
Line 3	Stop BT	-60	0	300	-60	0	0	0
	Station B, platform 2	-120	-60	240	-120	0	0	0

Table 12: Travel time differences for all stations/stops between case 2 (according to Table 10) and case 1 (according to Table 11). Bold numbers highlight the differences caused by the construction site. Small deviations (60 - 120 sec.) are caused by rounding effects in timetable event time calculation but are not a problem.

Origin-destination travel times [in seconds]		Line 1			Line 2		Line 3	
		Station A, platform 1	Stop A	Stop AT	Station A, platform 2	Station B, platform 1	Stop BT	Station B, platform 2
Line 1	Station A, platform 1	0.00	0.00	20.80	0.00	-3.69	0.00	-3.69
	Stop A	0.00	0.00	47.47	0.00	-2.35	0.00	-2.35
	Stop AT	20.80	47.47	0.00	20.76	7.27	5.99	7.27
Line 2	Station A, platform 2	0.00	0.00	20.76	0.00	-3.70	0.00	-3.70
	Station B, platform 1	-7.12	-2.47	7.54	-7.13	0.00	0.00	0.00
Line 3	Stop BT	-1.77	0.00	6.13	-1.77	0.00	0.00	0.00
	Station B, platform 2	-7.12	-2.47	7.54	-7.13	0.00	0.00	0.00

Table 13: Travel time differences between all stations/stops of test case 2 and test case 1 relative to the values of case 1. The bold numbers highlight again, where the delay has the highest impact.

b. Comparing case 3 and case 1

Origin-destination travel times [in seconds]		Line 1			Line 2		Line 3	
		Station A, platform 1	Stop A	Stop AT	Station A, platform 2	Station B, platform 1	Stop BT	Station B, platform 2
Line 1	Station A, platform 1	0	0	300	0	-60	-60	-60
	Stop A	0	0	300	0	1050	1050	1050
	Stop AT	300	300	0	300	1350	1350	1350
Line 2	Station A, platform 2	0	0	300	0	-60	-60	-60
	Station B, platform 1	-120	2550	2850	-120	0	0	0
Line 3	Stop BT	-120	2550	2850	-120	0	0	0
	Station B, platform 2	-120	2550	2850	-120	0	0	0

Table 14: Travel time differences for all stations/stops between case 3 and case 1. Bold numbers highlight the differences caused by cancelling the connections between Line 1 and Line 2. As a consequence, the timetable does not synchronise anymore between these two lines, which leads to some asymmetrically longer travel times from stations/stops along line 1 on the one hand, and along lines 2 and 3 on the other hand.

Origin-destination travel times [in seconds]		Line 1			Line 2		Line 3	
		Station A, platform 1	Stop A	Stop AT	Station A, platform 2	Station B, platform 1	Stop BT	Station B, platform 2
Line 1	Station A, platform 1	0.00	0.00	20.80	0.00	-3.69	-1.80	-3.69
	Stop A	0.00	0.00	47.47	0.00	41.14	24.64	41.13
	Stop AT	20.80	47.47	0.00	20.76	40.88	26.94	40.87
Line 2	Station A, platform 2	0.00	0.00	20.76	0.00	-3.70	-1.80	-3.70
	Station B, platform 1	-7.12	104.85	89.57	-7.13	0.00	0.00	0.00
Line 3	Stop BT	-3.53	61.56	58.26	-3.54	0.00	0.00	0.00
	Station B, platform 2	-7.12	104.81	89.54	-7.13	0.00	0.00	0.00

Table 15: Travel time differences between all stations/stops of test case 3 and test case 1 relative to the values of case 1. The bold numbers highlight again, where the delay has the highest relative impact.

c. Comparing case 4 and case 1

Origin-destination travel times [in seconds]		Line 1			Line 2		Line 3	
		Station A, platform 1	Stop A	Stop AT	Station A, platform 2	Station B, platform 1	Stop BT	Station B, platform 2
Line 1	Station A, platform 1	0	0	1200	0	-30	-30	-30
	Stop A	0	0	1200	0	-30	-30	-30
	Stop AT	1200	1200	0	1200	1200	1200	1200
Line 2	Station A, platform 2	0	0	1200	0	-30	-30	-30
	Station B, platform 1	-120	-60	1200	-120	0	0	0
Line 3	Stop BT	-30	30	1200	-30	0	0	0
	Station B, platform 2	-120	-60	1200	-120	0	0	0

Table 16: Travel time differences for all stations/stops between case 4 and case 1. Bold numbers highlight the differences caused by the shuttle bus introduced to connect Stop AT to Stop A, where trips are assumed to take 20 minutes longer than the journey by train. However, it shall be mentioned that the shuttle service is not part of the timetable.

Origin-destination travel times [in seconds]		Line 1			Line 2		Line 3	
		Station A, platform 1	Stop A	Stop AT	Station A, platform 2	Station B, platform 1	Stop BT	Station B, platform 2
Line 1	Station A, platform 1	0.00	0.00	83.22	0.00	-1.85	-0.90	-1.85
	Stop A	0.00	0.00	189.87	0.00	-1.18	-0.70	-1.18
	Stop AT	83.22	189.87	0.00	83.04	36.34	23.94	36.33
Line 2	Station A, platform 2	0.00	0.00	83.04	0.00	-1.85	-0.90	-1.85
	Station B, platform 1	-7.12	-2.47	37.71	-7.13	0.00	0.00	0.00
Line 3	Stop BT	-0.88	0.72	24.53	-0.88	0.00	0.00	0.00
	Station B, platform 2	-7.12	-2.47	37.70	-7.13	0.00	0.00	0.00

Table 17: Travel time differences between all stations/stops of test case 4 and test case 1 relative to the values of case 1. The bold numbers highlight again, where the delay has the highest relative impact.

4.1.3.4 Overall assessment of the system

The results of the small-scale test case presented in this section are summarised in Table 18. The aim is to show both the assessment from an operational as well as from a customer point of view. As outlined in the description of the four test cases in Table 9.

As test case 1 is the planned one, there are, of course, no travel time deviations, and the SII is 0, which is the optimal value. Case 2 has some delay between Stop A and Stop B leading to some minor decrease of the SII but, more important, to a critical state of the system with an eigenvalue of 60 minutes.

To simulate the relaxation performed usually (automatically) by the line planning procedure, with test cases 3 and 4 we introduced two operational options to get the system stable again and with moderate disadvantages for the customers regarding the travel time.

For case 3 we see that the state of the system remains critical but, due to the decoupling of Line 1 and Lines 2 and 3, disturbances along Line 1 have now no effects on the rest of the system. However, the increase in the sum of travel time and hence the SII is pretty high. For case 4, Line 1 and the rest of the system remains connected and with the shuttle bus introduced the system becomes even stable again. Furthermore, as the overall travel time and the SII are small compared to case 3, the relaxation applied with case 4 is preferable over the one of case 3.

Case	Assessment from an operational point of view			Assessment from a customer point of view	
	Eigenvalue of critical circuit [minutes]	Stability [-]	Buffer time [minutes]	Sum of travel time deviations [minutes]	Service Intention Index SII [%]
1	57	stable	3	0	0.00
2	60	critical	0	38	2.70
3	60	critical	0	402	28.59
4	57	stable	3	225	15.99

Table 18: Overview of the assessment of the four test cases.

In the small-scale test network, we demonstrate the basic principle of the automatic timetable generation and the process of finding a feasible solution in case of reduced infrastructure availability. In order to compare three different SI scenarios (theoretically resulting from use case IP-UC0) concerning the performance criteria introduced in section 2.3, we used the MPPA (Max-Plus Performance Analyser).

In the next section, we want to show how the planner can develop a feasible timetable which is flexible enough to tolerate a reduction of track capacity due to two different construction intervals at different locations. In agreement with the SR40 project team we decided for a real-world scenario, which was on one side simple enough to demonstrate the iterative timetable improvement sketched in the use case description of section 3.2, and on the other hand representing a realistic network scenario, with operational and commercial restrictions in the SI configuration. We also demonstrate the approach of the network segmentation described in section 2.5 for this specific scenario.

4.2 Real-world network Kerenzerberg

In this section, we describe in detail a case study related to IP-UC2. For the corridor between Ziegelbrücke and Sargans, we illustrate in detail:

- How to implement the traffic plan with a capacity time band for some subnetwork (of the overall train network) only.
- How to generate a traffic plan with capacity time band under normal operations (i.e. without construction intervals)
- How to generate a traffic plan with capacity time band feasible for different construction intervals.

The third point is already an extension of IP-UC2. Here we try to fix a traffic plan that is feasible for different construction intervals, which are planned for different time windows during the planning horizon. The goal here would be to communicate only ‘one’ traffic plan to the customers instead of several varying plans.

4.2.1 Introduction

We start with the description of mesoscopic infrastructure and SI on our test sector. According to our application concept, the SI is the result of IP-UC0 and IP-UC1 (see chapter 2 and 3 for details) and is maintained in Viriato and the ZHAW planning tool. For our case study, we adapted the existing SI for the timetable of 2018 in such a way, that we can prove that we can handle the basic IP requirements with the proposed IP use cases and the included algorithms for computer-aided timetable generation.

4.2.2 Description of the infrastructure

The infrastructure between Ziegelbrücke and Sargans under normal operations is summarised in the following table. The infrastructure table is maintained in Viriato (see chapters 0 and 3).

Station/Track ID	Number of tracks	Minimum travel time (Tracks)
Ziegelbrücke (ZGB)	12	
ZGB-WN	2	1.7
Weesen (WN)	2	
WN-MH	2	2.8
Mühlehorn (MH)	2	
MH-TIEF	1	1.3
Tiefenwinkel (TIE)	2	
TIE-MG	2	1.0
Murg (MG)	2	
MG-UNT	2	1.9
Unterterzen (UNT)	2	
UNT-MOL	2	1.0
Mols (MOL)	2	
MOL-WAL	2	1.6
Walenstadt (WAL)	3	

WAL-FMS	2	1.8
Flums (FMS)	2	
FMS-MEL	2	3.3
Mels (MEL)	2	
MEL-SA	2	1.5
Sargans (SA)	4	

Table 19: Infrastructure data of the sector ZGB-SA.

In the first column, we describe the stations (e.g. ZGB) and tracks (e.g. ZGB-MH). We see in Table 19 that there are always two tracks available, except between Tiefenwinkel and Mühlehorn, where only one track is available. Minimum travel times are derived from technical restrictions of the tracks.

4.2.3 Network segmentation

In order to generate a traffic plan with capacity bands, we have to segment the railway network into the relevant perimeter, as explained in chapter 2.5. The SI in the next section is also adapted to the segmented network. We illustrate the network related to our case study Kerenzerberg in Figure 22.

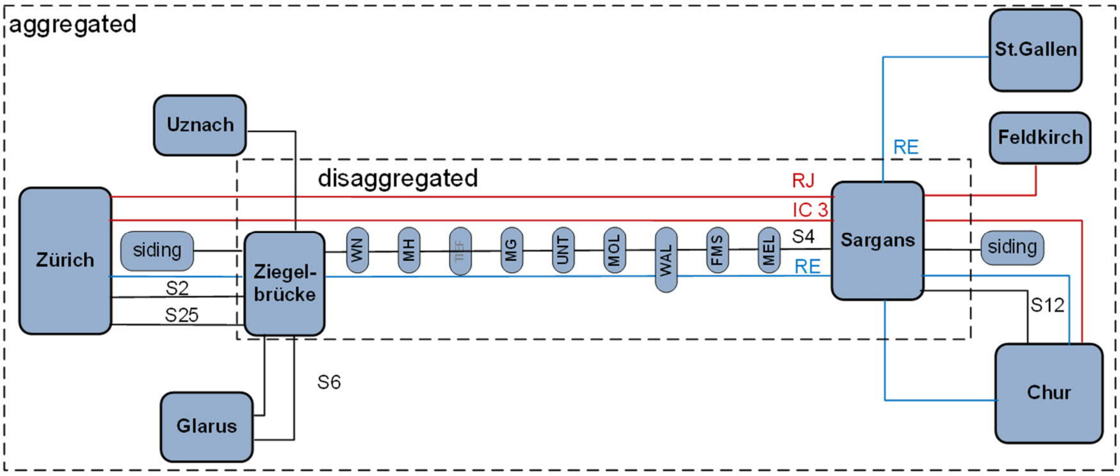


Figure 22: The network of the case study Kerenzerberg. In order to divide the relevant infrastructure for the IP timetabling scenario into a network partition with the relevant level of detail and a peripheral part with more coarse information, the railway network is divided into subnetworks. A disaggregated subnetwork contains the relevant infrastructure segments at the mesoscopic level, whereas an aggregated subnetwork represents the simplified infrastructure at the macroscopic level.

As the planned construction or maintenance work for our test scenario is located on the network section between Tiefenwinkel and Mels, we decided to use the corridor Ziegelbrücke-Sargans as the disaggregated partition of the test network, where we will generate a detailed traffic plan (see chapter 2.5). The western part of Ziegelbrücke is aggregated, i.e. we introduced the nodes Uznach, Zürich, Glarus and a siding of Ziegelbrücke and connecting tracks. The aggregated network will be used to maintain vehicle circulation (e.g. turnarounds) aspects of lines and to model connections to tangent lines (see the description of SI in the next section). The eastern part of Sargans is also aggregated. We introduced the nodes St.Gallen, Feldkirch, Chur and a siding of Sargans. In the aggregated network, we assume to have enough track capacity. Ziegelbrücke and Sargans can be considered as local hubs. At these stations the traffic plan has to account for passenger transfers between lines. Technically spoken, these transfer requirements result in connections constraints in our TCPESP-model.

4.2.4 Description of Service Intention

In chapters two and three, we explained that the SI is our main data structure and is maintained in Viriato and the planning tool ZHAW. The SI contains all the information needed to configure the Event Activity Network (EAN) and the Track Choice PESP model (TCPESP) respectively the Track Choice PESP model with flexibility (TCFPESP) (see chapters 2.4.2 and 2.4.3). We start with the lines considered. As mentioned before, our SI-lines represent an adaption of the lines in the corresponding timetable 2018. To demonstrate the turnaround operations, we decided that the line S4 makes a turnaround in a siding next to Ziegelbrücke and Sargans, respectively.

Stops / Lines	S4 $[D_{lo}, D^{up}]$ $[TT_{lo}, TT^{up}]$ $[TU_{lo}, TU^{up}]$	RJ	IC 3	RE 1	S12	S25	S6	RE 2	S 2
Glarus (GL)						[2, 58]	[2, 58]		
GL-ZGB						[0.5, 0.8]	[0.5, 0.8]		
St. Gallen (SG)								[2, 58]	
SA-SG								[0.5, 0.8]	
Zürich (ZUE)		[2, 58]	[2, 58]	[2, 58]		[2, 58]			[2, 58]
ZUE-ZGB		[0.5, 0.8]	[0.5, 0.8]	[0.5, 0.8]		[0.5, 0.8]			[0.5, 0.8]
Uznach (UZ)							[2, 58]		
UZ-ZGB							[0.5, 0.8]		
Siding (SZGB)	[2, 3.2]								
SZGB-ZGB	[0.5, 0.8]								
Ziegelbrücke (ZGB)	[2, 3]	[0, 1]	[0, 1]	[1, 1.5]		[2, 3]	[2, 3]		[2, 58]
ZGB-WN	[1.7, 2.6]	[1.7, 2.6]	[1.7, 2.6]	[1.7, 2.6]					
Weesen (WN)	[0, 1]	[0, 1]	[0, 1]	[0, 1]					
WN-MH	[2.8, 4.2]	[2.8, 4.2]	[2.8, 4.2]	[2.8, 4.2]					
Mühlehorn (MH)	[1, 1.5]	[0, 1]	[0, 1]	[0, 1]					
MH-TIE	[1.3, 2.00]	[1.3, 2]	[1.3, 2]	[1.3, 2]					
Tiefenwinkel (TIE)	[0, 1]	[0, 1]	[0, 1]	[0, 1]					
TIE-MG	[1, 1.5]	[1, 1.5]	[1, 1.5]	[1, 1.5]					
Murg (MG)	[0.5, 0.8]	[0, 1]	[0, 1]	[0, 1]					
MG-UNT	[1.9, 2.9]	[1.9, 2.9]	[1.9, 2.9]	[1.9, 2.9]					
Unterterzen (UNT)	[0.5, 0.8]	[0, 1]	[0, 1]	[0, 1]					
UNT-MOL	[1, 1.5]	[1, 1.5]	[1, 1.5]	[1, 1.5]					
Mols (MOL)	[0.5, 0.8]	[0, 1]	[0, 1]	[0, 1]					
MOL-WAL	[1.6, 2.4]	[1.6, 2.4]	[1.6, 2.4]	[1.6, 2.4]					
Walenstadt (WAL)	[0.5, 0.8]	[0, 1]	[0, 1]	[1, 1.5]					
WAL-FMS	[1.8, 2.7]	[1.8, 2.7]	[1.8, 2.7]	[1.8, 2.7]					
Flums (FMS)	[1, 1.5]	[0, 1]	[0, 1]	[0, 1]					
FMS-MEL	[3.3, 5]	[3.3, 5]	[3.3, 5.0]	[3.3, 5]					
Mels (MEL)	[0.5, 0.8]	[0, 1]	[0, 1]	[0, 1]					
MEL-SA	[1.5, 2.3]	[1.5, 2.3]	[1.5, 2.3]	[1.5, 2.3]					
Sargans (SA)	[2, 3]	[2, 3]	[1, 1.5]	[2, 3]	[2, 58]			[2, 3]	
SA-SSA	[0.5, 0.8]								
Siding SA (SSA)	[2, 3.2]								
SA-CH			[0.5, 0.8]		[0.5, 0.8]			[0.5, 0.8]	
Chur (CH)			[2, 58]	[2, 58]	[2, 58]			[2, 58]	
SA-FE		[0.5, 0.8]							
Feldkirch (FE)		[2,58]							

Table 20: The lines in the case study Kerenzerberg.

In Table 20, we summarised the upper and lower bound for dwell at every station ($[D_{lo}, D^{up}]$) and travel time for every track ($[TT_{lo}, TT^{up}]$). The routing can be derived from the entries in the table. A line visits all the stations and tracks from top to down and vice versa, where an upper and lower bound is given. Stations and tracks, which are not on the routing of a line, have no entry in the corresponding field. In the first and the last station, the lines perform a turnaround in the given interval ($[TU_{lo}, TU^{up}]$).

The minimum dwell D_{lo} and the minimum travel time TT_{lo} are given technical lower bounds. To compute the upper bounds D^{up} and TT^{up}, we multiplied the lower bounds with 1.5. This reserve will be used to derive flexible plans with the TCFPESP model.

The turnaround times are computed according to the approach of Liebchen and Möhring (2007). The turnaround intervals are computed such that a service with a minimal number of rolling stock gets possible. In our case study, line S4 is operating with one rolling stock. The other lines operate with more than one rolling stock due to longer round-trip times. These bounds are not computed according to Liebchen and Möhring (2007), they are set manually. These lines can cross themselves in opposite directions (as it is in the real-world timetable). Line Table 20 is mainly maintained in Viriato. Only the turnaround times are entered in the planning tool ZHAW.

The SI contains the following connections between the given lines:

Connection [C_{lo}, C^{up}] From/to at station	S4 (ZGB-SA)	IC 3 (ZGB-SA)	IC 3 (SA-ZGB)	RE1 (ZGB-SA)	RE 1 (SA-ZGB)	RJ (ZGB-SA)	RJ (SA-ZGB)	RE 2 (CH-SG)	S 12 (SA-CH)	S 25 (GL-ZUE)	S 25 (ZUE-GL)	S 2 (ZGB-ZUE)
S 4 (ZGB-SA)									[1,15] SA	[1,15] SA		[1,15] ZGB
S 4 (SA-ZGB)										[1,15] ZGB		
IC 3 (SA-ZGB)					[1,15] SA	[1,15] SA						
RE 1 (ZGB-SA)		[1,15] SA						[1,15] SA				
RE 1 (SA-ZGB)						[1,15] SA						
RE 2 (CH-SG)					[1, 15] SA	[1,15] SA						
RE 2 (SG-CH)			[1,15] SA		[1, 15] SA	[1,15] SA						
S 6 (GL- UZ)				[1,15] ZGB								
S 6 (UZ-GL)				[1,15] ZGB								
S 12 (CH-SA)			[1,15] SA		[1, 15] SA	[1,15] SA		[1,15] SA				
S 25 (ZUE-GL)	[1,15] ZGB											
S 2 (ZUE-ZGB)	[1,15] ZGB										[1,15] ZGB	

Table 21: Connections in the case study Kerenzerberg.

In Table 21, we find the implemented connections. The connections are the output of IP-UC0. The connections should take place in the time interval [C_{lo},C^{up}] from the line in the first column to line in the corresponding column, e.g. there should be a connection from the line S4 (direction ZGB-SA) to line IC 3 (direction ZGB-SA) in Sargans with a minimum and a maximum time, respectively, of 4 and 15 minutes, respectively. The connection Table 21 is maintained in the planning tool ZHAW.

Furthermore, the SI contains:

- A time separation of the lines S4 (ZGB-SA) and RE 1 (ZGB-SA) of [20, 40] minutes in Ziegelbrücke. This should guarantee a frequent service for passengers travelling from Ziegelbrücke to Sargans.

- Travel time restrictions for the lines S4, IC 3, RE 1 and RJ between Sargans and Ziegelbrücke, i.e. travel times should be between 17 and 21 minutes for the IC 3, RE 1 and RJ. Line S4 is restricted to be between 20 and 29 minutes.

The time separation and the travel time restrictions are an output of the line planning step IP-UC0.

4.2.5 Construction of traffic plan with capacity time band under normal operations

We describe in detail the process of generating a traffic plan with capacity time band. Based on the SI as defined above, the planning tool ZHAW generates the Event Activity Network (EAN) and the Track Choice PESP model with flexibility (TCFPESP) (see chapter 2.4.2 and 2.4.2).
To generate the traffic plan, we use the TCFPESP iteratively with different objective functions, namely:

- We minimise all passenger relevant times (i.e. trip, dwell and connections times). We will call the model in this case, according to Caimi et al. (2011b) MINTRAVEL.
- We maximise the flexibility in a certain range at all arrival and departure events at stations. We add a constraint on the passenger travel time. The passenger travel time has to be smaller than (1+p) times the best possible travel time from the model MINTRAVEL. We will call the model in this case, according to Caimi et al. (2011b) CONTRAVEL. Parameter p is controlling the quality of the schedule for the passengers' travel times.

By using the models MINTRAVEL and CONTRAVEL iteratively, we can generate a traffic plan covering stability and travelling time aspects.

Iteration scheme 1: Generate traffic plan with capacity time band under normal operations

Input: From IP-UC0, IP-UC3 and IP-UC4 we get the

- Size of flexibility for all arrival and departure nodes necessary for stability (in this context also refer to section
- Parameter p for controlling deviations of passenger travel times
- Bound on rolling stock per line

(at the beginning the size of flexibility and p are set to standard values)

1. Solve the model MINTRAVEL. We get a traffic plan with the best possible travel times TT_{best}.
2. Compare the necessary amount of rolling stock for the traffic plan with the given bound on rolling stock. If not feasible, go to IP-UC0 and adjust SI.
3. Solve the model CONTRAVEL, allow passenger travel times to be maximal (1+p) $TT_{best.}$ We get a traffic plan with capacity time band.
4. Compare the necessary amount of rolling stock for traffic plan with given bound on rolling stock. If not feasible, go to IP-UC0 and adjust SI.
5. Release traffic plan to IP-UC3 and IP-UC4.

The iteration scheme 1 above is only related to IP-UC2. It is embedded in the overall iterations of the use cases in chapter 3. In our case study, we show the results of the iteration scheme 1 related to IP-UC2.
In the case study Kerenzerberg, we have set the maximal flexibility to 10 seconds and p to 0.5. These values are based on the experience of planning experts. We get the following traffic plan with capacity time bands and resulting track allocation.

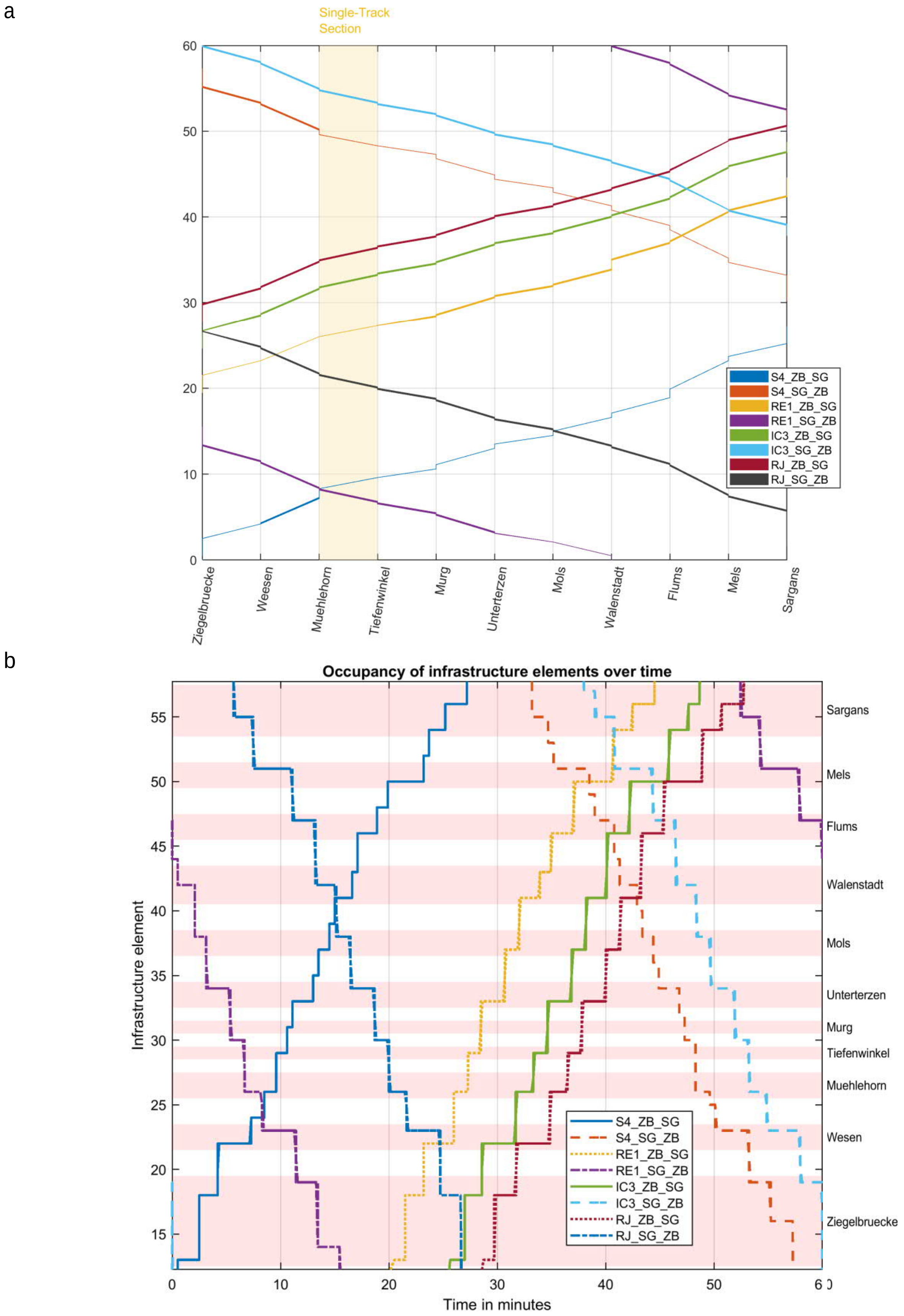


Figure 23: Traffic plan with capacity time band (a) and track assignment diagram (b) under normal operations.

In Figure 23, we see the traffic plan and the resulting track allocation. It can be verified that the SI is fulfilled in general. Especially we can see that

- line S4 operates with one rolling stock as requested;
- the service of S4 and RE 1 is separated in Ziegelbrücke to guarantee smooth services to Sargans;
- the track choice method TCFPESP is able to generate a feasible track allocation on the mesoscopic infrastructure.

Interesting for the next section is the fact that this traffic plan has crossings between Flums and Mels. It will not be feasible for the considered construction intervals.

4.2.6 Construction of traffic plan with capacity time band under several construction intervals

In this section, we want to demonstrate how to use the iteration scheme 1 from IP-UC2 above to generate a feasible traffic plan with capacity time band for two construction intervals. The construction sites are Tiefenwinkel to Unterterzen and Flums to Mels. The construction intervals take place during our planning horizon but in different time windows. Only one track is available during the construction intervals on the affected corridors.

To be more precise, we compute two traffic plans feasible for the respective construction interval, but the plans are so close to each other, that it is possible to communicate only one ‘commercial’ traffic plan to the customers. Of course, this is positive from a customer perspective, but also for the operator, since the free capacity in the network can be used for additional services (e.g. freight trains) during the whole planning horizon. The idea of this approach was introduced already to the SR40 project team by Laumanns (2017).

Iteration scheme 2: Generate traffic plan with capacity time band under several construction intervals

Input: From IP-UC0, IP-UC3 and IP-UC4 we get the

- Size of flexibility for all arrival and departure nodes necessary for stability
- Parameter p for controlling deviations of passenger travel times
- Bound on rolling stock per line
- Infrastructure restrictions for all n construction intervals
- Tolerance between traffic plans with capacity time band of the single construction intervals.

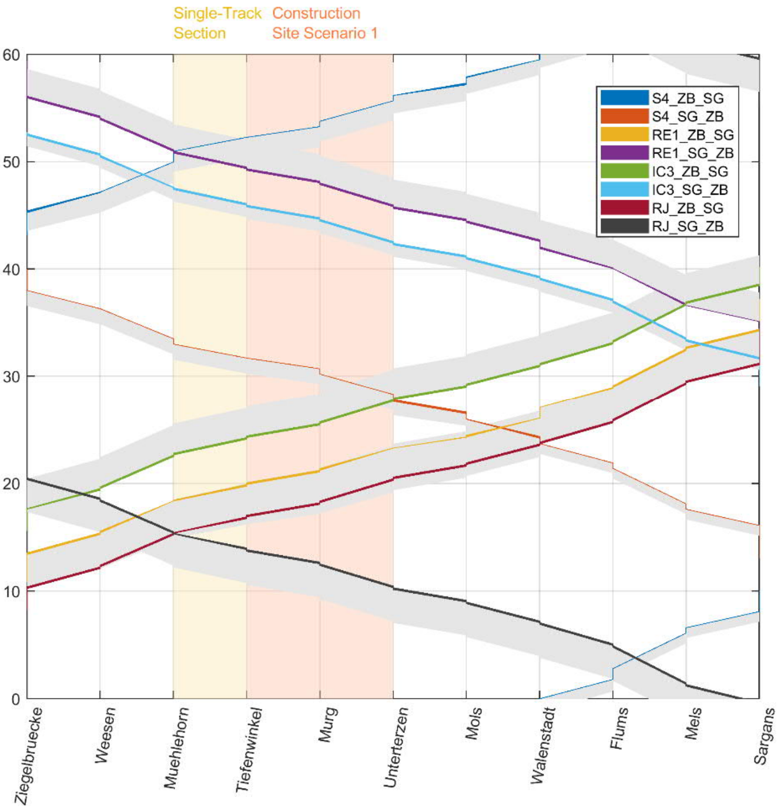
(The first three points are taken from the traffic plan under normal operations)

1. Try to generate a traffic plan with capacity time band feasible for all n construction intervals (i.e. feasible for all resource restrictions) with the help iteration scheme 1. If successful, stop the iteration.
2. Start with a first construction interval: Compute a traffic plan with capacity time band for this construction interval with the help iteration scheme 1. If not feasible, go to IP-UC0.
3. For each line take the passing times at the beginning and the end of the disaggregated network (i.e. Sargans and Ziegelbrücke in our case study) and add them to the SI with the expected tolerance from the input. The remaining construction intervals will be computed with this adapted SI.
4. Compute the traffic plans for all single construction intervals with the new SI from step 3 and with the help iteration scheme 1. If one is not feasible, go to IP-UC0.
5. We constructed a traffic plan with capacity time band for each construction interval. At every station and for every line we communicate the earliest departure and the latest arrival (concerning all construction intervals) as ‘commercial’ traffic plan to the customers.

In the case study Kerenzerberg, we have two construction intervals. Step 1 was not successful, i.e. we did not find a feasible traffic plan for both construction intervals. We then started with a construction interval between Flums and Mels in step 2, which lead to a feasible solution, given the SI under normal operations. In step 3, we took the passing times from all the lines in Ziegelbrücke and Sargans. For the second construction interval we allowed the lines to pass +/- 3 minutes with respect to passing times given from construction interval 1. We admitted a tolerance of 6 minutes. The second construction interval between Tiefenwinkel and Unterterzen was also feasible with the adapted SI. We get a 'commercial' traffic plan.

In Figure 24, we see the traffic plan with capacity time band for both construction intervals. Due to iteration scheme 2, the traffic plan for the lines is at the lower or the upper boundary of the grey band. The grey band is the 'commercial' traffic plan. The traffic plan for construction interval 1 (Figure 24a) is not feasible for construction interval 2 (Figure 24b) and vice versa, e.g. line RJ and line S4 have a crossing between Flums and Mels during construction interval 1. It is interesting to mention that the order of line RJ and line RE1 from Ziegelbrücke to Sargans change from construction interval 1 to 2.

a



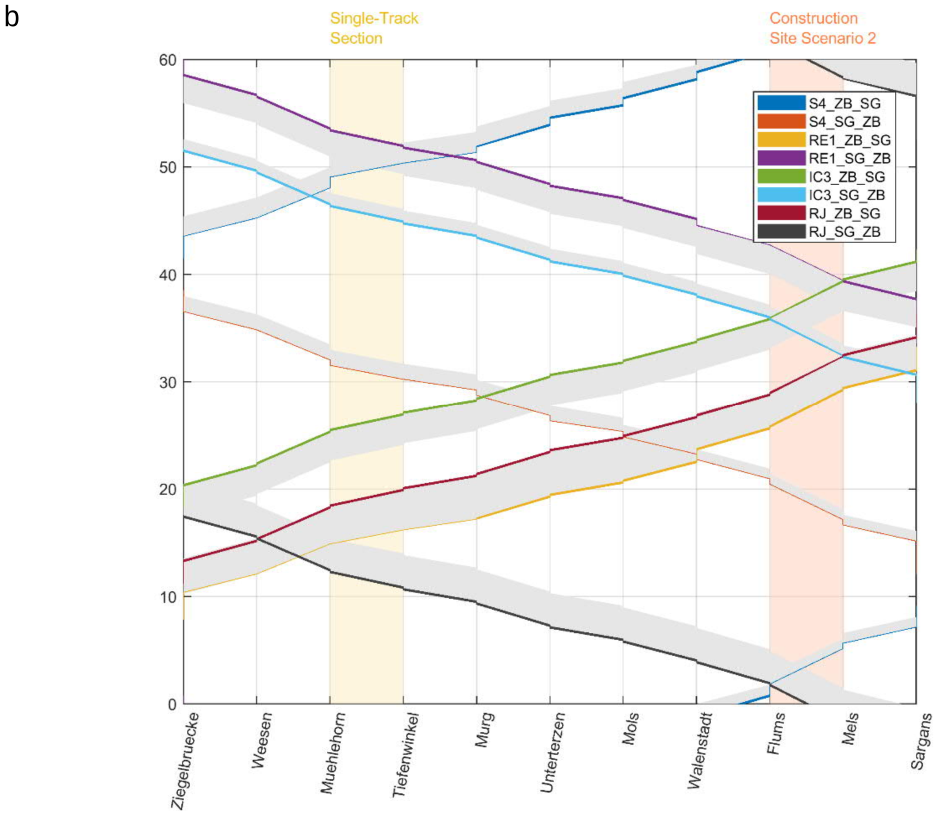


Figure 24: a) Traffic plan with capacity time bands for construction interval 1, b) Traffic plan with capacity time bands for construction interval 2.

In Figure 25 on the left (a), we see the ‘commercial’ traffic plan for the construction intervals. The departure times correspond to the lower boundary of the grey band and the arrival times to the upper boundary. During the planning horizon, we therefore, always find a feasible traffic plan for all construction intervals. On the right (b), we see a detailed view of the traffic plan of the line S4 between Unterterzen and Walenstadt. The blue bands represent the capacity time bands, e.g. during construction interval 1 we have around 10 seconds flexibility for the arrival and the departure in Mels. We generated a flexible traffic plan, which fulfils the SI. Therefore, we were able to integrate stability and passenger travel time aspects.

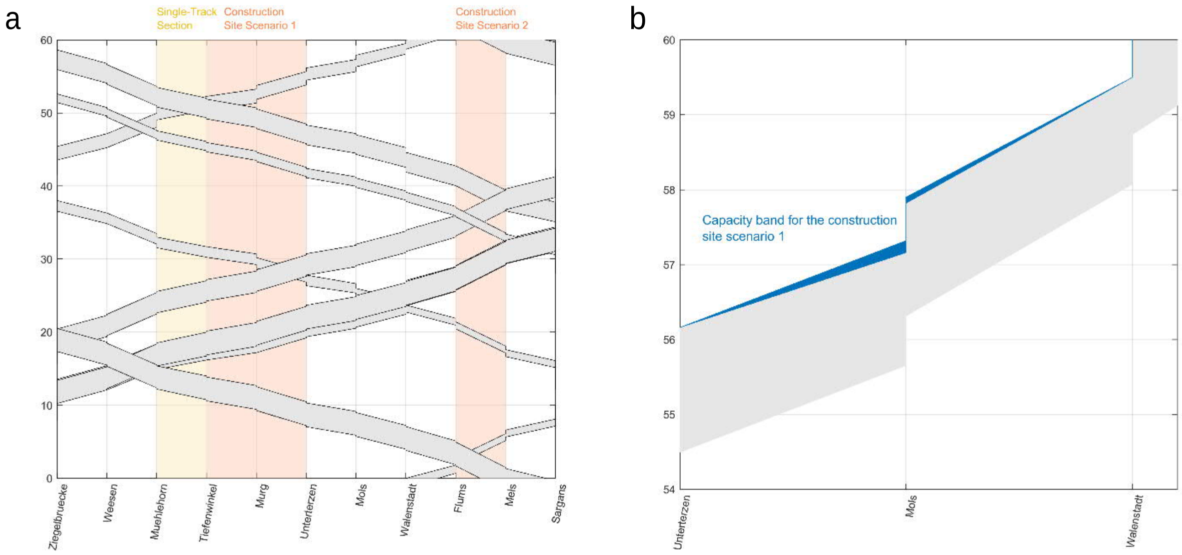


Figure 25: a) ‘Commercial‘ traffic plan, b) Capacity time band of line S4.

5 Discussion and outlook

5.1 Summary

The main goal of this project was to develop a concept for computer-aided timetable development, which refers to research questions RQ1 to RQ5 introduced in section 1.2.

Referring to existing evidence from academic research regarding

- an automated timetable creation (RQ1),
- the existing SR40 process model for timetable generation (RQ5) and
- the practical requirements for the 'progressive' feasibility assessment of algorithmically generated traffic plans (RQ3),

in this report, we introduce an SI-based application concept for interval planning (RQ2 and RQ5) that consists of 6 use cases IP-UC0 to IP-UC5.

We provide a detailed description of use cases IP-UC1 to IP-UC4. However, the elaboration of IP-UC0 and IP-UC5 were not part of this project.

In the use cases IP-UC1 "Consistency check of SI" and IP-UC2 "Traffic plan with capacity time band" we made use of an integrated timetabling model. The model in IP-UC1 is based on the well-known PESP method. The model we propose for IP-UC2 is based on a new version of the model that we call TCPESP (Track Choice Periodic Event Scheduling-Problem) and TCFPESP (Track Choice Periodic Event Scheduling-Problem with event flexibility), respectively. It is an extension of the PESP model and can be used to support the timetable planner for generating train and vehicle circulation schedules with track assignment. The TCFPESP model enables the required feasibility assessment of the resulting traffic plan as resource assignments are made on the mesoscopic level, which allows to solve the timetabling problem for practical problem sizes and keeps configuration efforts at a low level.

Moreover, feasible timetabling results computed with the TCFPESP procedure can be evaluated subsequently with the Max-Plus Performance Analyser (MPPA), which we use to analyse timetable stability and buffer requirements. In addition to the investigation of the stability of the timetable, we also assess the customer convenience by calculating the overall travel time resulting from traffic plans and subsequently compare the results of different plans. For this purpose, we introduced the so-called Service Intention Index SII (for details see Appendix F). With the SII of all relevant plans, we can provide the planner with a rating of different planning versions and hence some additional information as decision support. In this way, we expect to improve the quality of TCFPESP results (RQ4) and to make a relevant contribution for speeding up and facilitating the daily work of railway timetable planners.

In our opinion, an important business requirement is that there exists a close cooperation between TOC and IOP based on tightly coupled data exchanges and agreed service levels. In order to provide optimal transport services from a customer point of view (in terms of end to end transport times) as well as from an operational point of view (in terms of operational stability and cost efficiency) the process owner IOP needs extended data access (e.g. to line operation costs, origin-destination customer demand, product attributes and priorities etc). We describe the required data in detail in our interface specification in section 2.2 and the use case description in chapter 3.

The methods and model properties that we describe in chapter 0 and the use cases in chapter 3 are tested in a small-scale test and a real-world case study for IP in chapter 4. In the small-scale test case, we show, how the stepwise improvement of the traffic plan is achieved by an iterative execution of planning use case IP-UC2 (Traffic plan with capacity time band) and performance measurement use cases IP-UC3 (Stability of traffic plan) and IP-UC4 (SII of traffic plan). With this small-scale test case, we also point out how the planner can utilise customer oriented and operator-oriented performance measures in order to compare the timetabling result of different versions of relaxed SI configurations with each other.

In the real-world test case, we show how the concept of SI can be used to develop a customer timetable, which is valid during the complete timetable period but at the same time makes it possible that two different construction or maintenance intervals with different locations can be planned during this timetable period. This is of considerable practical relevance, especially with respect to the increasing number of intervals to be planned and executed under conditions of continued production of railway-services.

5.2 Outlook and future research

If timetabling requirements turn out to be infeasible to be solved by TCPESP because, for instance, the given SI is not realisable on the respective railway infrastructure (a typical situation during construction intervals), this situation has to be solved by relaxing the SI. This is described in the overview of IP use cases in section 3.1, showing that in this case, IP-UC0 has to be executed. In a next step which is part of the SR40 project supplement of this research, we will show, how the SI is generated using a standard line planning methods similar to those described by, e.g. Friedrich et al. (2017). Our preliminary investigations show that these methods can generate SI configurations that take reduced resource availability (for instance, because tracks are temporary out of service) into consideration. This part of our application concept also will allow deeper insights into the required data access for the IOC. This research will help to make detailed specifications of data interfaces and service levels between TOC and IOC in case of IP and operational disruptions in real-time conditions.

Another aim of future research concerns the method for the utilisation of timetable stability measures, such as CDI and CDS, obtained from the Max-Plus-Framework for assigning event flexibility in order to improve timetable robustness (see the detailed discussion in section 4.1.3.3). With this research step, we can provide a detailed description of the use case covering the iteration between use cases IP-UC3 and IP-UC2.

From our point of view, the results of the test our framework are encouraging enough to be elaborated in more detail and to form a solid basis for a functional and technical requirement catalogue in the SR40 project.

5.3 Publication of project results

The research described in this report has led to the following publications (in chronological order):

- Wüst, R.M, Bütikofer, St., Ess, S., Gomez, C., Steiner, A., Laumanns, M. and Szabo, J.: Periodic timetabling with 'Track Choice'-PESP based on given line concepts and mesoscopic infrastructure. Paper presented at the OR 2018 conference, 11 to 14 September 2018, Brussels, Belgium (2018)
- Bütikofer, St., Köchli, J., Frick, K. und Weber, Ch.: Automatisierte Linienplanung im öffentlichen Verkehr, Eisenbahntechnische Rundschau ETR, 4/19 (2019)
- Wüst, R.M, Bütikofer, St., Ess, S., Gomez, C., Steiner, A., Laumanns, M. and Szabo, J.: Improvement of maintenance timetable stability based on iteratively assigning event flexibility in FPESP. Paper presented at the RailNorrköping 2019 conference, 17 to 20 June 2019, Norrköping, Sweden (2019)
- Wüst, R.M, Bütikofer, St., Ess, S., Gomez, C., Steiner, A., Laumanns, M. and Szabo, J.: Maintenance timetable planning based on mesoscopic infrastructure and the transport service intention. Manuscript submitted to the Journal of Rail Transport Planning & Management (2019) (currently under revision)
- Wüst, R.M, Bütikofer, St., Köchli, J. and Ess, S.: Generation of the transport service offer with application to timetable planning considering constraints due to maintenance work. Manuscript to be submitted, 2nd International Railway Symposium Aachen (IRSA) 2019, 26 to 28 November 2019, Aachen, Germany (2019)

We assume that this research will lead to further conference contributions and/or publications in scientific journals.

List of abbreviations

Abbreviation	Description
BAV	Bundesamt für Verkehr
BFS	Bundesamt für Statistik
CH	Chur
CDI	Cumulative delay impact
CDS	Cumulative delay sensitivity
CONTRAVEL	Constrained Travel Time (Name of objective function corresponding to Caimi et al. (2011b)
DES	Discrete Event System
FE	Feldkirch
FPESP	Flexible Periodic Event Scheduling-Problem
FMS	Flums
GL	Glarus
IDP	Institute of Data Analysis and Process Design
IOC	Infrastructure operating company
IFIT	Integrated Fixed-Interval Timetable
IP	Interval planning
IP-UC	Interval planning use case
KPI	Key Performance Indicator
MH	Mühlehorn
ME	Mels
MP	Max-Plus
MPA	Max-Plus Algebra
MPPA	Max-Plus Performance Analyser
OD	Origin-Destination (used in the context of door-to-door transport demand)
OTP	Open Trip Planner
PESP	Periodic Event Scheduling-Problem
PETER	Performance Evaluation of Timed Events in Railways
POC	Proof of Concept
PTN	Public Transport Network
SA	Sargans
SG	St. Gallen
SI	Service Intention
SII	Service Intention Index
SR40	Smart Rail 4.0
SZGB	Siding Ziegelbrücke
TCPESP	Track Choice Periodic Event Scheduling-Problem
TCFPESP	Track Choice Periodic Event Scheduling-Problem with event flexibility

TIEF	Tiefenwinkel
TMS	Traffic Management System
TOC	Train operating company
TraMP	IBM Train Movement Planner
UNT	Unterterzen
MG	Murg
MH	Mühlehorn
MINTRAVEL	Minimum Travel Time (Name of objective function corresponding to Caimi et al. (2011b)
MOL	Mols
SA	Sargans
SBB	Schweizerische Bundesbahnen (Swiss Federal Railways)
SSA	Siding Sargans
TOC	Train operating company
UC	Use Case
UNT	Unterterzen
UZ	Uznach
WA	Walenstadt
WN	Weesen
ZUE	Zürich Hauptbahnhof
ZGB	Ziegelbrücke
ZHAW	Zürcher Hochschule für Angewandte Wissenschaften

References

Baccelli, F.L., Cohen, G., Olsder, G.J., Quadrat, J.P.: Synchronization and Linearity: An Algebra for Discrete Event Systems. Wiley, Chichester (1993)

BAV: Taktvoll durch die Schweiz. www.deutschland-takt.de. Presentation of Hauke Fehlberg, Sektionschef Planung (2011)

BFS: Umweltindikator - Modalsplit im Verkehr. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/raum-umwelt/ressourcen/umweltindikatorensystem/reaktionen-der-gesellschaft/modalsplit-verkehr.html (2018) (last access: 28.9.2018)

Bešinović, N., Goverde, R.M.P., Quaglietta, E., Roberti, R.: An integrated micro-macro approach to robust railway timetabling. Transportation Research Part B: Methodological, vol. 87, pp. 14-32 (2016)

Braker, J.G., Olsder, G.J.: The Power Algorithm in Max Algebra. Linear Algebra and its Applications, vol. 182, pp. 67–89 (1993)

Büker, T.: Ausgewählte Aspekte der Verspätungsfortpflanzung in Netzen. PhD Thesis, RWTH Aachen University (2010)

Büker, T., Seybold, B.: Stochastic modelling of delay propagation in large networks. Journal of Rail Transport Planning & Management, vol. 2, pp. 34–50 (2012)

Caimi, G.: Algorithmic decision support for train scheduling in a large and highly utilised railway network. Diss. ETH Zürich Nr. 18581, ETH Zürich, Zürich, Switzerland (2009)

Caimi, G., Laumanns, M., Schüpbach, K., Wörner, S., Fuchsberger, M.: The Periodic Service Intention as a Conceptual Frame for Generating Timetables with Partial Periodicity. Transportation Planning and Technology, vol. 34, issue 4, pp. 323-339 (2011a)

Caimi, G., Fuchsberger, M., Laumanns, M., Schüpbach, K.: Periodic railway timetabling with event flexibility. Networks, vol. 57, no. 1, pp. 3-18 (2011b)

Caimi, G., Kroon, L., Liebchen, Ch.: Models for railway timetable optimization: Applicability and applications in practice. Journal of Rail Transport Planning & Management 6, pp. 285-312 (2017)

Cochet-Terrasson, J., Cohen, G., Gaubert, S., Mc Gettrick, M., Quadrat, J.-P.: Numerical Computation of Spectral Elements in Max-Plus Algebra. In: IFAC Conference on System Structure and Control, Nantes, France, 1998 (1998)

De Fabris, S., Longo, G., Medeossi, G., Pesenti, R.: Automatic generation of railway timetables based on a mesoscopic infrastructure model. Journal of Rail Transport Planning & Management, vol. 4, pp. 2–13 (2014)

Desaulniers, G., Hickman M.: Public transit. In: Laporte, G., Barnhart, C. (eds.), Transportation. Handbooks in Operations Research and Management Science, vol. 14, pp. 69-127 (2007)

Franke, B.: railML.org, OnTime. URL: https://www.railml.org/en/introduction/applications/detail/ ontime.html (2018) (last accessed: 10 September 2018)

Friedrich, M., Hartl, M., Schiewe, A., Schöbel, A.: Integrating passengers' assignment in cost-optimal line planning. Technical Report 2017-5, Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Georg-August Universität Göttingen (2017). URL: http://num.math.uni-goettingen.de/ preprints/files/2017-5.pdf (last accessed: 19 September 2018)

GAMS: General Algebraic Modeling System GAMS, GAMS Software GmbH, Frechen, Germany (2018), URL: https://www.gams.com/ (last accessed: 19 September 2018)

Goverde, R.M.P.: Punctuality of Railway Operations and Timetable Stability Analysis. TRAIL Thesis Series no. T2005/10. Delft University of Technology, Delft, The Netherlands (2005)

Goverde, R.M.P.: Railway timetable stability analysis using max-plus system theory. Transportation Research Part B, vol. 41, no. 2, pp. 179-201 (2007)

Goverde, R.M.P.: A Delay Propagation Algorithm for Large-Scale Railway Traffic Networks. Transportation Research Part C, vol. 8, no. 3, pp. 269–287 (2010)

Goverde, R.M.P., Heidergott, B., Glenn, M.: A coupling approach to estimating the Lyapunov exponent of stochastic max-plus linear systems. European Journal of Operational Research, vol. 210, no. 2, pp. 249–257 (2011).

Hansen, I.A., Pachl., J. (eds.): Railway Timetable & Traffic. Analysis – Modelling – Simulation. Eurail Press, Hamburg, Germany (2008)

Heidergott, B., Olsder, G.J., van der Woude, J.: Max-Plus at Work. Modelling and Analysis of Sychronized Systems: A Course on Max-Plus Algebra and its Applications. Princeton University Press, Princeton, New Jersey, USA (2006)

Herrigel, S.: Algorithmic decision support for the construction of periodic railway timetables, Diss. ETH Zürich Nr. 22548, Department Bau, Umwelt und Geomatik, ETH Zürich, Zürich, Switzerland (2015)

Howald, P., Künzi, Th., Wild, P., Wieland, Th.: Grobkonzept «Linienplanung» Version 0.7, 19.10.2017. SBB Smart Rail 4.0 TMS-PAS, Document repository (2017)

Hürlimann, D.: Objektorientierte Modellierung von Infrastrukturelementen und Betriebsvorgängen im Eisenbahnwesen. Diss. ETH Zürich Nr. 14281 (2001)

Janecek, D.; Weymann, F., Schaer, Th.: "LUKS – Integriertes Werkzeug zur Leistungsuntersuchung von Eisenbahnknoten und -strecken", ETR, vol. 59, no. 1-2, pp. 25-32 (2010)

Karp, R.M.: A Characterization of the Minimum Cycle Mean in a Digraph. Discrete Mathematics, vol. 23, no. 3, pp. 309–311 (1978)

Laumanns, M. (2017). Modellierung der Planungsprozesse: Abschlussbericht Forschungsauftrag im Rahmen PPS2030 und TMS PAS, 31.8.2017 (2017)

Liebchen, C., Möhring, R.H.: The modeling power of the periodic event scheduling problem: Railway timetables – and beyond. In: Geraets, F., Kroon, L., Schöbel, A., Wagner, D., Zaroliagis, C. (eds.), Algorithmic Methods for Railway Optimization, Lecture Notes in Computer Science, vol. 4359, Springer, Berlin/Heidelberg, Germany, pp. 3–40 (2007)

Lusby, R.M., Larsen, J., Ehrgott, M., Ryan, D.: Railway track allocation: models and methods. OR spectrum, vol. 33, no. 4, pp. 843-883 (2011)

Middelkoop, A.D. and Bouwman, M.: SIMONE: Large scale train network simulations. In: Peters, B., Smith, J., Medeiros, D., Rohrer, M. (eds), Proceedings of the 2001 Winter Simulation Conference, pp. 1042–1047 (2001)

Noordeen, M.: Stability analysis of cyclic timetables for a highly interconnected rail network. PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland (1995)

OTP: Open Trip Planner – Multimodal Trip Planning, see: http://www.opentripplanner.org/ (last accessed: 19 September 2018) (2018)

Projektteam «PPS2030»: Ergebnisdokument. Zielarchitektur «PPS2030» (Business & IT), SBB Dokument Library, Version 0.6, 25.4.2017 (2017)

Robenek, T., Azadeh, S.S., Maknoon Y., Bierlaire, M.: Hybrid Cyclicity: Combining The Benefits of Cyclic and Non-Cyclic Timetables. Report TRANSP-OR 160510, Transport and Mobility Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland (2016)

Schöbel A.: Line planning in public transportation: models and methods, OR Spectrum, vol. 34, pp. 491-510, (2012)

Siefer, T., Radtke, A.: Railway Simulation: Key for Better Operation and Optimal Use of Infrastructure. In: Hansen, I.A., Dekking, F.M., Goverde, R.M.P., Hindergott, B., Meester, L.E. (eds.), Proceedings of the 1st International Seminar on Railway Operations Modelling and Analysis, International Association of Railway Operations Research, Delft, The Netherlands (2005)

Sinner, M., Weidmann, U.: Automatisierte Bahnproduktion. Identifikation der funktionalen Anforderungen an die Datengrundlagen und Datenflüsse im Hinblick auf TMS, Schlussbericht zum Forschungsprojekts ‚Automatisierte Bahnproduktion', durchgeführt vom IVT-VS (ETHZ), im Auftrag der Schweizerischen Bundesbahnen SBB (2016)

SMA: Viriato: software for railways. Info Folder: URL: https://www.railsystemsengineering.com/RSE/documents/SMA%20Viriato%20Brochure.pdf (2018) (last accessed: 7 September 2018)

SMA: Viriato - software for railways. User Manual (German): Viriato 8.9 Standard, Benutzerhandbuch (2016)

Subiono and Van der Woude, J.: Power Algorithms for (max,+)- and Bipartite (min,max,+)-Systems. Discrete Event Dynamic Systems, vol. 10, no. 4, pp. 369–389 (2000)

Tabak, M. M.: Timetable analysis with PETER, Master Thesis of Science in Applied Mathematics, Delft University of Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft, The Netherlands (2008)

Toletti, A., Weidmann, U.: Modelling customer inconvenience in train rescheduling. Proceedings of the 16th Swiss Transport Research Conference (STRC), Ascona, Switzerland (2016)

Wüst, R., Steiner, A., Gomez. C., Goecer, K.: Automatisierte Dispositionslösungen bei eingeschränkter Benutzbarkeit von Infrastrukturelementen, Projektbericht im Auftrag von SBB Infrastruktur, Betrieb, 20.3.2017 (2017)

Wüst, R., Bütikofer, S., Ess, Gomez, C., Steiner, A., Laumanns, M., Szabo, J.: Periodic timetabling with 'Track Choice'-PESP based on given line concepts and mesoscopic infrastructure. Operations Research 2018, Bruxelles, Conference presentation, 13.9.2018 (2018a)

Wüst, R., Bütikofer, S., Ess, s., Gomez, C., Steiner, A., Laumanns, M., Szabo, J.: Periodic timetabling with 'Track Choice'-PESP based on given line concepts and mesoscopic infrastructure, Operations Research 2018, Bruxelles, Conference proceedings (submitted: 12.7.2018) (2018b)

ZHAW-IDP: Proposal for a project advertised by the SBB Fund for Research into Management in the Field of Transport: Development of a prototype for the automated generation of timetable scenarios specified by the transport service intention (2017)

Appendix

A	Impact-Interest matrix of stakeholders
B	Types of network topologies to describe the system at different levels of detail
C	SBB SmartRail 4.0 process and system model: An overview
D	Methodological aspects of the Max-Plus Algebra
E	Computation of stability and performance measures
F	Computation of Service Intention Indices
G	Computation of the Cumulative Delay Impact and Cumulative Delay Sensitivity
H	Conference paper, OR 2018 conference, 11 to 14 September 2018, Brussels, Belgium.
I	Conference paper, RailNorrköping 2019 conference, 17 to 20 June 2019, Norrköping, Sweden.
J	Journal paper, submitted to the Journal of Rail Transport Planning & Management (under revision)
K	Conference paper, 2nd International Railway Symposium Aachen (IRSA), 26 to 28 November 2019, Aachen, Germany (to be submitted)

A Impact-Interest matrix of stakeholders

Figure copied from the document «SBB-Projekt SmartRail 4.0 TMS-PAS Team AFO, Grobkonzept «Linienplanung» Version 0.7 vom 19.10.2017».

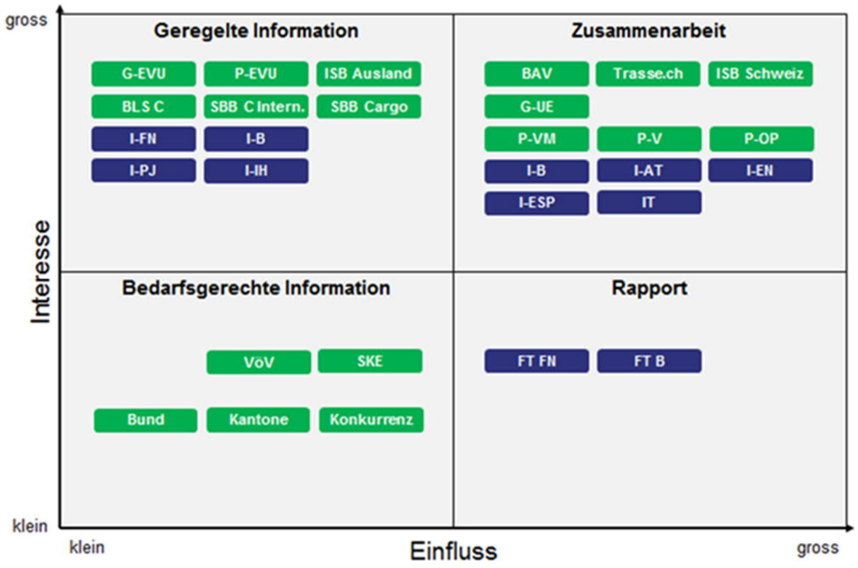


Figure 26: Influence-Interest-Matrix of affected stakeholders: Explanations to stakeholder groups and the corresponding interests are given in the Smart Rail 4.0 project requirements document: SR40_PaMa_Stakeholdermanagement_v3.0

B Types of network topologies to describe the system at different levels of detail

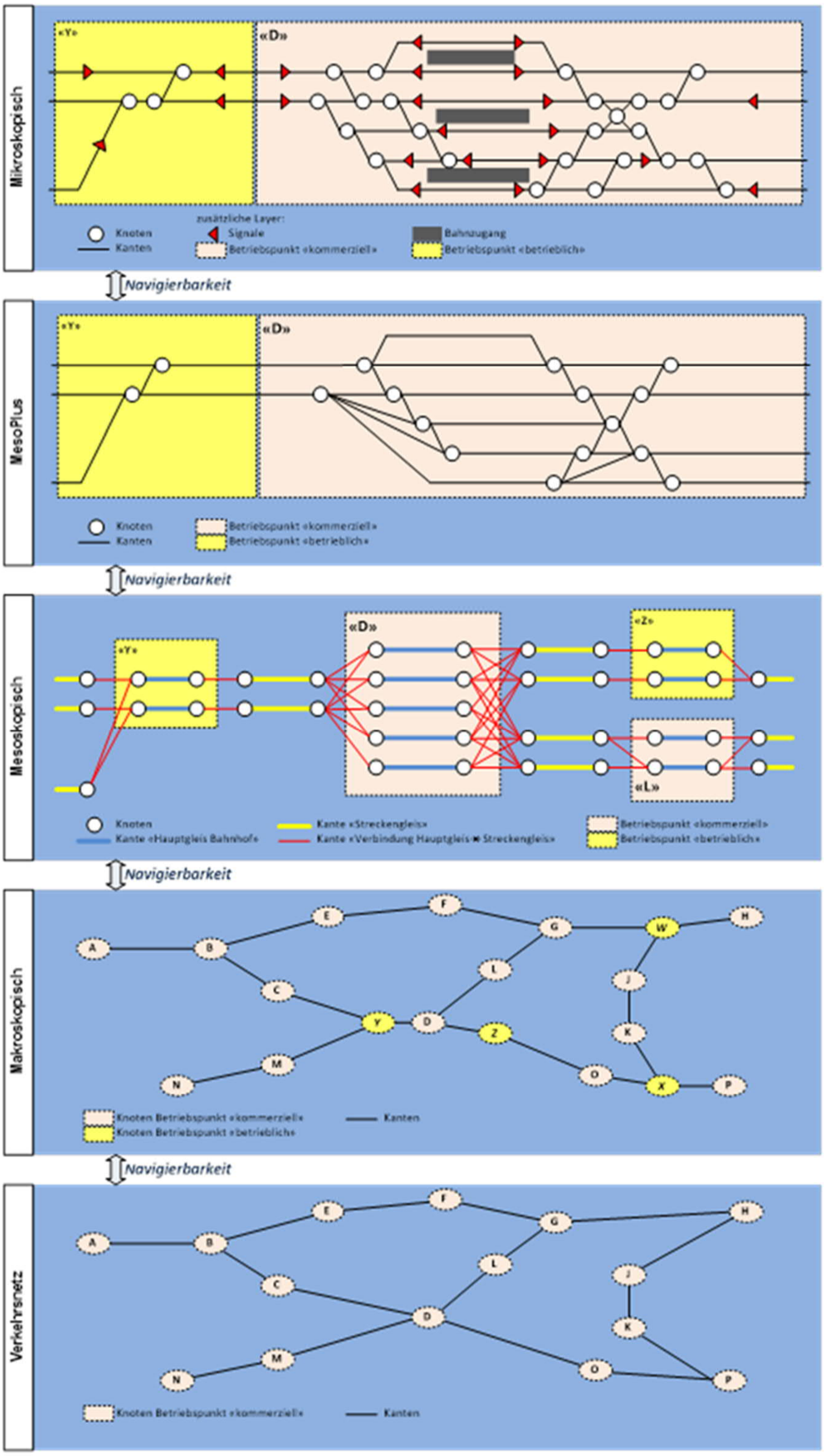
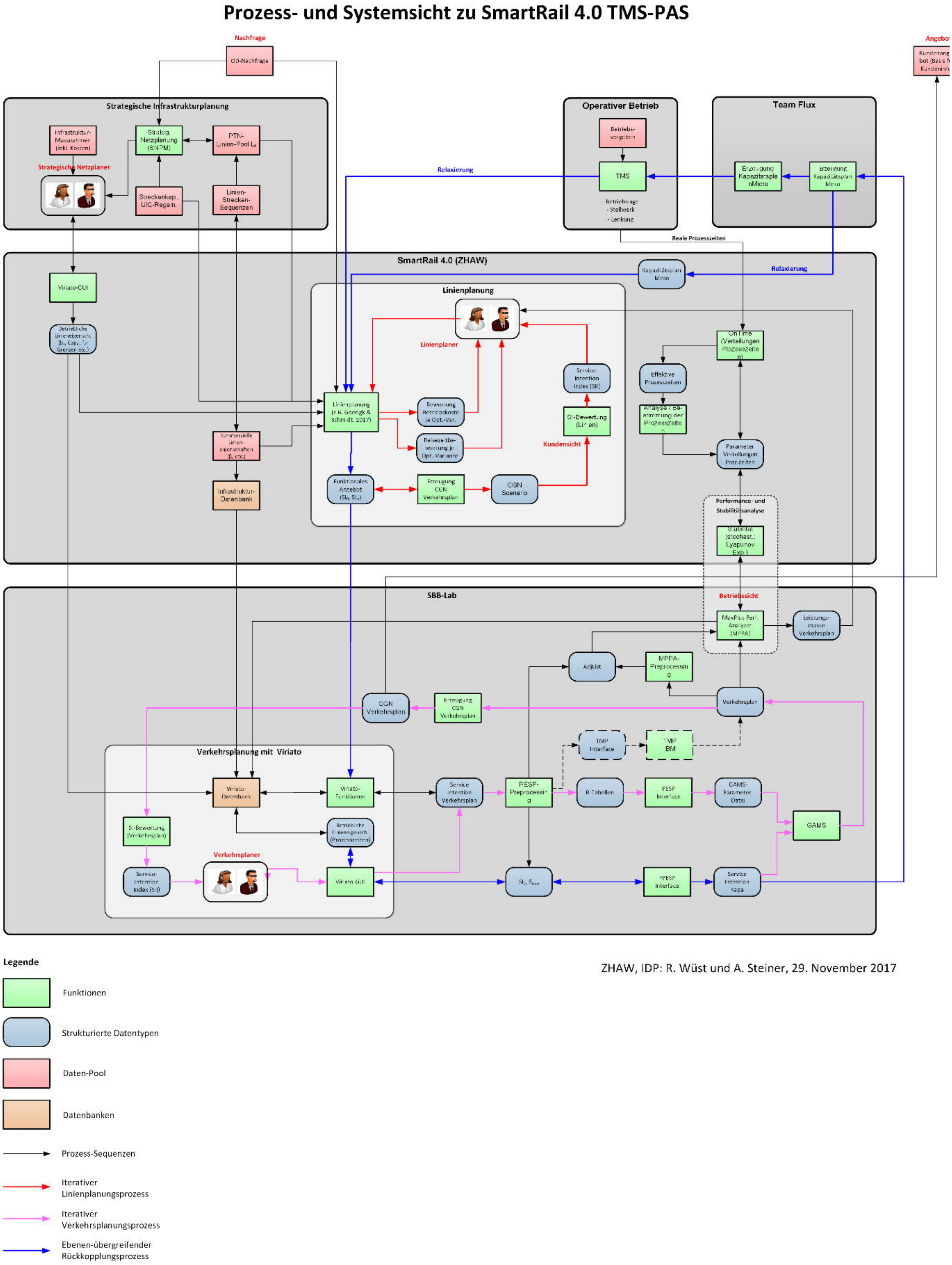


Figure 27: Topology. The topology is a classification system that makes the physical and logical elements of the railway network navigable as a logical graph.

C SBB SmartRail 4.0 process and system model: An overview



ZHAW, IDP: R. Wüst und A. Steiner, 29. November 2017

Figure 28: Overview system and information flow for all components of the current project (SBB-Lab) and the SR40 project supplement “ZHAW” (in German).

D Methodological aspects of the Max-Plus Algebra

For the mathematical description of time-controlled discrete event systems (DES) on a mesoscopic level, which also includes railway systems, the so-called Max Plus Algebra (MPA) has established itself over the last around fifteen years (see for example Goverde (2005), Heidergott et al. (2006), Hansen and Pachl (2008)). With this framework, this class of systems can be formally described elegantly, and at the same time, it offers a multitude of methods for system analysis. In this project, the MPA is mainly used for the performance and stability evaluation of timetables. Accordingly, in the following explanations, we limit ourselves to the basics required for this purpose. For further details on the method and additional examples, we refer to the references mentioned in this appendix.

D1 Mathematical symbols and variables

n	Number of events (nodes)
$\mathcal{N} \in \{1,...,n\}$	Set of all events
$i, j \in \mathcal{N}$	Event, where j and i denote the subsequent and preceding event, respectively
m	Number of processes (j,i) (edges)
p	Number of periods (order) of the system
T	Period duration of the system, in our case here this is typically $T = 60$ minutes
k	Index of the period considered, with $k \in \mathbb{N}_0$ and $\left[k \cdot T, (k+1) \cdot T\right)$ being the corresponding time window
d_i^0	Starting time (departure) of event i, with $d_i^0 \in [0, T)$ (initial time table)
d_0	Initial timetable vector $d_0 = \left(d_1^0, ..., d_i^0, ..., d_n^0\right)^{\mathrm{T}}$ (for period 0), where T denote the transpose of a vector or matrix
$d_i(k)$	Departure time according to the timetable, for event i in period k, where $d_i(k) \in [0, T)$ holds
$d(k)$	Timetable vector for period k: $d(k) = \left(d_1(k), ..., d_i(k), ..., d_n(k)\right)^{\mathrm{T}}$
$x_i(k)$	Start time of event i in period k
$x(k)$	State vector representing the system in period k: $x(k) = \left(x_1(k), ..., x_i(k), ..., x_n(k)\right)^{\mathrm{T}}$
s_{ij}	Slack time between the events j and i
r_{ij}	Recovery time between the events j and i
a_{ij}	Minimum processing time between the events j and i, where $a_{ij} \in \mathbb{R}_{\geq 0}$
$A \in \mathbb{R}_{\max}^{n \times n}$	General process matrix, where $a_{ij} = [A]_{ij}$
$A_0, ..., A_p$	Process matrices $A_l \in \mathbb{R}_{\max}^{n \times n}$ for $l = 0, ..., p$
a_{ij}^0	Planned processing time $a_{ij}^0 = a_{ij} + s_{ij}$
μ_{ij}	Period shift between event j and i

D2 Max-Plus Algebra: the basic idea

The Max-Plus-Algebra is an algebraic structure defined as $\mathbb{R}_{\max} = (\mathbb{R} \cup \{-\infty\}, \oplus, \otimes)$, i.e. the set $\mathbb{R} \cup \{-\infty\}$ together with the two operators addition $(\oplus)$ and multiplication $(\otimes)$, which are defined for all elements $a,b \in \mathbb{R} \cup \{-\infty\}$ according to $a \oplus b := \max(a,b)$ and $a \otimes b := a + b$ (on the left-hand-side you find the max-plus formulation of the corresponding operators in the conventional notation (right-hand-side)). Furthermore, the ε-element (the so-called zero element) has been established by $\varepsilon := -\infty$, and the element e (unit element) was defined by $e := 0$.

The above example shows how the two operators $\oplus$ and $\otimes$ are applied to two scalar values. In order to make the following mathematical formulations comprehensible, the application of the operators to matrices will now be shown. The general rule also applies here: Multiplication before addition.

Let $A, B \in \mathbb{R}_{\max}^{n \times n}$ be two $n \times n$-matrices with the elements in $\mathbb{R}_{\max}$. The matrix addition and matrix multiplication are defined for any matrices of form $A = (a_{ij}), B = (b_{ij}) \in \mathbb{R}_{\max}^{n \times n}$ as follows:

$$[A \oplus B]_{ij} = a_{ij} \oplus b_{ij} = \max(a_{ij}, b_{ij}) \tag{1}$$

and

$$[A \otimes B]_{ij} = \bigoplus_{k=1}^{n} a_{ik} \otimes b_{kj} = \max_{k=1,\ldots,n} (a_{ik} + b_{kj}). \tag{2}$$

Since the extensive possibilities of the MPA cannot be described in greater detail in this report, we refer once again to some standard literature with a very close relation to railway systems: Goverde (2005), Heidergott et al. (2006), Hansen and Pachl (2008).

D3 Definition of a periodic timetable

The so-called timetable vector defines the initial periodic timetable d_0. As shown above, d_i^0 denotes the time of the event i in the period 0, where $d_i^0 \in [0, T)$ applies. In general, the time of event i in any period $k = 0,1,\ldots$ is calculated as follows:

$$d_i(k) = d_i^0 + k \cdot T\,. \tag{3}$$

D4 Constraints to be considered

In order to calculate the start time of an event i, various constraints must be considered, which result from the requirements of the timetable and preceding events. These constraints are now briefly explained.

a. Constraints concerning the timetable

An event i shall never begin earlier than specified in the timetable. Therefore:

$$x_i(k) \geq d_i(k) \tag{4}$$

b. Consideration of preceding events

In order to maintain the minimum process time of a process (j,i) for a given start time x_j of a preceding event, the following condition must always be fulfilled:

$$x_i(k) \geq a_{ij} + x_j\left(k - \mu_{ij}\right), \tag{5}$$

wherein the period shift is calculated by the initial schedule d_0, the minimum process time a_{ij} and the schedule period T as follows:

$$\mu_{ij} = \frac{a_{ij}^0 + d_j^0 - d_i^0}{T} \in \mathbb{N}_0 \tag{6}$$

Eight different types of minimum process times a_{ij} between the events j and i are distinguished:

- minimum running times,
- pass times
- minimum turnaround times
- minimum dwell times,
- reservation times,
- release times,
- minimum transfer times (connections),
- minimum departure headway times.

c. Generalisation of constraints

The start time of event i in period k as a function of all preceding events and under consideration of the mentioned constraints (4) and (5) can be written in conventional notation as follows:

$$x_i(k) = \max\left(\max_j\left(a_{ij} + x_j\left(k - \mu_{ij}\right)\right), d_i(k)\right), \quad i = 1,\dots,n. \tag{7}$$

Formulated in Max Plus notation, the equation reads as:

$$x_i(k) = \bigoplus_{j=1}^{n}\left(a_{ij} \otimes x_j\left(k - \mu_{ij}\right)\right) \oplus d_i(k), \quad i = 1,\dots,n. \tag{8}$$

D5 State of the time-controlled Max-Plus linear system

In order to be able to describe the state of the overall system in period k, the state vector is finally calculated as follows:

$$x(k) = \bigoplus_{l=0}^{p} A_l x(k-l) \oplus d(k) = \bigoplus_{l=0}^{p} A_l \otimes x(k-l) \oplus d(k), \text{ with } d(k) = d_0 + k \cdot T. \tag{9}$$

The use of the MPA in the context of this work is limited to the analysis of the stability and performance of the system. For further uses, i.e. investigations of the dynamics system with MPA and further analyses such as disturbance propagation we refer to corresponding literature, e.g. Goverde (2010).

E Computation of stability and performance measures

Based on the explanations above, we now show how the performance and stability measures considered are calculated and which criteria must be fulfilled for stable operation.

E1 Input variables

In order to calculate the measures considered in this context, the so-called adjacency list and the timetable vector d_0 are required. The adjacency list contains for each of the processes $r = 1 \ldots m$ (edges) between the events j and i (nodes) a row vector of the form:

$$AL_r = \left(j(r), i(r), \mu_{ij}(r), a_{ij}(r) \right). \tag{10}$$

The four elements have the following meaning: $j(r)$ and $i(r)$ denote the number of the preceding or following event of the process r, $\mu_{ij}(r)$ indicates, whether there is a periodic jump between the events $j(r)$ and $i(r)$, and $a_{ij}(r)$ finally contains the process time of the event r. The minimum process times are used to calculate the eigenvalue and the associated eigenvector as well as the critical cycle, whereas the planned process times are used to test the feasibility of the schedule.

E2 Characteristics for assessing stability and performance

In order to evaluate the stability of a system, which is defined by the so-called process matrix, it is necessary to solve the eigenvalue problem or in this case, the generalised eigenvalue problem. If the problem can be solved, this results in the eigenvalue and eigenvector of the system. In the following, it will be shown how these values can be calculated and how additional values for the evaluation of the stability can be determined in further steps.

a. Eigenvalue and eigenvector of the system

The eigenvalue problem related to a quadratic state matrix $A \in \mathbb{R}_{\max}^{n \times n}$ describes the problem of searching for a scalar $\lambda \in \mathbb{R}_{\max}$ and a corresponding vector $v \in \mathbb{R}_{\max}^{n} \setminus \{\varepsilon\}$, so that (in Max-Plus-notation) the following equation is fulfilled:

$$A \otimes v = \lambda \otimes v. \tag{11}$$

If a solution (λ, v) exists, $\lambda = \lambda(A)$ denotes the eigenvalue and v the corresponding eigenvector.

For the systems considered in this work, the generalised eigenvalue problem for the quadratic, polynomial and non-reducible matrix $A(\gamma)$ of the form

$$A(\gamma) = \bigoplus_{l=0}^{p} A_l \gamma^l \in \mathbb{R}_{\max}^{n \times n}[\gamma] \tag{12}$$

has to be solved and a scalar $\lambda \in \mathbb{R}_{\max} \setminus \{\varepsilon\}$ and a corresponding vector $v \in \mathbb{R}_{\max}^{n} \setminus \{\varepsilon\}$ need to be found such that the following equation is fulfilled (with $\gamma = \lambda^{-1}$):

$$A\left(\lambda^{-1}\right) \otimes v = v, \tag{13}$$

where γ denotes the so-called backward shift operator. The application of this operator on $x(k)$ leads to $\gamma x(k) = x(k-1)$. Furthermore γ^l denotes a shift backwards in time by $l(l \geq 1)$ periods, and corresponds to an l-fold application of γ. Accordingly, this results in $\gamma^l x(k) = x(k-l)$ for all integer values $l \geq 1$. Furthermore $\gamma^0 = e$ applies. According to these definitions (12) can be written as:

$$x(k) = \bigoplus_{l=0}^{p} A_l x(k-l) \oplus d(k) = \bigoplus_{l=0}^{p} A_l \gamma^l x(k) \oplus d(k) = A(\gamma) x(k) \oplus d(k). \tag{14}$$

The solution of the generalised eigenvalue problem can be interpreted here as follows: If the eigenvector v is used as initial schedule (i.e. $d_0 = v$), the Max-Plus linear system described by equation 9 has a periodic behaviour with cycle time $T = \lambda_0$, where λ_0 is the maximum mean cycle time. Thus λ_0 represents the maximum cycle time over all possible cycles of the system. Unless otherwise noted, $\lambda = \lambda_0$ is assumed. Matrix $A\left(\lambda^{-1}\right) \in \mathbb{R}_{\max}^{n \times n}$ stands for the polynomial matrix A evaluated at λ^{-1} and is calculated as follows:

$$A_\lambda = A\left(\lambda^{-1}\right) = \bigoplus_{l=0}^{p} A_l \lambda^{-l} = \bigoplus_{l=0}^{p} A_l - l \cdot \lambda. \tag{15}$$

For the case with $p = 1$, a case that is often encountered here, this results:

$$A_\lambda = A_0 \oplus (A_1 - \lambda). \tag{16}$$

Various methods exist to solve the generalised eigenvalue problem (16), for example, Karp (Karp, 1978) and various extensions, Power Algorithm (Braker and Olsder, 1993 and Subiono and Van der Woude, 2000) or Policy Iteration Algorithm (Cochet-Terrasson et al., 1998). For the method described here, the fast and reliable policy iteration algorithm was implemented and used.

To assess the stability of the system, the following three cases can be distinguished, depending on the eigenvalue λ_0 of the critical cycle:

$\lambda_0 < T$ → the system is stable

$\lambda_0 = T$ → the system is critical

$\lambda_0 > T$ → the system is unstable

Two measures that can be calculated from the eigenvalue and the period duration are the so-called network throughput, which can also be interpreted as the utilisation of the available capacity:

$$\rho = \frac{\lambda_0}{T} \tag{17}$$

and the average total buffer of the critical cycle in comparison to the period duration,

$$\Delta_1 = \lambda_0 - T. \tag{18}$$

For a railway system to be in a stable state, $0 < \rho < 1$ must be fulfilled. If $\rho = 1$, the system is in a critical state.

Δ_1 is a measure of the robustness or sensitivity of the system concerning disturbances. Δ_1 can be interpreted as the additional active delay of any event that is part of the critical cycle that brings the timetable or the system into a critical state.

Since this case study with the MPA does not include any further analyses of system stability, we will not b discuss it here. However, we refer to detailed explanations on timetable stability, for example, in Goverde (2007, 189 ff.).

b. Critical path/cycle

The critical path is the sequence of events that results in the maximum mean cycle time. The critical (or longest) path matrix is defined as

$$A^+ = \bigoplus_{l=1}^{\infty} A^l = A \oplus A^2 \oplus A^3 \oplus \cdots, \tag{19}$$

where A^l is the matrix that has the maximum weights of the paths with length l. If there are no positively weighted cycles, one can write:

$$A^+ = \bigoplus_{l=1}^{n} A^l, \tag{20}$$

because every critical path with a length of $> n$ must contain a path with a weight of 0.
Further literature with detailed explanations of the Max-Plus algebra and its connection to graph theory can be found, for example, in Baccelli et al. (1993).

In order to finally determine the critical nodes of the system and the associated sequence, the critical path matrix is calculated according to (20) with A_λ (15) instead of A:

$$A_\lambda^+ = \bigoplus_{l=1}^{n} A_\lambda^l. \tag{21}$$

A node belongs to the critical cycle if the corresponding diagonal element corresponds to the unit element e, i.e. has a value of 0. Accordingly, the set of critical nodes results from

$$\mathcal{N}_{\text{crit}} = \left\{ i \in \mathcal{N} \middle| \left[A_\lambda^+ \right]_{ii} = e \right\}. \tag{22}$$

The temporal sequence defines the order of the nodes forming the critical cycle according to the adjacency list, from which vector c results:

$$c = \left(i_{(1)}, i_{(2)}, \ldots, i_{(n_{\text{crit}})} \right)^{\mathrm{T}}.$$

For example, $i_{(2)}$ denotes the second of a total $n_{\text{crit}} = \left| \mathcal{N}_{\text{crit}} \right|$ (ordered in ascending order over time) events in the critical cycle, and $\left| \cdot \right|$ stands for cardinality, i.e. the number of elements in a set.

c. Feasibility, stability and robustness of a timetable

Realizability

The feasibility of a timetable is a vital requirement for planning and operation. For a timetable, the following condition must be fulfilled for all processes of the system:

$$d(k) \geq d'(k) = \bigoplus_{l=0}^{p} A_l \otimes d(k-l), \text{ for all } k \in \mathbb{N}. \tag{23}$$

Now, the following holds for a single event i:

$$d_i(k) \geq d_i'(k) = [A_l]_{ij} \otimes d_j(k-l), \quad \text{for all } i, j, k, l . \tag{24}$$

If the timetable is periodic, as in this case, it can be written in simplified form:

$$d_0 \geq d_0' = A_T \otimes d_0 , \tag{25}$$

with

$$A_T = A\left(T^{-1}\right) = \bigoplus_{l=0}^{p} A_l T^{-l} = \bigoplus_{l=0}^{p} A_l - l \cdot T . \tag{26}$$

Robustness of a timetable

In subsection a (eigenvalue and eigenvector of the system) above, some explanations were already provided on the stability of the system due to the maximum mean cycle time λ_0 in relation to the length of the timetable period T.

In this section, we further introduce two very important measures: the slack $S = \left(s_{ij}\right) \in \mathbb{R}_{\max}^{n \times n}$ and the recovery matrix $R = \left(r_{ij}\right) \in \mathbb{R}_{\max}^{n \times n}$.

The elements s_{ij} of the slack matrix indicate how large the time buffer is between two elements j and i. The calculation is as follows:

$$s_{ij} = d_j^0 - d_i^0 + [A_T]_{ij} , \tag{27}$$

d_i^0 and d_j^0 are defined by the initial timetable, and A_T is defined by equation (29).

If the Max-Plus linear system (12) is realisable and the slack matrix is defined, it fulfils the $s_{ij} \leq 0$ for all $i, j \in \mathcal{N}$.

To be able to assess the robustness of a timetable in the event of disruptions, the so-called recovery matrix is determined. Their elements (j,i) are calculated for each process according to

$$r_{ij} = d_i^0 - d_j^0 - \left[A_T^+\right]_{ij} \tag{28}$$

where A_T^+ can be calculated with:

$$A_T^+ = \bigoplus_{l=1}^{n} A_T^l . \tag{29}$$

r_{ij} represents the maximum permitted delay of an event j (with $x_j(k)$), with which a subsequent event $x_i(k')$ is not influenced for all periods $k' \geq k$. If $r_{ij} = \infty$, no path exists from j to event i.

In the case of a feasible timetable, the condition $r_{ij} \geq 0$ is met for all $i, j \in \mathcal{N}$.

Furthermore, it can be shown for this case (see also Goverde, 2005) that the slack and recovery matrix are connected as follows:

$$r_{ij} = -\left[S^{+}\right]_{ij} \tag{30}$$

with

$$S^{+} = \bigoplus_{l=1}^{n} S^{l}\,. \tag{31}$$

F Computation of Service Intention Indices

F1 Travel times

In this section, we define the notation and computation of travel times, given a feasible timetable and adjacency list.

Basic travel times

The travel time T_{uv} between an origin node u and a destination node v for a specific trip k is defined by $T_{uvk} = t_{vk}^{\mathrm{arr}} - t_{uk}^{\mathrm{dep}}$, i.e., the difference between the arrival time (superscript arr) at node v and the departure time (superscript dep) at node u, with $u, v \in \mathcal{V}$ and $\mathcal{V}$ being the set of all stops considered. In the context here, a stop is defined as a platform at a station. For the two cases (planned and disponed), the travel times are determined according to

$$T_{uvk}^{\mathrm{plan}} = t_{vk}^{\mathrm{plan,arr}} - t_{uk}^{\mathrm{plan,dep}} \tag{32}$$

and

$$T_{uvk}^{\mathrm{dispo}} = t_{vk}^{\mathrm{dispo,arr}} - t_{uk}^{\mathrm{dispo,dep}}. \tag{33}$$

Due to the fact, that not all trips can be executed as planned, either with delays or even cancellations of trips, the disponed travel time of the simple form (33) needs to be extended as follows:

$$T_{uvk}^{\mathrm{dispo}} = \theta_k^{\mathrm{dispo}} \cdot \left(t_{vk}^{\mathrm{dispo,arr}} - t_{uk}^{\mathrm{dispo,dep}} \right) + \left(1 - \theta_k^{\mathrm{dispo}} \right) \cdot T_{uvk}^{\mathrm{estim}}. \tag{34}$$

where function $\theta_k^{\mathrm{dispo}}$ indicates if trip k can be carried out or not. It is defined as

$$\theta_k^{\mathrm{dispo}} = \begin{cases} 1 & \text{if } k \in \mathcal{N}_{uv}^{\mathrm{dispo}} \\ 0 & \text{otherwise} \end{cases}, \tag{35}$$

where $\mathcal{N}_{uv}^{\mathrm{dispo}}$ denotes the set of all trips provided and T_{uvk}^{estim} represents a travel time of the affected trip that is estimated by the planner. T_{uvk}^{estim} is determined, amongst others, by the waiting time until departure and the travel time of alternatives (e.g. bus shuttles) or, if the trip needs to be cancelled, the expected arrival time of the next possible connection.

Travel times can be computed by any journey planning and/or travel time engine. All travel times in this project were computed by using OpenTripPlanner (OTP) (for details see OTP, 2018).

Travel time deviations

In the real-world case, travel times can, unfortunately, deviate from the planned ones due to various operational reasons. The most basic measure of a deviation for a single trip k between the origin node u and the destination node v is determined by

$$\Delta T_{uvk} = T_{uvk}^{\mathrm{dispo}} - T_{uvk}^{\mathrm{plan}}. \tag{36}$$

Using the definition in (3), we can rewrite (5) as

$$\Delta T_{uvk} = \theta_k^{\mathrm{dispo}} \cdot \left(t_{vk}^{\mathrm{dispo,arr}} - t_{uk}^{\mathrm{dispo,dep}} \right) + \left(1 - \theta_k^{\mathrm{dispo}} \right) \cdot T_{uvk}^{\mathrm{estim}} - T_{uvk}^{\mathrm{plan}}. \tag{37}$$

Aggregated and weighted travel times

To assess a timetable, we first compute the sum (total) of all travel times for both scenarios (planned, disponed) independently and for each specific origin-destination pair (u,v), using the variables T_{uvk}^{plan} and T_{uvk}^{dispo}, respectively. This leads to

$$TT_{uv}^{\text{plan}} = \sum_{k \in \mathcal{N}_{uv}^{\text{plan}}} T_{uvk}^{\text{plan}} \tag{38}$$

and

$$TT_{uv}^{\text{dispo}} = \sum_{k \in \mathcal{N}_{uv}^{\text{plan}}} T_{uvk}^{\text{dispo}}, \tag{39}$$

where the corresponding set of all planned trips is denoted by $\mathcal{N}_{uv}^{\text{plan}}$. To get meaningful results, it is important that a reasonable planning horizon is specified in advance. We define it as $T_{\text{h}} = h \cdot T_{\text{p}}$, where T_{p} denotes the period of the system, usually one hour, and $h \in \mathbb{N}$ represents the number of periods considered. Hence $\mathcal{N}_{uv}^{\text{plan}} = \mathcal{N}_{uv}^{\text{plan}}(T_{\text{h}})$ represents the number of trips starting and ending within the planning horizon.

F2 Service Intention Indices

The idea of the Service Intention Index (SII) is to assess the extent to which a realised (disponed) timetable deviates from a planned one. This represents, in an aggregated way, the view of the customers. There are various alternatives for how an SII can be defined. In the following, we present one specific way:

$$SII = \frac{\sum_{u \in \mathcal{V}} \sum_{\substack{v \in \mathcal{V}, \\ v \neq u}} \sum_{k \in \mathcal{N}_{uv}^{\text{plan}}} T_{uvk}^{\text{dispo}} - T_{uvk}^{\text{plan}}}{\sum_{u \in \mathcal{V}} \sum_{\substack{v \in \mathcal{V}, \\ v \neq u}} \sum_{k \in \mathcal{N}_{uv}^{\text{plan}}} T_{uvk}^{\text{plan}}}. \tag{40}$$

Index SII is useful to assess the overall system over the time window T_{h}. However, as for large systems, the sums of travel times might lead to large numbers for both scenarios, and hence, the ratio tends towards 1, it becomes difficult to capture real differences. Furthermore, to analyse the quality of the travel time of individual origin-destination pairs, we need an option to quantify deviations at a lower quantity.

To achieve this, we next introduce an absolute and a relative Service Intention Index for each origin-destination pair (u,v).

First, ΔTT_{uv} captures the sum of deviations of disponed trips and planned ones, i.e.

$$\Delta TT_{uv} = \sum_{k \in \mathcal{N}_{uv}^{\text{plan}}} T_{uvk}^{\text{dispo}} - T_{uvk}^{\text{plan}}. \tag{41}$$

Next, SII_{uv} represents the absolute deviations relative to the sum of all planned travel times for origin-destination pair (u,v):

$$SII_{uv} = \frac{\sum_{k \in \mathcal{N}_{uv}^{\text{plan}}} T_{uvk}^{\text{dispo}} - T_{uvk}^{\text{plan}}}{\sum_{k \in \mathcal{N}_{uv}^{\text{plan}}} T_{uvk}^{\text{plan}}} = \frac{\Delta TT_{uv}}{\sum_{k \in \mathcal{N}_{uv}^{\text{plan}}} T_{uvk}^{\text{plan}}} . \tag{42}$$

With the relative version, i.e. SII_{uv} , we can directly compare the quality of services offered. Finally, given all ΔTT_{uv} , we can easily see the relation to SII :

$$SII = \frac{\sum_{u \in \mathcal{V}} \sum_{\substack{v \in \mathcal{V}, \\ v \neq u}} \Delta TT_{uv}}{\sum_{u \in \mathcal{V}} \sum_{\substack{v \in \mathcal{V}, \\ v \neq u}} \sum_{k \in \mathcal{N}_{uv}^{\text{plan}}} T_{uvk}^{\text{plan}}} . \tag{43}$$

G Computation of the Cumulative Delay Impact and Cumulative Delay Sensitivity

The Cumulative Delay Impact (CDI) is a measure to quantify the overall impact of a delay κ at a specific event (node) j on all other events (nodes). Formally the CDI is computed as follows:

$$CDI_j\left(\mathbf{R}\middle|\kappa,\gamma\right) = \sum_{i\in\mathcal{N}\setminus j} \max\left(\kappa - r_{ij}, 0\right)^{\gamma} , \tag{44}$$

where $\mathcal{N}$ denotes the set of all events of the system, $\mathbf{R}$ represents the recovery matrix of size $N \times N$, with $N = |\mathcal{N}|$. r_{ij} represents the actual buffer time r_{ij} between events j (column in matrix $\mathbf{R}$), and i (row in matrix $\mathbf{R}$), κ is the parameter that denotes the initial delay (in minutes) applied to node j, for which the CDI_j shall be calculated, and finally $\gamma \geq 1$ is a parameter to increase the impact of positive differences between the delay κ and r_{ij}. In this study, γ was always set to 1. Furthermore, CDI_j is strictly monotonically increasing, with $CDI_j\left(\mathbf{R}\middle|0,\gamma\right) = 0$.

The Cumulative Delay Sensitivity (CDS) is a measure to quantify the impact of a delay κ applied to all events j except i on a specific event i. Formally the CDS is defined as follows:

$$CDS_i\left(\mathbf{R}\middle|\kappa,\gamma\right) = \sum_{j\in\mathcal{N}\setminus i} \max\left(\kappa - r_{ij}, 0\right)^{\gamma} , \tag{45}$$

where parameters, sets and variables are defined as above. $\gamma = 1$ holds here as well. Furthermore, is strictly monotonically increasing, with $CDS_i\left(\mathbf{R}\middle|0,\gamma\right) = 0$.

Periodic timetabling with ‘Track Choice’-PESP based on given line concepts and mesoscopic infrastructure

Raimond Wüst[1], Stephan Bütikofer[1], Severin Ess[1], Claudio Gomez[1], Albert Steiner[1],
Marco Laumanns[2], Jacint Szabo[3]

[1] Institute for Data Analysis and Process Design, Zurich University of Applied Sciences
ZHAW, 8401 Winterthur, Switzerland, Switzerland, `wura@zhaw.ch`
[2] Bestmile SA, 1007 Lausanne, Switzerland, `marco.laumanns@bestmile.com`
[3]IBM Research, now at Google Switzerland, Brandschenkestrasse 110, Zurich, Switzerland
`szabojacint@gmail.com`

Abstract. We present a track-choice and vehicle scheduling extension of the commonly known method for the generation of periodic event schedules ‘PESP’. The extension makes use of the mesoscopic track infrastructure representation widely used by public transport planners and operators. Taking into consideration the technical and operational constraints given by rolling stock, station and track topology data on the one hand, and the commercial requirements defined by a given line concept on the other, the method presented generates periodic timetables including train-track assignments. Due to the utilization of infrastructure based track capacities, we are also able to assess the feasibility of the line concept given. Additionally, the method allows for handling temporary resource restrictions (e.g. caused by construction sites or operational disturbances) up to a certain degree.

Keywords: Periodic Event Scheduling Problem, Mesoscopic railway topology, Service Intention, Track Choice

1 Introduction

In the operational management of railway networks, an important requirement is the fast adaptation of timetable scenarios, in which operational disruptions or time windows with temporary unavailability of infrastructure, for instance during maintenance time windows, are taken into account. In those situations, easy and fast reconfiguration and recalculation of timetable data is of central importance. This local and temporal rescheduling results in shifted departure and arrival times and sometimes even in modified stop patterns at intermediate stations of train runs. In order to generate reliable timetabling results it is a prerequisite that train-track assignments, as well as operational and commercial dependencies are taken into consideration and that all these dependencies are not conflicting with each other. Hence, finding the right level of detail for modelling track infrastructure and train dynamics is crucial for supporting the planning process in an optimal way. This requirement motivated several research groups to combine common timetabling procedures with constraints resulting from mesoscopic infrastructure information in recent years.

From the existing approaches, we will discuss below some that are relevant to our work. Hansen and Pachl [6] show how running, dwell and headway times at critical route nodes and platform tracks must be taken into account for train processing and present a deep timetable quality analysis depending on these parameters. De Fabris et al. [4] calculate arrival and departure time, platform and the route in stations and junctions that trains visit along their lines. Bešinović et al. [1] present a micro–macro framework based on an integrated iterative approach for computing a microscopically conflict-free timetable that uses a macroscopic optimization model with a post-processing robustness evaluation. Caimi et al. [3] extend PESP (see e.g. [7]) by proposing the flexible periodic event scheduling problem (FPESP), where intervals are generated instead of fixed event times. By applying FPESP, the output does not define a final timetable but an input for finding a feasible timetable on a microscopic level, ([2] and [3]).

Our modelling approach is also based on an extension of PESP and takes the service intention (SI) as input data structure. The SI was first described in Caimi [2] and integrates commercial timetabling requirements given by the respective line concept on one side and technical constraints on the other. It largely corresponds to the ‘line concept’, and represents functional timetabling requirements including line data, line frequencies and separations as well as line transfers at specific stations. Similarly to de Fabris et al. [4], we call this level of abstraction of the available resources ‘mesoscopic topology’. Together with the functional requirements of the SI this mesoscopic infrastructure data model of a given scenario is entered into a standard timetable editor (see, e.g. SMA Viriato, [8]).

2 Methodology

The investigation of feasible event times for individual train runs and the corresponding resource allocations fitting into an integrated clock face timetable is usually done manually in a time consuming way. On the other hand, algorithmic approaches for solving this task computationally require models based on microscopic information about track capacity. This capacity information can be aggregated to headway constraints that are used for solving standard periodic timetable problems. In order to facilitate this step, we present a generic approach, which makes use of the mesoscopic infrastructure. We call this approach Track-Choice PESP (TCPESP), as it can be considered as an extension of PESP, which includes the selection of relevant headway constraints into the optimization problem.

The SI is defined by a set of train runs. Each train run belongs to a line L and is characterized by the sequence of sections that are traversed and a corresponding time interval, which is required for either running or stopping on a corresponding track section. Each time interval has a minimal and maximal value. Stop nodes typically provide a service for boarding or de-boarding. Together, a pair of train runs moving in opposite directions makes up a train circulation.

In the TCPESP model, the mesoscopic infrastructure consisting of sections is summarized as a set I of operation points. Operation points are largely tracks and stations but can also be other critical resources as junctions (see example below). As mentioned before, each operation point $i \in I$ is associated to a capacity consisting of a set of tracks T_i. A train run $l \in L$ is described by a sequence of operation points of I.

Based on our mesoscopic model we form an event-activity network (E, A). The set E of events consists of an arrival event arr_{li} and a departure event dep_{li} for each train run $l \in L$ and operation point $i \in l$. The activities $a \in A$ are directed arcs from $E \times E$ and describe the dependencies between the events. For every train run we have arcs between arrival and departure events at the same operation points (dwell times or trip times) and arcs between departure and arrival events of successive operation points (time needed for the travel between operation points). Further arcs include connections between train runs, headways and turnaround operations (see section 3). We refer to [7] for a detailed overview of the modelling options of dependencies. Fig. 1 provides a sample of such an event graph.

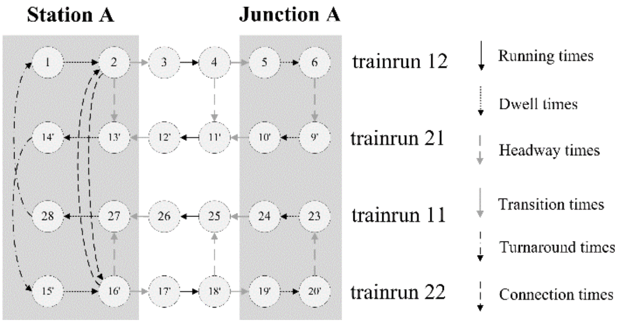


Fig. 1 Sample of an event activity network, where arcs connect arrival and departure events. Nodes belonging to grey shaded boxes indicate events at operation point type operation points. Other nodes indicate track type arrival and departure events. Arrow line styles indicate different types of time dependencies.

Headway arcs $a \in A_H$ are especially important for explaining the 'track-choice PESP model' below. Headway arcs are used to model safety distances between trains running in the same and in opposite directions (see example in Fig. 1). For the sake of simplicity we consider in TCPESP (1) below only headways related to one operation point, i.e. we omit headways for train runs in opposite directions over several successive operation points. The problem formulation (1) can be easily extended to include general headways.

The classical PESP tries to determine a periodic schedule on the macroscopic level (i.e. without using the tracks at an operation point) within a period T. Event $e \in E$ takes place at time $\pi_e \in [0, T)$. The schedule is periodic with time T, hence each event is repeated periodically $\{\dots, \pi_e - T, \pi_e, \pi_e + T, \pi_e + 2T, \dots\}$.

The choices of the event times π_e depend on each other. The dependencies are described by arcs $a = (e, f)$ in A and modeled as constraints in the PESP. The constraints always concern the two events e and f and define the minimum and maximum periodic time difference l_a and u_a between them. These bounds are given as parameters in the PESP model. We therefore look for the event times π_e for every $e \in E$ that fulfill all constraints of the form

$$l_a \leq \pi_e - \pi_f + p_a T \leq u_a, \text{ for all } a = (f, e) \in A, \tag{1}$$

where p_a is an integer variable that allows the constraints of the form (2) to be met in a periodic sense.

Track-choice PESP model. We extended the classical PESP model by using the number of tracks T_i at each operation point $i \in I$. The track-choice PESP model assigns the arrival event arr_{li} and the departure event dep_{li} of train run l at operation point i uniquely to a track in T_i. We can use these assignments to switch on headway arcs $a \in A_H$ by using the following big-M-approach. In addition to variables π and p from the classical PESP model we need: (i) Binary variables tc_{et} (track choice) for each event $e \in E$ and track $t \in T_{i(e)}$, where operation point $i(e)$ is associated to event e, i.e. e is equal to arr_{li} or dep_{li} for a train run l. (ii) Binary variables h_a for every headway edge $a = (f, e) \in A_H$. Headway edges are always between events at the same operation point, therefore $T_{i(e)} = T_{i(f)}$ holds. The track-choice model is defined by:

$$\begin{array}{llr}
\min f(\pi, p) & & \\
s.t. \quad l_a \leq \pi_e - \pi_f + p_a T \leq u_a, & \forall\ a = (f, e) \in A \setminus A_H, & (1) \\
l_a - (1 - h_a)M \leq \pi_e - \pi_f + p_a T \leq u_a + (1 - h_a)M, & \forall\ a = (f, e) \in A_H, & (2) \\
\sum_{t \in T_{i(e)}} tc_{et} = 1, & \forall\ e \in E, & (3) \\
tc_{arr_{li}t} = tc_{dep_{li}t}, & \forall\ l \in L, i \in l,\ t \in T_i, & (4) \\
h_a \geq tc_{et} + tc_{ft} - 1, & \forall\ a = (f, e) \in A_H,\ t \in T_{i(e)} & (5) \\
tc_{et}, h_a \in \{0,1\}, \pi_e \in [0, T), p_a \in \mathbb{Z}, & \forall\ e \in E, t \in T_{i(e)}, a \in A, &
\end{array}$$

where M is a big enough natural number.

There are many different objective functions $f(\pi, p)$ described in literature [7]. In our test case below we minimize the total passenger travel time. In (1) are the normal PESP constraints summarized (without headway arcs). In (2) are the headway constraints, which can be switched off with a big-M technique. The assignment of the events to the tracks is done in (3). (4) is used to assign the corresponding arrival and departure events to the same track. In (5) the headway variable is set to 1, if the events take place on the same track, i.e. the headway is required at this operation point.

3 Case study

In order to validate the proposed TCPESP model we designed a simple test case. The relationship between the macroscopic timetable events of three train lines are illustrated by means of a simplified network graph (see Fig. 2a). To validate the model, a virtual railway network was defined for which the service intention was implemented (see Fig. 2b).

The test network contains the two main station nodes (Station A and Station B) connected by line 2, and three stop stations, Stop A, Stop AT (served by line 1) and Stop BT (served by line 3). The planning-relevant secondary conditions for the case study are limited to stations A and B. The period of each train run is indicated in Table 1.

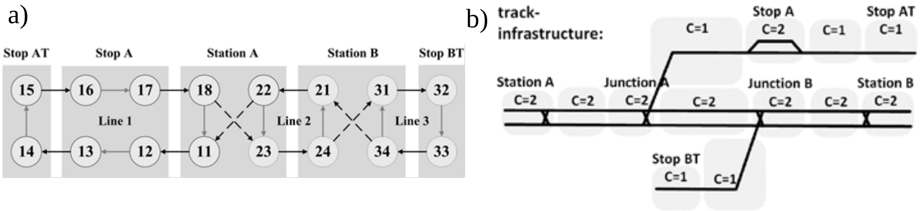


Fig. 2. a) Schedule activity network with commercial dependencies modified from Goverde. Line 1, serving Stop A and Stop AT and connecting to Line 2 in Station A. Line 2 connecting Stations A and B. Line 3, serving Stop BT and connecting to Line 2 in Station B. b) Track infrastructure of the test scenario with an indication of track capacities at each operation point. Operation points indicated as grey shaded boxes.

Fig. 2a shows the service intention including train lines and commercial dependencies between single train runs of each line. Table 1 below provides an example of constraints related to the hourly service of line 2 running from station A (St A) to station B (St B). Fig. 2b shows the track infrastructure of the scenario together with the mesoscopic section topology indicating the section capacities by the corresponding number of horizontal lines.

The SI of test case A offers an hourly service of line 2 between major Stations A and B with connections to and from line 1 in station A and to and from line 3 in station B. A complete rotation of line 1 and 2 takes 120 minutes, one of line 3 takes 60 minutes. Therefore two vehicles are needed for rotations of line 1 and 2 and one is needed for line 3. Line services with train runs and corresponding periodicity and minimum circulation times are indicated in Table 1.

Table 1. Line services with minimal trip times and periods. Odd numbers indicate train runs in one direction; even numbers indicate train runs of the same line in the opposite direction.

Line ID	Service ID	Minimum trip time	Period
1	11	50	60
1	12	50	60
1	13	50	60
1	14	50	60
2	21	50	60
2	22	50	60
2	23	50	60
2	24	50	60
3	31	20	60
3	31	20	60
3	32	20	60
3	32	20	60

Fig. 3 illustrates the results of the TCPESP algorithm for the given test scenario. In addition to the output of the conventional PESP algorithm given by arrival and departure event times, the result that we obtain from the TCPESP model includes track assignment information for each train run. The rail infrastructure of the test scenario consists of two single-track lines (line 1 and 3) and one double track line (line 2). We indicate the resulting track assignment by track numbers (T1 and T2) to each train run during run time on a given track section (see track diagram above each line diagram). There, the number of grey bold horizontal lines is identical to the number of tracks available at a corresponding operation points (T1 or both T1 and T2, respectively). From Fig. 3 it can be seen that the TCPESP algorithm only permits contra rotating train runs to meet in double track sections (line 1) and connecting train runs to meet in a station on neighboring tracks (platforms; St A: line 1 and 2, St B: line 2 and 3). Line styles correspond to directed train runs in both, the track diagrams and the time diagrams.

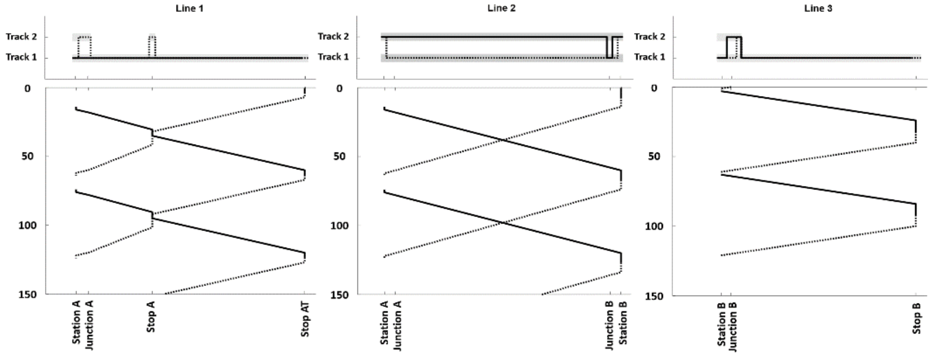


Fig. 3 Scheduling results obtained from of our TCPESP model. A train diagram with the arrival and departure event times is plotted together with the track assignment. Vertical axis: time between 0 and 150 minutes, horizontal axis: sequential locations. St A: station A, St B: station B, Stp A: Stop A, Stp AT: Final stop at AT. T1 and T2 with grey shaded horizontal lines above each location-time diagram indicate track assignments for each vehicle circulation of the three given lines.

4 Discussion and outlook

We introduced and successfully applied the new timetabling model TCPESP, which can be used to support timetable planners for generating train and vehicle schedules with track assignment. This model is based on an extension of the well-known PESP model and can be configured by using a standard schedule editor. Future developments include (i) the generation of the SI using a standard line planning method (see e.g. [5]); (ii) the evaluation of timetable stability. In that way, we expect to further improve the quality of TCPESP results and contribute for speeding up and facilitating practical railway timetabling.

References

1. Bešinović, N., Goverde, R.M.P., Quaglietta, E., Roberti, R.: An integrated micro-macro approach to robust railway timetabling. Transportation Research Part B: Methodological, vol. 87, pp. 14-32 (2016)
2. Caimi, G.: Algorithmic decision support for train scheduling in a large and highly utilised railway network. Diss. ETH Zürich Nr. 18581 (2009)
3. Caimi, G., Fuchsberger, M., Laumanns, M., Schüpbach, K.: Periodic railway timetabling with event flexibility. Networks, vol. 57, issue 1, pp. 3-18 (2011b)
4. de Fabris, S., Longo, G., Medeossi, G., Pesenti, R.: Automatic generation of railway timetables based on a mesoscopic infrastructure model. Journal of Rail Transport Planning & Management, vol. 4, pp. 2–13 (2014)
5. Friedrich, M., Hartl, M., Schiewe, A. and Schöbel, A.: Integrating passengers' assignment in cost-optimal line planning. Technical Report 2017-5, Preprint-Reihe, INAM, Georg-August Universität Göttingen, (2017).
6. Hansen, I.A., Pachl., J. (eds.): Railway Timetable & Traffic. Analysis – Modelling – Simulation. Eurail Press, Hamburg, Germany (2008)
7. Liebchen, C., Möhring, R.H.: The modeling power of the periodic event scheduling problem: Railway timetables – and beyond. In: Geraets, F., Kroon, L., Schoebel, A., Wagner, D., Zaroliagis, C. (Eds.), Algorithmic Methods for Railway Optimization, Lecture Notes in Computer Science, vol. 4359, pp. 3–40, Springer, Berlin/Heidelberg, D (2007)
8. SMA: Viriato - software for railways. http://www.sma-partner.ch, Zurich. last accessed 2018/07/01

Improvement of maintenance timetable stability based on iteratively assigning event flexibility in FPESP

Raimond Wüst [a,1], Stephan Bütikofer [a], Severin Ess [a], Claudio Gomez [a],
Albert Steiner [a],
Marco Laumanns [b], Jacint Szabo [c]
[a] Institute for Data Analysis and Process Design, Zurich University of Applied Sciences
ZHAW, 8401 Winterthur, Switzerland
[1] E-mail: wura@zhaw.ch, Phone: +41 (0) 58 934 65 81
[b] Bestmile SA, 1007 Lausanne, Switzerland
[c] IBM Research, now at Google Switzerland, Brandschenkestrasse 110, Zurich,
Switzerland

Abstract

In the operational management of railway networks, an important requirement is the fast adaptation of timetable scenarios, in which operational disruptions or time windows with temporary unavailability of infrastructure, for instance during maintenance time windows, are taken into consideration. In those situations, easy and fast reconfiguration and recalculation of timetable data is of central importance. This local and temporal rescheduling results in shifted departure and arrival times and sometimes even in modified stop patterns at intermediate stations of train runs. In order to generate reliable timetabling results it is a prerequisite that train-track assignments, as well as operational and commercial dependencies are taken into consideration. In order to refer to the right level of detail for modelling track infrastructure and train dynamics in the computer aided planning process we present a generic model that we call Track-Choice FPESP (TCFPESP), as it implements suitable extensions of the established PESP-model. We show how the service intention (the timetable specification resulting from line planning) together with resource capacity information can be utilized in order to configure the TCFPESP model.

In addition, we are able to calculate quantitative performance measures for assessing timetable quality aspects. In order to achieve this we present a method for evaluating timetable robustness. By utilizing delay impact values resulting from max-plus algebraic performance analysis, we are thus able to iteratively distribute event flexibility in such a way that overall robustness of the maintenance timetable is improved.

This approach supports the planner to generate integrated periodic timetable solutions in iterative development cycles.

Keywords

Flexible PESP, Mesoscopic railway topology, Service Intention, Timetabling with track assignment, Timetable stability analysis

1 Introduction

1.1 Generating and investigating temporary timetable scenarios

In the operational management of railway networks, an important requirement is the fast adaptation of timetable scenarios, in which operational disruptions or time windows with temporary unavailability of infrastructure, as for instance during maintenance time windows (‘possessions’, see RailNetEurope 2017), have to be accounted for. In those situations, easy and fast reconfiguration and recalculation of timetable data is of central importance. This local and temporal rescheduling results in shifted departure and arrival times and sometimes even in modified stop patterns at intermediate stations of train runs. Only recently, van Aken et al. presented a PESP based macroscopic model for solving train timetable adjustment problems (TTAP) under infrastructure maintenance possessions (2017a). They show, that by applying TTAP, they are able to adjust a given timetable to a specified set of station track and complete open-track possessions by train retiming, reordering, short-turning and cancellation. In (2017b) they apply several network aggregation techniques to reduce the problem size and thus enable the model to solve large instances within short computation times with instances of the complete Dutch railway network.

However, in order to generate reliable timetabling results it is prerequisite that besides train-track assignments, also operational and commercial dependencies are taken into consideration. Hence, finding the right level of detail for modelling track infrastructure and train dynamics is crucial for supporting the planning process in an optimal way.

In day-to-day business, determining the feasible event times for individual train runs and the corresponding resource-allocation fitting into efficient transport chains resulting from an integrated clockface timetable is time-consuming and is carried out manually. On the other hand, algorithmic approaches for solving this task computationally require models based on microscopic information about track capacity. This capacity information is aggregated to (normative) minimum headway constraints that are used for solving standard periodic timetabling problems. In order to facilitate this step, several research groups made suggestions, how to combine common timetabling procedures with constraints resulting from mesoscopic infrastructure information. Hansen and Pachl (2008) show how running, dwell and headway times at critical route nodes and platform tracks must be taken into account for train processing and present a deep timetable quality analysis depending on these parameters. De Fabris et al. (2014) calculate arrival and departure times, platform and route assignments in stations and junctions that trains visit along their lines. Bešinović et al. (2016) present a micro–macro framework based on an integrated iterative approach for computing a microscopically conflict-free timetable that uses a macroscopic optimization model with a post-processing robustness evaluation. Caimi et al. (2011) extend PESP (see e.g. Serafini and Ukovich (1989) and Liebchen and Möhring (2007)) and propose the flexible periodic event scheduling problem (FPESP), where intervals are generated instead of fixed event times. By applying FPESP, the output does not define a final timetable but an input for finding a feasible timetable on a microscopic level, (Caimi (2009) and Caimi et al. (2009)).

1.2 Service Intention based approach for timetable specification

To improve customer value even under limited operating conditions, such as those encountered during infrastructure maintenance intervals, our modelling approach for creating temporary schedules is also based on an extension of PESP and takes the ‘service intention’ (SI) as input data. The SI was first described in Wüst et al. (2008), formally specified in Caimi (2009) and integrates commercial timetabling requirements given by the respective demand oriented ‘line concept’ on one side and technical constraints on the other. The ‘line concept’ results from a strategical planning process which is executed by the transport carrier. In this process, the available amount, the dynamics and the circulation of rolling stock are taken into account. In Switzerland, the integrated fixed-interval timetable (IFIT) is created on the basis of SI's. The required system times (minimum travel times between node stations, see for example Herrigel (2015) and BAV (2011)) are a prerequisite (see e.g. Liebchen and Möhring (2007)).

The maintenance interval planning step (denoted as IP in the sequel) is executed by the infrastructure manager. In this step, the functional requirements of the SI are brought together with this mesoscopic infrastructure data model of a given scenario. Altogether these data can be maintained in a standard timetable editor (see for instance SMA Viriato, 2018). In this way, the SI represents functional timetabling requirements including line data, line frequencies and separations as well as line transfers at specific stations. Hence, it contains explicit information about intended transport chains but is still flexible enough, to allow different ways of operational planning and resource allocation. Like de Fabris et al. (2014), we call this level of abstraction of the available resources 'mesoscopic topology'. We call our FPESP model that we apply to this mesoscopic topology 'Track-Choice FPESP' (TCFPESP).

In order to evaluate timetable robustness criteria we use a special algebraic approach that is commonly known as max-plus algebra. This approach has been elaborated in mathematical detail by Goverde (2007) who also demonstrates the benefits of this algebraic approach for timetable stability analysis in practical applications. We apply this method for evaluating the robustness of our resulting timetable and try to improve the timetable based on this performance evaluation in successive re-planning iterations. More specifically, we show how the max-plus-delay impact analysis can help to improve timetable stability by iteratively adjusting local flexibility constraints in the configuration of the TCFPESP model.

1.3 Structure of this article

This article is structured as follows: In chapter 2, we describe the methodology for achieving the research goals. In section 2.1 we summarize the FPESP model which implements the idea of periodic timetabling with event flexibility. Extending this FPESP to our proposed mesoscopic model we present in Section 2.2 our TCFPESP-model. For the iterative configuration of the event flexibility in the TCFPESP we make use of the delay impact vector that we obtain from max-plus analysis. This is shown in section 2.3. In section 2.4 we describe the TCFPESP heuristic for reducing the overall delay impact. In chapter 3 (Case Study 'Kerenzerberg') we present the results from applying the methods introduced in chapter 2 and the coordinated application in a real-world scenario from eastern Switzerland. Finally, in chapter 4 we conclude with a summary of the findings and an outlook on future work.

2 Methodology

2.1 Periodic Timetabling with Event Flexibility

The classical PESP tries to determine a periodic schedule on the macroscopic level (i.e. without using the tracks at an operation point) within a period T. Event $e \in E$ takes place at time $\pi_e \in [0, T)$. The schedule is periodic with time period T, hence each event is repeated periodically $\{\dots, \pi_e - T, \pi_e, \pi_e + T, \pi_e + 2T, \dots\}$.

The choices of the event times π_e depend on each other. The dependencies are described by arcs $a = (e, f)$ from a set A and modelled as constraints in the PESP. The constraints always concern the two events e and f and define the minimum and maximum periodic time difference l_a and u_a between them. These bounds are given as parameters in the PESP model. We therefore look for the event times π_e for every $e \in E$ that fulfill all constraints of the form

$$l_a \leq \pi_e - \pi_f + p_a T \leq u_a$$

for all $a = (e, f) \in A$,, where p_a is an integer variable that makes sure, that these constraints are met in a periodic sense.

In order to avoid tedious iterations between the process steps “microscopic capacity planning” and “mesoscopic capacity planning” in case of infeasibility of the micro-level problem, one can improve the chance of finding a feasible solution by enlarging the solution space in the micro-level. This approach has been described in detail in Caimi et al. (2011b). We also implement this event flexibility method by adding some flexibility for the events of the event and activity network (E, A) by introducing lower and upper bounds to the event times of the arrival and departure nodes in Figure 1b. The final choice of the event times in the range between the lower and upper bound shall be independent for each event such that each value of the end of an activity arc should be reachable from each time value at beginning of that activity arc (see Figure 1a).

We are not forced to add this flexibility to all the events, but we can select the nodes where we want to add it, for instance only nodes corresponding to events in a main station area with high traffic density, where it is more difficult to schedule trains on the microscopic level. In general, one can say, that this placement of flexibility is the timetable configuration feature, which has the highest level of influence on improving operational stability. This is where the information provided by the max-plus measures of delay impact (see section 2.3 et seq.) can be utilized in order to achieve timetable robustness. For more details regarding the FPESP method, we refer to the article of Caimi et al. (2011b).

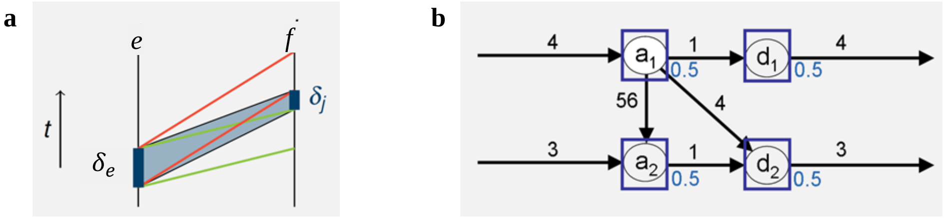


Figure 1: Target oriented placement of time reserves. a) Time frames $[\pi_e, \pi_e + \delta_e]$ in place of time points π_e. By implementing this method, the normal PESP constraints $l_a \leq \pi_e - \pi_f + p_a T \leq u_a$ now become $l_a + \delta_f \leq \pi_e - \pi_f + p_a T \leq u_a - \delta_e$ (see next section). In the example b) this means that instead of planning time points $(\pi_{a_1}, \pi_{d_1}, \pi_{a_2}, \pi_{d_2})$ we plan time frames $[\pi_e, \pi_e + 0.5]$ for $e \in \{a_1, d_1, a_2, d_2\}$.

2.2 Track-choice FPESP.

For our proposed timetabling model, we extend the FPESP method with events at track-level in order to generate event slot timetables on a mesoscopic level. In the TCFPESP model, the mesoscopic infrastructure consisting of sections is summarized as a set I of operation points. Operation points are largely tracks and stations but can also be other critical resources such as junctions (see OP ‘Tiefenwinkel’ in Figure 2b). As mentioned before, each operation point $i \in I$ is associated to a capacity consisting of a set of tracks T_i. A train run $l \in L$ is described by a sequence of operation points of I.

Based on this mesoscopic model we form an event-activity network (E, A). The set E

of events consists of an arrival event arr_{li} and a departure event dep_{li} for each train run $l \in L$ and operation point $i \in l.$ The activities $a \in A$ are directed arcs from $E \times E$ and describe the dependencies between the events. For every train run we have arcs between arrival and departure events at the same operation points (dwell times or trip times) and arcs between departure and arrival events of successive operation points (running time between operation points). Further arcs include connections between train runs, headways and turnaround operations. Headway arcs $a \in A_H$ are especially important for explaining the ′track-choice FPESP model′ below. Headways are used to model safety distances between trains running in the same and in opposite directions. For the sake of simplicity in the formal description of the TCFPESP we consider only headways related to one operation point, i.e. we omit headways for train runs in opposite directions over several successive operation points. The TCFPESP-model can be easily extended to include general headways.

We extended the classical PESP resp. FPESP model by using the number of tracks T_i at each operation point $i \in I.$ The track-choice FPESP model assigns the arrival event arr_{li} and the departure event dep_{li} of train run l at operation point i uniquely to a track in T_i. We can use these assignments to switch on headway arcs $a \in A_H$ by using the following big-M-approach. In addition to variables π and p from the classical PESP model we need: (i) Binary variables tc_{et} (track choice) for each event $e \in E$ and track $t \in T_{i(e)}$, where operation point $i(e)$ is associated to event e, i.e. e is equal to arr_{li} or dep_{li} for a train run l. (ii) Binary variables h_a for every headway edge $a = (e,f) \in A_H$. As mentioned before, headway edges are always between events at the same operation point, therefore $T_{i(e)} = T_{i(f)}$ holds. (iii) Positive variables δ_e for each event $e \in E$ to model the event flexibility.

The track-choice model is defined by:

$$\min f(\pi, \delta)$$

$$\begin{array}{rll}
s.t. \qquad l_a + \delta_f \le \ \pi_e - \pi_f + p_a T \le u_a - \delta_e, & \forall\ a = (e,f) \in A \setminus A_H, & (1)\\
l_a + \delta_f - (1 - h_a)M \le \pi_e - \pi_f + p_a T & \forall\ a = (e,f) \in A_H, & (2)\\
\le u_a - \delta_e + (1 - h_a)M, & & \\
\sum_{t \in T_{i(e)}} tc_{et} = 1, & \forall\ e \in E, & (3)\\
tc_{arr_{li}t} = tc_{dep_{li}t}, & \forall\ l \in L, i \in l,\ t \in T_i, & (4)\\
h_a \ge tc_{et} + tc_{ft} - 1, & \forall\ a = (e,f) \in A_H, & (5)\\
 & t \in T_{i(e)} & \\
tc_{et}, h_a \in \{0,1\}, \pi_e \in [0,T), p_a \in \mathbb{Z}, \delta_e \ge 0, & \forall\ e \in E, t \in T_{i(e)}, a \in A, &
\end{array}$$

where M is a big enough natural number.

In (1) the normal FPESP constraints are summarized (without headway arcs). In (2) are the headway constraints, which can be switched off with a big-M technique. The assignment of the events to the tracks is done in (3). (4) is used to assign the corresponding arrival and departure events to the same track. In (5) the headway variable is set to 1, if the events take place on the same track, i.e. the headway is required at this operation point.

There are many different objective functions $f(\pi, \delta)$ suggested by Caimi et al. (2011b) for the FPESP model. To generate the traffic plan for our test scenario we use iteratively the TCFPESP with different objective functions (see Wüst et al (2018b)), namely:

- We minimize all passenger relevant times (i.e. $t \in A_T$ the set of trip arcs, $d \in A_D$ the set of dwell arcs and $c \in A_C$ the set of connections times). The weights w_t, w_d and w_c can be used for prioritizing certain times, e.g. connection times. We will call the model in this case MINTRAVEL, according to Caimi et al. (2011b). The objective function is defined as follows:

$$\min f_{TT}(\pi) = \sum_{t \in A_T} w_t \pi_t + \sum_{d \in A_D} w_d \pi_d + \sum_{c \in A_C} w_c \pi_c \tag{6}$$

- We maximize the flexibility in a certain range at certain arrival and departure events. The objective function is defined as follows:

$$\max f_{flex}(\delta) = \sum_{e \in V} w_e \, \delta_e, \tag{7}$$

  where $V \subseteq E$ is the set of all events where flexibility is introduced.
  Furthermore we add two constraints. The passenger travel time has to be smaller than $(1+\epsilon)$ times the best possible travel time from the model MINTRAVEL. The flexibility for all events is bounded by a maximal flexibility δ_{max} for a better distribution of the flexibility to all events. The two constraints are given by

$$f_{TT}(\pi) \leq (1+\epsilon) f_{TT}^{*} \quad \text{and} \quad \delta_e \leq \delta_{\max} \; \forall \; e \in E, \tag{8}$$

  where f_{TT}^{*} is the optimal value found for f_{TT} in (6).
  We will call the model in this case CONTRAVEL according to Caimi et al. (2011b). ϵ is a parameter controlling the quality of the schedule for the passengers' travel times and w_e will be used for individual adjustments in event flexibility in order to maximize timetable robustness (see section 2.3 and 2.4).

By using the models MINTRAVEL and CONTRAVEL iteratively we can generate a traffic plan covering stability and travelling time aspects (see Wüst et al (2018b)).

2.3 Computation of the Cumulative Delay Impact

The Cumulative Delay Impact (CDI) is a measure to quantify the overall impact that a certain delay κ at a specific event f has on all other events e. Formally the CDI is computed as follows:

$$CDI_f(\mathrm{R}) = \sum_{e \in E \setminus f} \max\left(\kappa - r_{ef}, 0\right)^{\gamma}, \tag{9}$$

where E denotes the set of all events. R represents the recovery matrix of size $|E| \times |E|$, where r_{ef} represents the actual buffer time between events f and e (see Goverde (2005) and (2007) for details) given a periodic timetable π. κ is the parameter that denotes the initial delay (in minutes) applied to node f, for which CDI_f shall be calculated. Finally $\gamma \geq 1$ is a parameter to increase the impact of positive differences between the delay κ and r_{ef}. In this study γ was always set to 1. Furthermore, CDI_f is strictly monotonically increasing in κ and $CDI_f(\mathrm{R}) = 0$ for $\kappa = 0, \gamma \geq 1$. The initial delay κ can of course be set for each event $f \in E$ individually. E.g. when κ is determined with the help of a statistical delay analysis for each event $f \in E$.

2.4 Heuristic for improvement of delay impact

We measure the robustness of a periodic timetable π by the sum of all cumulative delay impacts, i.e. we consider $f_{rob}(\pi) = \sum_{f \in E} CDI_f(\mathrm{R})$. Given an acceptable κ (from an operational point of view), we would like to have this measure as small as possible. From the definition of CDI it follows, that $f_{rob}(\pi)$ is bounded from below by 0.

It would therefore be natural to use $f_{rob}(\pi)$ in the CONTRAVEL model as objective function. Since we don't have a direct solution approach for this case, we propose the following heuristic.

Iteratively we try to use the weights w_f in the function $f_{flex}(\delta)$ to give more flexibility to the events $f \in E$, where $CDI_f(\mathrm{R})$ is not zero. We will use the following formula to compute w_f:

$$w_f = \left(\frac{CDI_f(\mathrm{R})}{\max_{\mathrm{f \in E}} CDI_f}\right)^{\alpha}, \text{ if } \max_{\mathrm{f \in E}} CDI_f > 0 \text{ and } CDI_f(\mathrm{R}) \geq \sigma;\ w_f = 0 \text{ otherwise} \qquad (10)$$

where $\alpha \geq 1$ is a parameter to increase the impact of relatively large $CDI_f(\mathrm{R})$ and σ is a threshold value, which is set to $\sigma = 0.4$ in Figure 5a and to $\sigma = 0$ in Figure 5b in the two improvement scenarios in section 3.4.

Iteration scheme: Improving delay impact

Input:

- Periodic timetable π computed with the CONTRAVEL model.
- Initial delay κ and parameter γ.

Iteration steps:

1. Compute $CDI_f(\mathrm{R})$ for all $f \in E$ and $f_{rob}(\pi)$.
   If $f_{rob}(\pi) = 0$, **stop and accept timetable $\boldsymbol{\pi}$.**
2. For timetable π set the weights w_f according to (10).
3. Recompute a new timetable π_{new} with the help of the CONTRAVEL model.
4. Compute $f_{rob}(\pi_{new})$ for the new timetable π_{new}.
   If $f_{rob}(\pi_{new}) = 0$, **stop and accept timetable $\boldsymbol{\pi_{new}}$.**
5. If $f_{rob}(\pi_{new})$ is smaller than $f_{rob}(\pi)$, set $\pi = \pi_{new}$ and go to step 2.
   If $f_{rob}(\pi_{new})$ is bigger or equal than $f_{rob}(\pi)$, **stop and accept timetable $\boldsymbol{\pi}$.**

All timetables during the iterations fulfil the same service intention (see section 1.2), but the resulting timetable is the most robust one with respect to the cumulative delay impact measure (among the constructed timetables during iterations). We illustrate this iteration scheme in our case study in section 3.

3 Case study 'Kerenzerberg'

In order to illustrate the iterative improvement of timetable stability for IP scenarios, we selected a railway corridor in the eastern part of Switzerland. We call the case study 'Kerenzerberg' and the maintenance work is planned on the network section between Flums and Mels. The impact on the schedule is that there is a reduced velocity on that section during normal operation hours.

3.1 Network segmentation

In order to avoid putting too much effort into entering information that is not needed and rather focus on the relevant perimeter for the IP timetabling scenario, one has to identify which part of the entire railway network has to be investigated and which part will be assumed to remain as given by the ordinary timetable. In a first step, the relevant lines and services operating on the subnetwork, which will be affected by the construction sites, have to be identified. In a second step, those lines, which are coupled (e.g. by transfers or technical dependencies) to these affected lines have to be found.

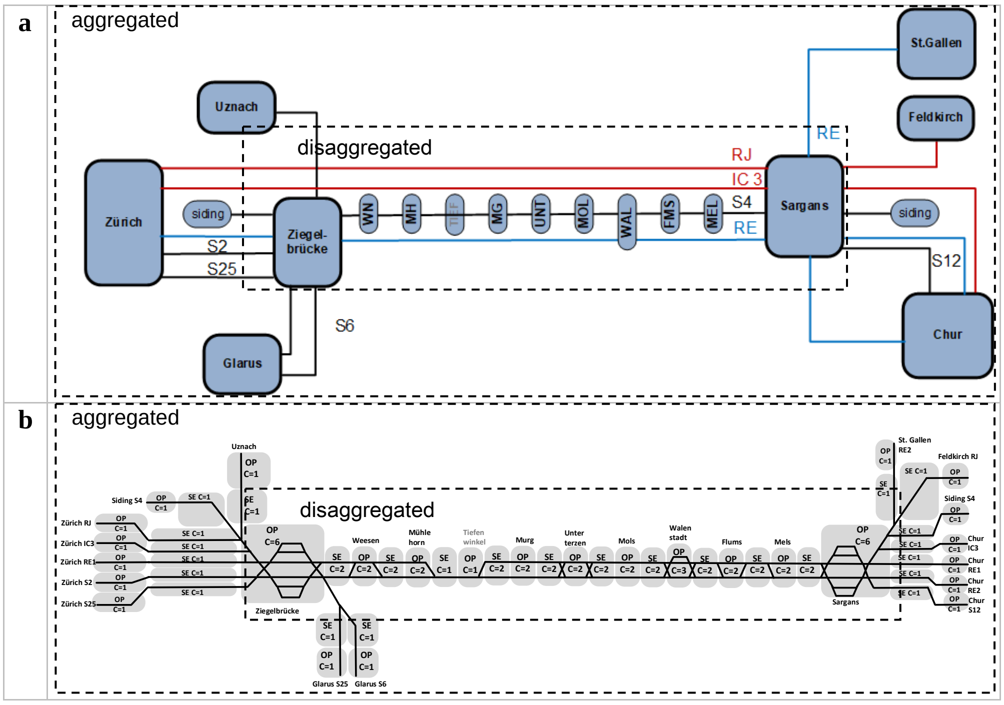


Figure 2: Case study Kerenzerberg a) In order to divide the relevant infrastructure for the IP timetabling scenario into a network partition with the relevant level of detail and a peripheral part with more coarse information, the railway network is divided into subnetworks. A disaggregated subnetwork containing the relevant infrastructure segments at mesoscopic level and an aggregated subnetwork, representing infrastructure on the macroscopic level. b) Shows the track topology for the both, the aggregated and disaggregated network partitions. The grey shaded topology points are indicated with it's type (operation point 'OP', or section point 'SE') and it's number of tracks (track capacity 'C'). In order to avoid treating line interactions outside the disaggregated partition, each line has an individual peripheral OP and the section between the final destination OP and the boundary OP that separates the two partitions from each other is configured with aggregated running times and dwell times of the respective line.

One has to identify the sub-network nodes which isolate the relevant infrastructure partitions from the fixed periphery. In this way one obtains a disaggregated subnetwork containing the relevant infrastructure segments and an aggregated subnetwork, representing infrastructure on the macroscopic level (see outer dashed square areas in

Figure 2 a and b). The disaggregated subnetwork is configured with all mesoscopic details. On this disaggregated subnetwork all train movements are planned in detail. For each line coming from or going beyond the boundary nodes of the disaggregated subnetwork we create a virtual end station node which is connected by a single section to the corresponding boundary node. The section lengths with the appropriate trip times, the turnaround times of the line outside the disaggregated subnetwork together with the run- and dwell times within the disaggregated subnetwork have to sum up to the proper roundtrip time. The mesoscopic track topology of the disaggregated subnetwork is illustrated in Figure 2 b).

3.2 Network of the case study Kerenzerberg.

The planned construction or maintenance work for our test scenario ‘Kerenzerberg’ is located on the network section between Flums and Mels. During the IP interval, trains are running with reduced speed in both directions. We decided to use the corridor Ziegelbrücke-Sargans as the disaggregated partition of the test network. It has to be mentioned, that there is a single track section between the operation points ‘Mühlehorn’ and ‘Tiefenwinkel’. For this disaggregated network partition we iteratively generate IP timetable scenarios (see section 3.4). The western part of Ziegelbrücke is aggregated, i.e. we introduced the nodes Uznach, Zürich, Glarus and a siding of Ziegelbrücke and connecting tracks. The aggregated network will be used to maintain vehicle circulation (e.g. turnarounds) aspects of lines and to model connections to peripheric lines (see the description of SI in section 3.3). The eastern part of Sargans also belongs to the aggregated partition. We introduced the nodes St.Gallen, Feldkirch, Chur and a siding of Sargans. In the aggregated network we assume to have enough track capacity to compensate for delays.

3.3 Description of Service Intention

The configuration of the SI is mainly done in the planning system Viriato. Additional information like upper boundaries of time intervals and flexibility of event times as required in the TCFPESP model is maintained in an R-based table editor (see chapter 2.2). The SI-lines represent the lines in the corresponding timetable 2018 with minor adaptations. In order to demonstrate the turnaround operations within our test scenario, we decided that the line S4 makes a turnaround in a siding next to Ziegelbrücke and Sargans, respectively. Minimal line rotation times and line frequencies are indicated in Table 1.

Table 1: Line rotations and line frequencies

Line ID	**Min. Line rotation time (modulo 60, min)**	**Line frequency (repetition / hour)**
S4	56.6	1
RJ	45.8	0.5
IC 3	43.8	1
RE 1	52.8	1
S12	9	1
S25	14	1
S6	14	1
RE 2	14	1
S 2	9	1

Table 1: Line rotations and frequencies. The minimum line rotation times are computed according to the approach of Liebchen and Möhring (2007). The corresponding turnaround intervals are computed in such a way, that a service with a minimal number of rolling stock is possible. In our case study the line S4 is operating with one rolling stock. The other lines operate with more than one rolling stock due to longer round-trip times. These bounds are not computed according to Liebchen and Möhring (2007), they are set manually.

Ziegelbrücke and Sargans are considered as local hubs. At these stations the traffic plan has to account for passenger transfers between lines. Technically, these transfer requirements result in connections constraints in our TCFPESP-model. These line connections are indicated in Table 2. For a detailed definition of the infrastructure and the SI specification including time intervals of running times, dwell times, turnaround times, separation times etc. see Wüst et al. (2018b).

Table 2: Line Connections at Stations

Connection [1, 15] From/To at station	**S 25 (ZB-GL)**	**S 25 (GL-ZB)**	**S 4 (ZGB-SA)**	**S 4 (SA-ZGB)**	**S 12 (SA-CH)**	**RE 2 (CH-SG)**	**IC 3 (ZGB-SA)**	**RE 1 (SA-ZGB)**	**RE 1 (ZGB-SA)**	**S 6 (GL-UZ)**	**S 6 (UZ-GL)**
S 4 (ZGB-SA)	ZGB	ZGB			SA						
S 4 (SA-ZGB)		ZGB			SA						
S 25 (GL-ZB)			ZGB								
S 25 (ZB-GL)				ZGB							
IC 3 (ZGB-SA)						SA					
S 12 (CH-SA)							SA				
RE 2 (CH-SG)								SA			
RE 2 (SG-CH)							SA	SA			
RE 1 (ZGB-SA)											ZGB
RE 1 (SA-ZGB)										ZGB	ZGB
S 6 (GL-UZ)								ZBG	ZGB		
S 6 (UZ-GL)								ZGB			

Table 2: Case study Kerenzerberg: Line connections at stations are dependent on the direction of the train runs. The time intervals for connection arcs [lb, ub] is configured identically for all connections: [1 min, 15 min].

3.4 Iterative improvement of timetable robustness

Once the configuration of the SI and the mesoscopic infrastructure is complete it is transformed into the TCFPESP model which was implemented in GAMS by applying the the CONTRAVEL model as indicated by equations (7) and (8) with parameter $\epsilon = 0.5$. In case the SI is feasible with respect to the capacity constraints given by the infrastructure, GAMS returns the timetable π_{it_1} with capacity bands. These are plotted as time-distance diagram as shown in Figure 3. This represents the result of the first iteration it_1 of the timetable event flexibility adjustment. As can be seen in the diagram, the capacity time bands of the train runs are quite narrow which is due to a small $\delta_{max} = 10s$, but show variable width within a certain range up to 10 seconds. They are quite homogenously

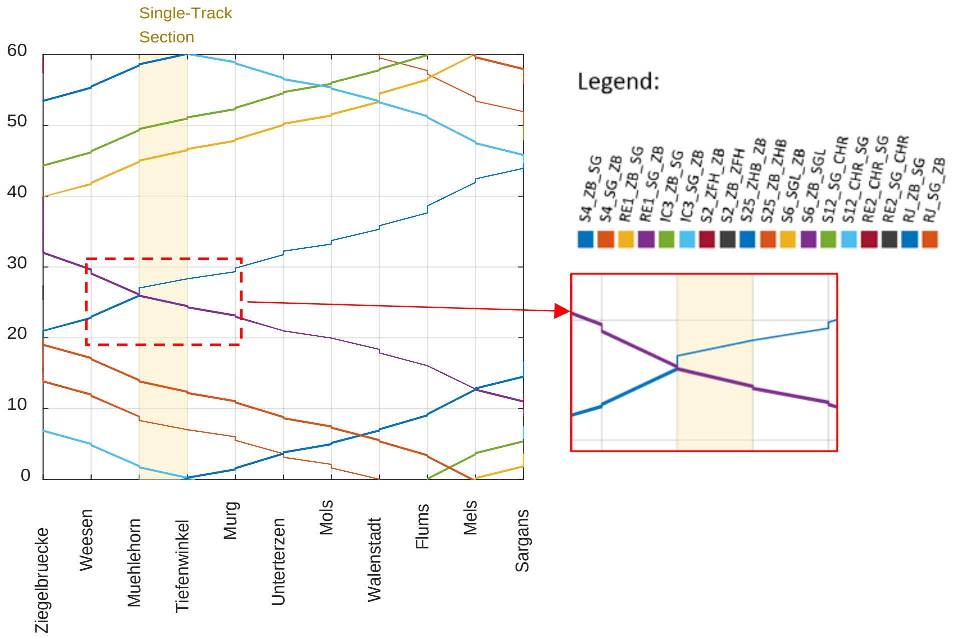


Figure 3: Time-distance diagram: GAMS output for TCFPESP applying the iteration scheme of section 2.4. Line names and directions are indicated by colours as shown in the legend. One can also see the narrow but variable width of the capacity time bands indicating a low flexibility of each train run in a range below one minute.

distributed, indicating some, but low flexibility in all timetable events. The robustness of this result is assessed by calculating the $CDI_f(\mathrm{R})$ for an initial delay κ of 3 minutes. Figure 4a shows the delay impact of each timetable event to all other network events indicated by the corresponding colour (dark colours indicate higher delay impacts) together with the inter dependencies (connecting arrows) in the event activity network. In order to demonstrate the influence a target oriented adjustment of the event flexibility we selected all delay impacts above a threshold of $\sigma = 0.4$ and used them to calculate new weights w_f for the next iteration for calculating a more robust timetable π_{it_2} (see equation (10) and iteration scheme in section 2.4). This time the parameter δ_{max} is set to 60s in order to assign more flexibility to the critical events. The selected weights are shown at the right in Figure 4 for nodes with numbers above 150. Here node numbers are sorted with higher node numbers for high delay impacts. The time-distance diagram of the resulting timetable $\pi_{it_2}$with $\sigma = 0.4$ is shown in Figure 5a. One can clearly see that here certain timetable events have much more flexibility than others. If we sum up the delay impacts of all events in this scenario, we obtain a $f_{rob}(\pi)$ $-$value reduced to 87.0% compared to the one of π_{it_1} (see Figure 5d, result ‘with few weights’).

Figure 5b shows the time-distance diagram for the result of timetable iteration $\pi_{it_2}$with $\sigma = 0$. This means that in this iteration we selected all weights. In this scenario, the effect on the cumulative delay impact is even stronger. The resulting $f_{rob}(\pi)$ $-$value is reduced to 79.3% compared to the one of π_{it_1} (see Figure 5d, result ‘with all weights’).

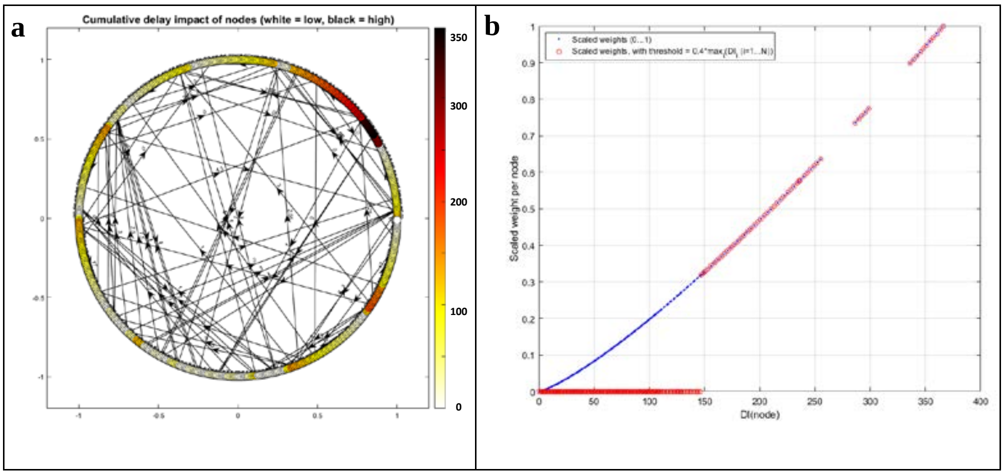
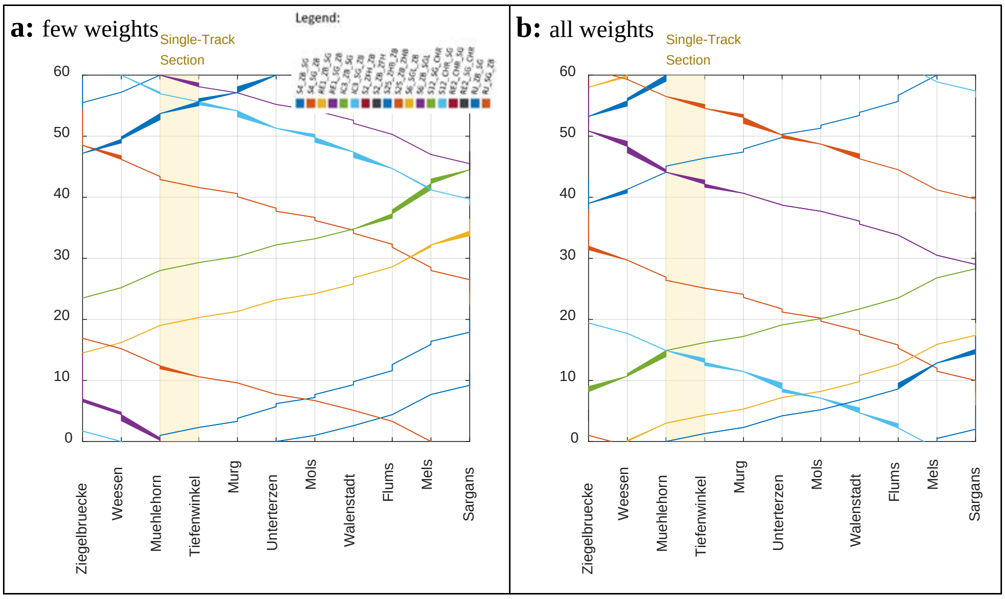
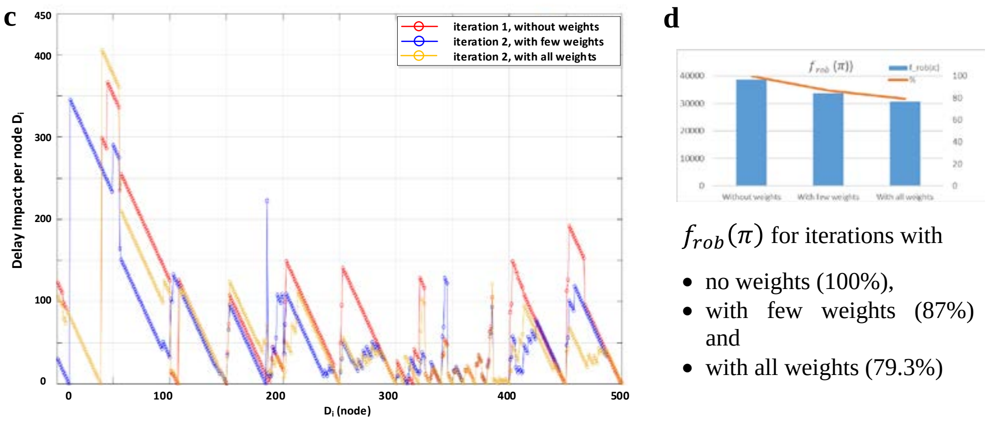


Figure 4: a) The values of the $CDI_f(R)$ for all event nodes indicated by a colour code ranging from 0 to 350 seconds and the interdependencies between the event nodes. b) shows the weights (calculated according to (10)) scaled to values in the range [0…1] for all event nodes sorted from left (low values) to right (high values). One can see that some event nodes (all events of the Railjet RJ) do not have any delay impact, as they do not have influence on other events.





$f_{rob}(\pi)$ for iterations with

- no weights (100%),
- with few weights (87%) and
- with all weights (79.3%)

Figure 5: a) and b) show time-distance diagrams for the second iteration of the timetable calculation π_{it_2} with a selection of weights w_f (a: σ = 0.4) and with all weights (b: σ = 0). Line colours are the same as in Figure 3. C) shows the distribution of the delay impact values across the event nodes f for timetable iteration it_1 with no weights applied (red curve), and for timetable iteration it_2 with few selected weigths (blue curve, see text for selection criteria) and for timetable iteration it_2 with all weigths applied (orange curve). d) indicates the improvement of iteration it_2 with respect to the iteration it_1 for both weighting scenarios.

4 Conclusions

The aim of this research was to show, that using the service intention as the demand oriented functional requirement for timetable generation one can generate timetable scenarios for maintenance interval planning in a fast and easy way. An additional requirement for generating reliable timetable scenarios is the usage of a mesoscopic infrastructure model for input to FPESP. Temporary changes of infrastructure properties like the number or the maximum allowed speed of tracks and switches reduce the available capacity for track assignments to train runs. For this reason, we introduce an extension of the FPESP model that we call 'TCFPESP' model. The TCFPESP model allows to make a target oriented adjustment of event flexibility by applying weights to the TCFPESP objective function. We obtain those weights from the calculation of the cumulative delay impacts for all timetable events and use them in an iterative manner for improving timetable stability.

We show results for a few scenarios which demonstrate that we can reduce the overall delay impact of timetable events by a significant amount (a reduction of more than 20% in the second iteration compared to the first iteration). We consider these preliminary results as promising for making target oriented improvements of timetable robustness, especially in cases where variability of process times is high and cannot be reduced by operational measures. Timetable events that have a strong influence on many other timetable events should be planned with more flexibility than those with low cumulated impact. In a next step we want to further investigate this observation with the help of simulations on microscopic level.

Acknowledgements

This research has been funded by the SBB-Research Fund and was part of the SmartRail 4.0 Development Initiative in order to introduce new railway technologies and improve the efficiency of a competitive Swiss Public Transport. We are very grateful also to IBM Research, who contributed with financial support to M. Laumanns and J. Szabo to develop the concept for the TCPESP-model. The technical implementation of the TCPESP-model and the CDI_f(R) as well as the execution of this case study was done by IDP in collaboration with the SBB project team of SR4.0-TMS-PAS. Here we want to thank especially Thomas Wieland and Thomas Künzi for many inspiring discussions.

References

Bešinović, N., Goverde, R.M.P., Quaglietta, E., Roberti, R., 2016. *An integrated micro-macro approach to robust railway timetabling.* Transportation Research Part B: Methodological, vol. 87, pp. 14-32.

Caimi, G.,2009. *Algorithmic decision support for train scheduling in a large and highly utilised railway network*. Diss. ETH Zürich Nr. 18581

Caimi, G., Fuchsberger, M., Laumanns, M., Schüpbach, K., 2011. “Periodic railway timetabling with event flexibility” *Networks*, vol. 57, issue 1, pp. 3-18

deFabris, S., Longo, G., Medeossi, G., Pesenti, R., 2014. “Automatic generation of railway timetables based on a mesoscopic infrastructure model” *Journal of Rail Transport Planning & Management*, vol. 4, pp. 2–13

Friedrich, M., Hartl, M., Schiewe, A. and Schöbel, A., 2017. *Integrating passengers' assignment in cost-optimal line planning.* Technical Report 2017-5, Preprint-Reihe, INAM, Georg-August Universität Göttingen

GAMS 2018. General Algebraic Modeling System GAMS, GAMS Software GmbH, Frechen, Germany, https://www.gams.com/ . last accessed: 19 September 2018

Goverde, R.M.P., 2005. *Punctuality of Railway Operations and Timetable Stability Analysis.* TRAIL Thesis Series no. T2005/10. Delft University of Technology, Delft, The Netherlands

Goverde, R.M.P., 2007. “Railway timetable stability analysis using max-plus system theory” *Transportation Research Part B*, vol. 41, no. 2, pp. 179-201

Hansen, I.A., Pachl., J. (eds.), 2008. *Railway Timetable & Traffic. Analysis – Modelling – Simulation.* Eurail Press, Hamburg, Germany

Herrigel, S., 2015. *Algorithmic decision support for the construction of periodic railway timetables*, Diss. ETH Zürich Nr. 22548, Department Bau, Umwelt und Geomatik, ETH Zürich, Switzerland. https://doi.org/10.3929/ethz-a-010412035

Liebchen, C., Möhring, R.H., 2007. “The modeling power of the periodic event scheduling problem: Railway timetables – and beyond” In: Geraets, F., Kroon, L., Schoebel, A., Wagner, D., Zaroliagis, C. (Eds.), *Algorithmic Methods for Railway Optimization*, Lecture Notes in Computer Science, vol. 4359, pp. 3–40, Springer, Berlin/Heidelberg, Germany

RailNetEurope, 2017. Glossary of Terms Related to Network Statements [online]. RailNetEurope.http://www.rne.eu/rneinhalt/uploads/RNE_NetworkStatementGlossary__V8_2016_web.pdf [Accessed 22 Jan. 2019].

Serafini, P., Ukovich, W., 1989. “A mathematical model for periodic scheduling problems”, *SIAM Journal on Discrete Mathematics*, vol. 2, pp. 550-581. http://dx.doi.org/10.1137/0402049

SMA, 2018. *Viriato - software for railways.* http://www.sma-partner.ch, Zurich. last accessed 2018/07/01

Van Aken, S., Bešinović, N., Goverde, R.M.P., 2017a. “Designing alternative railway timetables under infrastructure maintenance possessions” *Transportation Research Part B* vol. 98, pp. 224–238

Van Aken, S., Bešinović, N., Goverde, R.M.P., 2017b. “Solving large-scale train timetable adjustment problems under infrastructure maintenance possessions” *Journal of Rail Transport Planning & Management* vol. 7, pp. 141-156

Wüst, R. M., Laube, F., Roos, S., Caimi, G., 2008. *Sustainable Global Service Intention as objective for Controlling Railway Network Operations in Real Time*, In: Proceedings of the WCRR 2008. Seoul

Wüst, R. M., Bütikofer, S., Ess, S., Gomez, C, Steiner A., Laumanns, M., Szabo, J., 2018a. *Periodic timetabling with ‘Track Choice’-PESP based on given line concepts and mesoscopic infrastructure* In: Operations Research Proceedings 2018, In press, Springer, Berlin/Heidelberg, Germany

Wüst, R. M., Bütikofer, S., Ess, S., Gomez, C, Steiner A., 2018b. *Development of a prototype for the automated generation of timetable scenarios specified by the transport service intention.* Research Report of SBB Research Fund St.Gallen http://www.hsg

Maintenance timetable planning based on mesoscopic infrastructure and the transport service intention

Raimond Wüst[1], Stephan Bütikofer[1], Severin Ess[1], Claudio Gomez[1], Albert Steiner[1],
Marco Laumanns[2], Jacint Szabo[3]

[1] Institute of Data Analysis and Process Design, Zurich University of Applied Sciences ZHAW,
8401 Winterthur, Switzerland, Switzerland, `wura@zhaw.ch`
[2] Bestmile SA, 1007 Lausanne, Switzerland, `marco.laumanns@bestmile.com`
[3]IBM Research, now at Google Switzerland, Brandschenkestrasse 110, Zurich, Switzerland
`szabojacint@gmail.com`

Abstract. Planners of maintenance intervals and operations have a strong need for rapid development and assessment of comprehensive and reliable timetable scenarios, which are able to satisfy the requirements of both, the train operating company and the infrastructure operating company. In this work, we present a use case that enables the integration of long-term and short-term process steps in the transport service value chain. It is based on functions executed by human actors and system components for computer aided timetable generation. The use case takes the central planning object ‘service intention’ as input and generates a feasible timetable scenario as output. The service intention is a data structure representing a line concept that takes into account passenger transfers for efficient transport chains. It supports an iterative timetable development relying on a ‘progressive feasibility assessment’, a feature that is requested in practice.

Our use case for generating maintenance-interval-timetable scenarios is part of an application concept that was developed together with SBB Infrastruktur* and is based on the ‘track-choice’ and line rotation extension of the commonly known method for the generation of periodic event schedules ‘PESP’. The extension takes into account event flexibility requirements of the service intention and makes use of a mesoscopic track infrastructure. Both, the service intention criteria as well as the mesoscopic infrastructure representation can be configured in the line planning and timetabling system Viriato. The level of detail of the considered data in Viriato is well suited for specifying the input of our timetabling model. This system is widely used by public transport planners and operators. It is therefore possible to control our timetable model by a standard planning tool from industry.

Taking into consideration the technical and operational constraints given by rolling stock, station and track topology data on the one hand, and the commercial requirements defined by a given line concept on the other, the method presented generates periodic timetables, including train-track assignments. The data structure ‘service intention’ represents the line concept consisting of train paths, frequencies and connections. Due to the utilization of infrastructure-

---

* Infrastructure department of Swiss National Railways (Schweizerische Bundesbahnen, SBB)

based track capacities, we are also able to assess the feasibility of the line concept given on a mesoscopic level. Additionally, we can handle temporary resource restrictions, e.g. caused by construction sites or operational disturbances. After the description of the methods developed, we provide a practical proof of concept by testing the use case for different maintenance scenarios. Thereby we can show that, i) based on the service intention planners are able to quickly develop feasible timetable scenarios for maintenance intervals, and that ii) in case no feasible timetables can be found a new iteration in the timetable development process can be initiated by changing the functional requirements for the planning process.

The use case presented in this paper refers to generating short-term timetable scenarios is part of a planning framework, which is currently developed together with SBB and which also contains use cases for long-term process steps.

Keywords: Periodic Event Scheduling Problem, Mesoscopic railway topology, Service Intention, Timetabling with track assignment

Declarations of interest: none

1 Introduction

1.1 State-of-the-art

In the operational management of railway networks, an important requirement is the fast adaptation of timetable scenarios, in which operational disruptions or time windows with temporary unavailability of infrastructure, for instance during maintenance time windows, are taken into consideration. In those situations, easy and fast reconfiguration and recalculation of timetable data is of central importance. This local and temporal rescheduling results in shifted departure and arrival times and sometimes even in modified stop patterns at intermediate stations of train runs. We will refer to this scheduling process as interval planning (IP). In order to generate reliable timetabling results, it is a prerequisite that train-track assignments as well as operational and commercial dependencies are taken into consideration and that all these aspects are not conflicting with each other. Hence, finding the right level of detail for modelling track infrastructure and train dynamics is crucial for supporting the planning process in an optimal way. In recent years, this requirement motivated several research groups to combine common timetabling procedures with constraints resulting from mesoscopic infrastructure information.

Some of the most important approaches that are relevant to our work are discussed in the following. Hansen and Pachl (2008) show how running, dwell and headway times at critical route nodes and platform tracks must be taken into account for train processing and present an in-depth timetable quality analysis depending on these parameters. De Fabris et al. (2014) calculate arrival and departure time, the platform and the route within stations and on junctions that trains visit along their lines. Bešinović et al. (2016) present a micro–macro framework based on an integrated iterative approach for computing a microscopically conflict-free timetable that uses a macroscopic optimization model with a post-processing robustness evaluation. Caimi et al. (2011b) extend PESP (see, e.g. Liebchen and Möhring (2007) for a detailed introduction) by proposing the flexible periodic event scheduling problem (FPESP), where intervals are generated instead of fixed event times. By applying FPESP, the output does not define a final timetable but an input for finding a feasible timetable on a microscopic level (Caimi (2009) and Caimi et al. (2011a)).

1.2 Research goals

In this article we present solutions for two research goals: (i) We develop a method for the timetable generation based on mesoscopic infrastructure and the transport service intention (SI, see section 2.1); (ii) From the future SBB timetable planning process we derive a use case for interval planning, which utilizes the method mentioned in (i) for timetable generation.

The use case includes an iteration scheme for applying the timetable method with slightly changing objectives for timetable generation. The necessary data structures can be managed in a standard timetable editor. Finally, the use case is applied in a practical case study.

Similar to Caimi et al. (2011b), our modelling approach for the timetable generation is also based on an extension of the periodic event scheduling problem (PESP) and takes the SI as input data structure. The SI was first described in Wüst et al. (2008), formally specified in Caimi (2009) and integrates commercial timetabling requirements given by the respective line concept on the one side and technical constraints on the other. It largely corresponds to the 'line concept', and represents functional timetabling requirements including line data, line frequencies and separations. In order to preserve acquired knowledge about customer flows for the subsequent planning step, customer transfers between lines at specific stations are also included in the SI. In accordance to de Fabris et al. (2014), we call our level of abstraction of the available resources 'mesoscopic topology'. Along with the functional requirements of the SI, this mesoscopic infrastructure data model of a given scenario is entered via a standard timetable editor (see, e.g., SMA Viriato (2018)). We demonstrate the detailed sequence of the planning actions that planners must execute in order to generate timetable results for different scenarios of maintenance intervals for the interval planning (IP) use case.

1.3 Structure of this document

This article is structured as follows: In chapter 2, we describe the methodology for achieving the research goals mentioned. In section 2.1 (IP business requirements) we provide a short description of the special business requirements of interval planning, which are based on the future SBB process model for timetable generation. This process model assumes a close cooperation between the different Train Operating Companies (TOC) and the Infrastructure Operating Company (IOC), including (to a certain degree) a barrier free data access for the IOC. This improved data transparency is needed to insure a specified service level that holds also in case of IP or operational disruptions. Based on these requirements, in section 2.2 (based on mesoscopic infrastructure and the transport service intention) we propose an IP use case for computer aided interval planning. In sections 2.3 (Network segmentation), 2.4 (Method for generating traffic plan with flexibility in IP) and 2.5 (Generation of traffic plan based on a standard planning tool) we describe how data that are required for input in the proposed use case can be handled in a standard planning tool, like Viriato.

In chapter 3 (Case study ‘Kerenzerberg’) we present the results of applying the methods introduced in chapter 2 as well as their interdependencies in a real-world test scenario. The chapter contains a detailed description of the actions belonging to the proposed use case and to the application of the proposed method.

In chapter 4 we conclude with a summary of the encouraging results of the presented case study. We consider them as a proof of concept for our proposed use case. Finally, we provide a brief outlook on future work.

2 Methodology

One of the most important requirements for public transport services is its usability compared with competing transport modes. Two factors have a significant impact on the usability. The first one is the aspect of the regularity or periodicity of a timetable, which allows travellers for easily remembering departure and arrival times and hence making travel planning much simpler, especially for regular travellers. The second aspect deals with the integrated transport chain. The transport chain is characterized by changing transport modes between local (de-)feeding lines like bus or streetcar lines and high-performance train lines with higher speeds and capacities. Of course, this includes the change between different line types (e.g. far distance line, commuter line) or lines as well. Here the realization of short connection times is the main objective. The set of relevant transport chains can be obtained by combining origin-destination-demand matrices with potential line pools in the line planning process step (see Figure 1, ‘line planning, line concept’, Schöbel and Scholl (2006) and Friedrich et al. (2017)). The integrated information regarding lines and transport chains represents the intended transport service and is called service intention (SI). Technically, both usability aspects can be realized by introducing a countrywide integrated fixed interval timetable (IFIT) (see for example Herrigel (2015) for an explanation of the fundamental idea), which synchronizes the service schedules of almost all carriers. The integration of services of different TOCs results in a highly iterative process of timetable generation. This process will be explained in the following section.

2.1 IP business requirements

At the beginning of the collaboration project with our industry partner SBB-Infrastructure, business analysts reviewed the planning process resulting in a description of the streamlined future timetable generation process which is summarized in Figure 1.

The desired transport service is based on requests from various stakeholders and is consolidated by the TOCs. This functional description is represented by the SI. The functional requirements represent scenarios of transport chains, which have been consolidated before. In the first set of process steps, the SI is defined by the involved TOCs and translated into a capacity requirement, mapped onto railway lines and stations. It can be visualized in terms of net graphs, line diagrams and passenger-flow tables (see TOC 'Passenger and Freight Assignment' and 'line planning, line concept' in Figure 1). In the second set of process steps, the SI's of the different TOCs and the capacity requirements of the different train lines must be consolidated and checked for operational feasibility. This is done by the integrating IOC. SBB calls these process steps 'traffic planning' and 'capacity planning' (see Figure 1). The result of process step 'traffic planning' corresponds to the consolidated SI. In the scope of this paper we will call the consolidated SI 'commercial timetable'. The commercial timetable is the basis for the communication of the timetable to the customer. Additionally, the consolidated SI requires a feasibility check based on the following process step 'capacity planning'. The result of this process step is a validated version of the service intention, which accounts for capacity constraints defined by track occupation, headway, transfer and line rotation time requirements. In addition, constraints resulting from maintenance and construction requirements are accounted for. All these aspects of capacity consumption are integrated in the capacity plan, which we will call 'timetable with event flexibility'.

As one can see in Figure 1, in cases of reduced capacity (compared to resource conditions of the standard timetable), the IOC has the responsibility of providing the best service quality possible. That means that in the use case of interval planning or operational disruption, the IOC must have access to demand and service specific data (managed by the TOCs) which determine the input for the process of generating a consolidated SI.

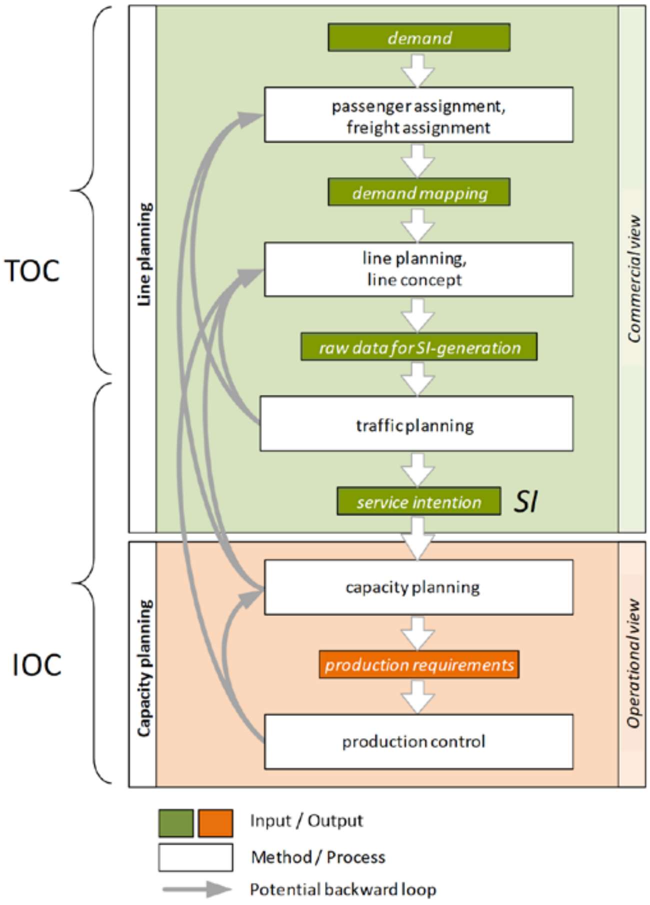


Figure 1. The SBB timetable planning process: overview and integration of line planning into the timetable planning model. The grey arrows in the figure demonstrate, that when going back in the in the process chain, data referring to the operation of train lines and passenger demand have to be shared by TOC and IOC. Further explanations are given in the text. Adaptation from Howald et al. (2017)

Figure 1, also shows that the planning process is an iterative one, indicated by the grey backward loop arrows. During this process iteration, the level of detail of the resulting plan is progressive because under conditions of long- and medium-term timetable planning, knowledge and decisions regarding functional (commercial) and non-functional (technical and operational) timetabling requirements are getting more concrete. In case of reduced resource availability, resulting from short term operational disruptions or maintenance work, this may cause timetable conflicts that can only be resolved by relaxing the functional timetabling requirements. Hence, the SI has to be adapted temporarily.

With the methods that will be presented in the following sections, we attempt to operationalize the SI based planning step (Figure 1: ‘traffic planning’) in terms of the

IP use case and the proposed method for computer aided interval planning based on mesoscopic infrastructure and the transport service intention.

2.2 Method for computer aided interval planning

The generation and investigation of feasible event times for individual train runs and corresponding resource allocations fitting into the structure of an IFIT is usually done manually. For this reason, timetabling is considered a time consuming and challenging task even for experienced planners. On the other side, algorithmic approaches for solving this task computationally require models based on microscopic information about track capacity, like discussed, for instance in Bešinović et al. (2016) or, in an intermediary step, by defining possible train routes as outlined in de Fabris et al. (2014), from which headway constraints for trains can be derived. We present a generic approach, which makes use of the mesoscopic infrastructure, for setting headway constraints and other operational dependencies like turnaround times, and the service intention. The corresponding data are implemented and managed in a standard timetable planning system like Viriato (see, e.g., SMA (2018)).

2.2.1 Mesoscopic infrastructure model

To illustrate the level of detail of the respective infrastructure mapped onto a mesoscopic topology we refer to an example of the SBB “Grobkonzept Linienplanung” in Howald et al. (2017, see Figure 2a). The mesoscopic topology consists of operation points linked by route-sections. At each operation points and route-section there is a given number of tracks. Each location that provides an option to change tracks there is assigned a new operation point. If there are customer services assigned to an operation point, it is classified as ‘commercial’, otherwise it is classified as ‘operational’. In our topology model, we introduce graph nodes for both, operation points and route-sections connecting two operation points. The capacity ‘C’ of each node is defined by the number of enumerated tracks of the operation point (see Figure 2b). The connectivity of the tracks at each node are additional node attributes and can be configured in Viriato (see Figure 7). From this node topology we derive our event activity network representing all potential track specific event dependencies (see section 2.2.2).

Timetables that are mapped to mesoscopic infrastructure enable a much better feasibility assessment of the result compared to considering only the macroscopic infrastructure. On the other hand, the gap in the level of detail to microscopic infrastructure in terms of feasibility assessment can be reduced substantially, if the event times that are assigned to mesoscopic topology nodes are within a certain range of flexibility (see Caimi (2009) and Caimi et al. (2011b)). We make use of the mesoscopic topology together with the event flexibility according to the FPESP model, introduced in section 2.2.3. This allows generating periodic timetables with a reasonably good assessment of feasibility. The model generates results with flexibility to find a conflict free resource allocation taking a micro-topological level of detail into consideration in a subsequent planning step (Figure 1: ‘production control’) or if planning has to account for slightly different individual conditions, for example during the course of a day or to consider the operational variability.

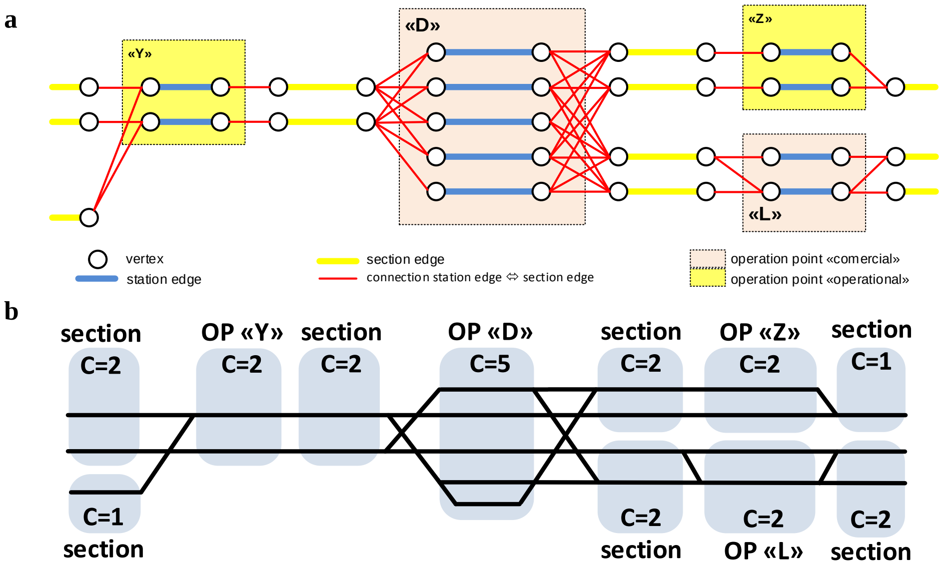


Figure 2: a) Mesoscopic infrastructure example from the SBB “Grobkonzept Linienplanung” in Howald et al. (2017). b). Extracted topology information: Each operation point, and each linking track segment is mapped into a graph node, represented by a grey shaded box. The node attribute ‘C’ indicates the track capacity of each node. Switches between node tracks allow for changing tracks when moving from one node to the other.

2.2.2 Event activity network and periodic timetabling

The event-activity network (EAN) is the input for our timetable model. It is constructed based on mesoscopic infrastructure information and the SI.

The SI is defined by a set of train runs. Each train run belongs to a line L and is characterized by the sequence of sections that are traversed and a corresponding time interval, which is required for either running or stopping on a corresponding track section. Each time interval has a minimal and maximal value. Stop nodes typically provide a service for boarding or de-boarding a train. Together, a pair of train runs moving in opposite directions makes up a train circulation.

The mesoscopic infrastructure consisting of sections is summarized as a set I of operation points. Operation points are largely tracks and stations but can also be other critical resources as junctions (see example below). As mentioned before, each operation point $i \in I$ is associated to a capacity consisting of a set of tracks T_i. A train run $l \in L$ is described by a sequence of operation points of I.

Based on our mesoscopic model we algorithmically create an event-activity network (E, A). The set E of events consists of an arrival event arr_{li} and a departure event dep_{li} for each train run $l \in L$ and operation point $i \in l$. The activities $a \in A$ are directed arcs from $E \times E$ and describe the dependencies between the events. For every train run we have arcs between arrival and departure events at the same operation points (dwell times or trip times) and arcs between departure and arrival events of successive operation points (time needed for the travel between operation points). Further arcs include connections between train runs, headways and turnaround operations (see chapter 3). Connections and turnaround information are given in the SI. Headways are derived from the mesoscopic infrastructure and the train runs in the SI. We refer to Liebchen and Möhring (2007) for a detailed overview of the modelling options of dependencies. Figure 3 provides a sample of such an event-activity network.

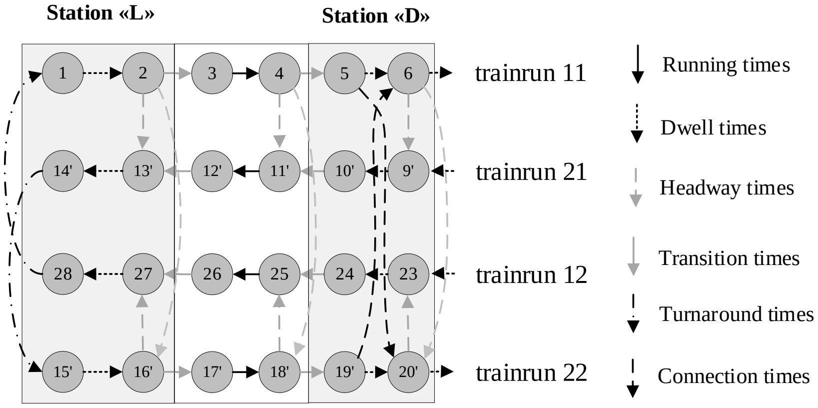


Figure 3: Sample of an event-activity network, where arcs connect arrival and departure events. Nodes belonging to grey shaded boxes indicate events at operation points (here Station «L» and Station «D»). Other nodes indicate track type arrival and departure events. Arrow line styles indicate different types of time dependencies.

Headway arcs $a \in A_H$ are especially important for explaining the timetable model below. Headway arcs are used to model safety distances between trains running in the same and in opposite directions (see example in Figure 3). For the sake of simplicity we consider in the formal description of our timetable model in section 2.2.3 only headways related to one operation point, i.e. we omit headways for train runs in opposite directions over several successive operation points. These headways can be easily included in the event-activity network. They are included in our implementation of the timetable model.

The classical PESP model tries to determine a periodic schedule on the macroscopic level (i.e. without using the tracks at an operation point) within a period T. Event $e \in \mathrm{E}$ takes place at time $\pi_e \in [0, T)$. The schedule is periodic with time period T, hence each event is repeated periodically $\{\dots, \pi_e - T, \pi_e, \pi_e + T, \pi_e + 2T, \dots\}$.

The choices of the event times π_e depend on each other. The dependencies are described by arcs $a = (e, f)$ in A and modeled as constraints in PESP. The constraints always concern the two events e and f and define the minimum and maximum periodic time difference l_a and u_a between them. These bounds are given as parameters in the PESP model. We therefore look for the event times π_e for every $e \in E$ that fulfill all constraints of the form

$$l_a \leq \pi_f - \pi_e + p_a T \leq u_a,$$

for all $a = (e, f) \in A$, where p_a is an integer variable that makes sure, that these constraints are met in a periodic sense.

2.2.3 Periodic Timetabling with Event Flexibility

In order to avoid tedious iterations between the process steps “microscopic capacity planning” and “mesoscopic capacity planning” in case of infeasibility of the micro-level problem, one can improve the chance of finding a feasible solution by enlarging the solution space in the micro-level. This approach has been described in detail in Caimi et al. (2011b). We also implement this event flexibility model by (optionally) adding some flexibility for the events of the EAN by introducing lower and upper bounds to the event times of the arrival and departure nodes in Figure 4. The final choice of the event times in the range between the lower and upper bound shall be independent for each event such that each value of the end of an activity arc should be reachable from each time value at beginning of that activity arc.

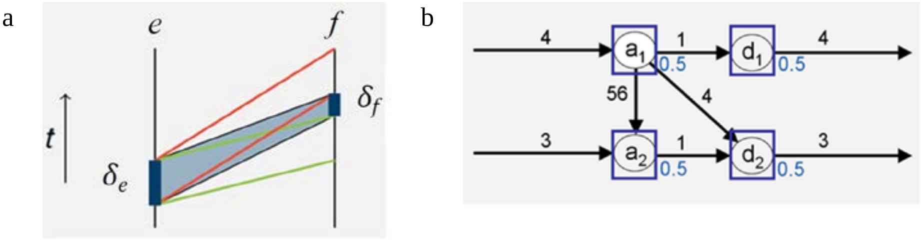


Figure 4: Target oriented placement of time reserves (adapted from Caimi (2011b)): a) Time frames $[\pi_e, \pi_e + \delta_e]$ in place of time points π_e. By implementing this method, the normal PESP constraints $l_a \leq \pi_f - \pi_e + p_a T \leq u_a$ now becomes $l_a + \delta_e \leq \pi_f - \pi_e + p_a T \leq u_a - \delta_f$ (see section 2.2.4). b) In this EAN example this means that instead of planning time points $(\pi_{a_1}, \pi_{d_1}, \pi_{a_2}, \pi_{d_2})$ we plan time frames $[\pi_e, \pi_e + 0.5]$ for $e \in \{a_1, d_1, a_2, d_2\}$.

We are not forced to add this flexibility to all the events, but we can select the nodes where we want to add it based on user defined rules, for instance only nodes corresponding to events in a main station area with high traffic density, where it is more difficult to schedule trains on the microscopic level. In general, one can say, that this placement of flexibility is the timetable configuration feature that has the highest impact on improving operational stability. For our proposed timetabling model, we integrate an extended PESP model (based on mesoscopic infrastructure, see section 2.2.4) with the “flexible PESP” (FPESP) model in order to generate timetables with event slots on a mesoscopic level. For more details regarding the FPESP model, we refer to the article of Caimi et al. (2011b).

2.2.4 Track-choice PESP model with event flexibility.

We extend the classical PESP resp. FPESP model by using the number of tracks T_i at each operation point $i \in I$. We will refer to this extended model as the track-choice FPESP model (TCFPESP). It assigns the arrival event arr_{li} and the departure event dep_{li} of train run l at operation point i uniquely to a track in T_i. We can use these assignments to switch on headway arcs $a \in A_H$ by using a big-M-approach.

In addition to the variables π and p from the PESP model we need:

(i) Binary variables tc_{et} (track choice) for each event $e \in E$ and track $t \in T_{i(e)}$, where operation point $i(e)$ is associated to event e, i.e. e is equal to arr_{li} or dep_{li} for a train run l.

(ii) Binary variables h_a for every headway edge $a = (e,f) \in A_H$. Headway edges are always between events at the same operation point, therefore $T_{i(e)} = T_{i(f)}$ holds.

(iii) Positive variables δ_e for each event $e \in E$ to model the event flexibility.

The TCFPESP model is then defined by:

TCFPESP model:

$$
\begin{array}{llr}
\min f(\pi, p, \delta) & & \\
s.t. \quad l_a + \delta_e \leq \pi_f - \pi_e + p_a T \leq u_a - \delta_f, & \forall\ a = (e,f) \in A \setminus A_H, & (1) \\
l_a + \delta_e - (1 - h_a)M \leq \pi_f - \pi_e + p_a T \leq u_a - \delta_f + (1 - h_a)M, & \forall\ a = (e,f) \in A_H, & (2) \\
\sum_{t \in T_{i(e)}} tc_{et} = 1, & \forall\ e \in E, & (3) \\
tc_{arr_{li}t} = tc_{dep_{li}t}, & \forall\ l \in L, i \in l,\ t \in T_i, & (4) \\
h_a \geq tc_{et} + tc_{ft} - 1, & \forall\ a = (e,f) \in A_H,\ t \in T_{i(e)} & (5) \\
tc_{et}, h_a \in \{0,1\}, \pi_e \in [0,T), p_a \in \mathbb{Z}, \delta_e \geq 0, & \forall\ e \in E, t \in T_{i(e)}, a \in A, &
\end{array}
$$

where M is a big enough natural number.

In (1) the normal FPESP constraints are summarized (without headway arcs). (2) defines the headway constraints, which can be switched off with a big-M technique. The assignment of the events to the tracks is done in (3). (4) is used to assign the corresponding arrival and departure events to the same track. In (5) the headway variable is set to 1, if the events take place on the same track, i.e. the headway is required at this operation point.

There are many different objective functions $f(\pi, p, \delta)$ described in literature (see Liebchen and Möhring (2007) for the general PESP model and Caimi et al. (2011b) for the FPESP model). In our test case below we use two objective functions.

Objective Functions:

- MINTRAVEL: We minimize all passenger relevant times (i.e. $t \in A_T$ the set of trip arcs, $d \in A_D$the set of dwell arcs and $c \in A_C$ the set of connections times). The weights w_t, w_d and w_c can be used for prioritizing certain times, e.g. connection times. The objective function f_{TT} is defined as follows.

$$f_{TT}(\pi) = \sum_{t \in A_T} w_t \pi_t + \sum_{d \in A_D} w_d \pi_d + \sum_{c \in A_C} w_c \pi_c \qquad (6)$$

  According to Caimi et al. (2011b) we will call the TCFPESP model with this objective function MINTRAVEL.

- CONTRAVEL: We maximize the flexibility in a given range at certain arrival and departure events. The objective function f_{flex} is defined as follows:

$$f_{flex}(\delta) = \sum_{e \in V} w_e\, \delta_e, \tag{7}$$

where $V \subseteq E$ is the set of all events where flexibility is introduced. Furthermore, we add two constraints. The passenger travel time has to be smaller than $(1 + \epsilon)$ times the best possible travel time from the model MINTRAVEL. The flexibility for all events is bounded by a maximal flexibility δ_{max} for a better distribution of the flexibility to all events. The two constraints are given by

$$f_{TT}(\pi) \le (1 + \epsilon) f_{TT}^* \quad \text{and} \quad \delta_e \le \delta_{\max}\ \forall\ e \in E, \tag{8}$$

where f_{TT}^* is the optimal value found for f_{TT} in the MINTRAVEL model. We will call the TCFPESP model with the objective function in (1) and the additional constraints in (3) CONTRAVEL according to Caimi et al. (2011b). ϵ is a parameter controlling the quality of the schedule for the passengers' travel times and the weights w_e can be used for individual adjustments in event flexibility to maximize timetable robustness.

Both models MINTRAVEL and CONTRAVEL are mixed integer linear problems. In this paper all the weights in the objective functions of MINTRAVEL and CONTRAVEL are set to 1. We provide further details about the implementation and the size of these models in chapter 3. The TCPESP variant of the model (i.e. without event flexibility) has been presented recently in Wüst et al. (2018).

2.3 Network segmentation

In order to avoid putting too much effort into entering information that is not needed and rather focus on the relevant perimeter for the IP timetabling scenario, one has to identify which part of the entire railway network has to be accounted for. The relevant lines and services operating on the subnetwork which will be affected by the construction or maintenance sites have to be identified in a first step. In a second step, those lines, which are coupled (e.g. by transfers or technical dependencies) to these affected lines have to be found.

In the second step, one has to identify the sub-network nodes which isolate the relevant infrastructure segments from the irrelevant periphery. In this way one obtains a disaggregated subnetwork containing the relevant infrastructure segments and an aggregated subnetwork, representing infrastructure on the macroscopic level (see dashed square area on the top of Figure 5a). The disaggregated subnetwork is configured with all mesoscopic details. On this disaggregated subnetwork all train movements are planned in detail for every single IP-scenario. For each line coming from or going beyond the boundary nodes of the disaggregated subnetwork we create a virtual end station node which is connected by a single section to the corresponding boundary node. The section lengths with the appropriate trip times, the turnaround times of the line outside the disaggregated subnetwork together with the run- and dwell times within the disaggregated subnetwork have to sum up to the proper roundtrip time. This segmentation of disaggregated subnetwork and aggregated subnetwork into a new mesoscopic infrastructure model is illustrated in Figure 5b.

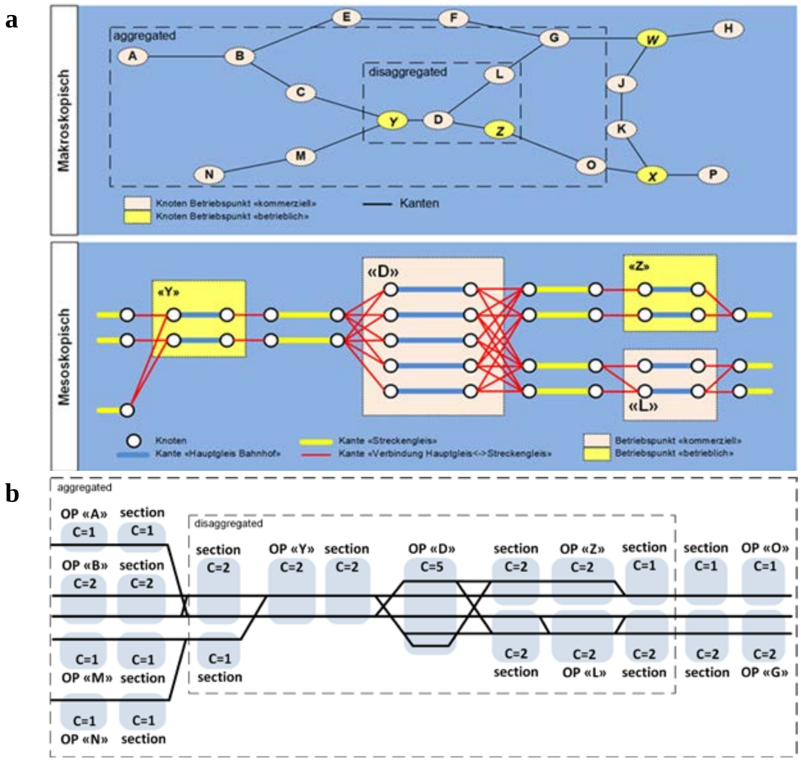


Figure 5: a) Example of mesoscopic and macroscopic topology from Howald et al. (2017). In order to divide the relevant infrastructure for the IP timetabling scenario into a segment with the relevant level of detail and a peripheral part with more coarse information, the railway network is partitioned into subnetworks with different topology levels. A disaggregated subnetwork containing the relevant infrastructure segments on mesoscopic level and an aggregated subnetwork, representing infrastructure on the macroscopic level. b) The illustrates the disaggregated subnetwork representation in our model which can be configured with detailed mesoscopic information.

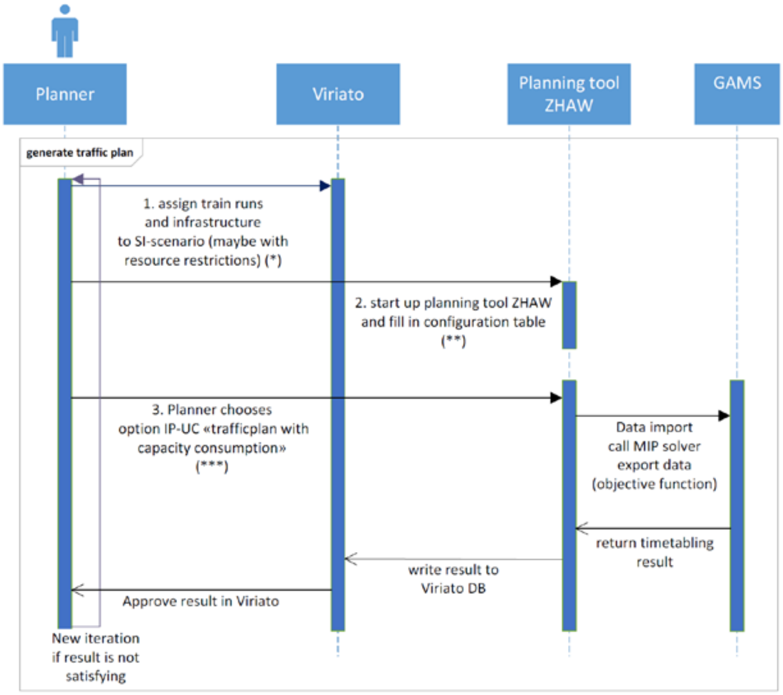
2.4 Method for generating a timetable with event flexibility

In the IP use case a timetable with event flexibility, including track assignment, is generated, if possible. In this section we will describe the implementation and the software used in detail. First, we describe the actions of the IP use case from the point of view of a planner (see Figure 6):

1. In the first step, the planner has to enter SI data into the planning system (Viriato).
2. Additional SI data (not implemented in Viriato) like transfer times between lines, turnaround times and time dependencies which implement e.g. service frequencies are entered in the ZHAW planning tool (in-house development).
3. The planner initiates the timetable calculation with the TCFPESP model for

the IP-use case by selecting the appropriate objective function and parameters according to the iteration schemes of sections 2.4.1 and 2.4.2 below. In the ZHAW planning tool, this induces the building of the EAN and the data import to the Algebraic Modeling System GAMS (GAMS, 2018). The TCFPESP model is implemented in GAMS. GAMS attempts to automatically calculate a solution based on the chosen model with a MIP solver (we use CPLEX 12.6.3).

A feasible solution is then returned to Viriato. If no feasible solution can be generated, the planner has to review and eventually relax the SI. This may lead to a backward loop to ‘line planning, line concept’ in Figure 1.



(*) Configure: dwell times, segment trip times, service travel times
(**) Configure: connection times, turnaround times, -time dependencies
(***) Choose an objective function according to iteration scheme: MinTravel or ConTravel

Figure 6: Sequence diagram and system functions of actions during the use case IP. The sequence diagram shows the tasks and functions of the involved planner and system components. Next to the planner there are several system components involved. Viriato (see SMA (2018)) is used as timetable data editor. The planning tool ZHAW implements all timetable data preparation functions not included in Viriato. The main actions within this use case are executed according to the explanation in the text and the iteration schemes described in section 2.4.1. The TCFPESP-model is solved in GAMS. For GAMS see GAMS (2018)

In the following two sections, we describe the methods for the construction of a timetable with event flexibility under normal (unrestricted) operational conditions and under restricted conditions due to maintenance.

2.4.1 Construction of a timetable with event flexibility for periods with normal operations

The heuristic iteration scheme 1 is the standard iteration scheme for applying TCFPESP (see section 2.2.4) under normal operations. If we are able to compute a timetable with iteration scheme 1, we refer to it as reference timetable. If the availability of the involved resources during interval planning is restricted compared to the reference timetable, a new (temporary) interval timetable must be generated with the help of the heuristic iteration scheme 2 in section 2.4.2, below.

Aim: Try to generate a feasible reference timetable with event flexibility.
Input:

- SI-data (Line data, line transfers, time dependencies, track infrastructure, rolling stock, train properties, etc.)
- Maximal size of flexibility δ_{max} for all arrival and departure nodes (available for planning on the micro-level or for stability reasons)
- Parameter ϵ for controlling maximal deviations of optimal passenger travel times
- Bound on rolling stock per line

(at the beginning the sizes of the event flexibility δ_{max} and parameter ϵ are set to default values and adapted during the iteration in order to achieve the feasibility or improve the stability of the timetable scenario)

Iteration scheme 1

1. Solve the model MINTRAVEL. We get a timetable with best possible travel times f_{TT}^{*}. If the model MINTRAVEL is not feasible adjust SI and go to step 1.
2. Compare the necessary amount of rolling stock per line for the timetable from the MINTRAVEL model with the given bound on rolling stock. If one line needs too much rolling stock, adjust SI and go to step 1.
3. Solve the model CONTRAVEL, allow passenger travel times to be maximal $(1+\epsilon)f_{TT}^{*}$. We get a timetable with (maximal) event flexibility.
4. Compare the necessary amount of rolling stock per line for the timetable from the CONTRAVEL model with the given bound on rolling stock. If one line needs too much rolling stock, reduce ϵ by multiplying ϵ with a positive factor smaller than 1 and go step 3.
5. Release timetable with event flexibility as reference timetable.

Iteration scheme 1 is clearly a heuristic scheme. The adjustment of the SI in step 1 and step 2 corresponds to the backward loop to ‘line planning, line concept’ in Figure 1. This loop is not part of this paper but of ongoing research. A description of the state of the work on this loop can be found in Wüst et al. (2018) and Bütikofer et al. (2019).

The computed event flexibility in the reference timetable can be tested with respect to stability or feasibility on the micro level. If the flexibility is not satisfying, we may loop the CONTRAVEL model and adjust the weights w_e in the objective function f_{flex}. This loop is also not part of this paper, but it is described in detail in Wüst et al. (2018) and Wüst et al. (2019).

In step 2 and 4 we control the necessary amount of rolling stock. This is possible since we are including turnaround activities in our EAN according to Liebchen and Möhring (2007) (see section 3.3).

2.4.2 Construction of a commercial timetable for periods with maintenance intervals

In this section we want to demonstrate how to adapt the iteration scheme 1 in order to generate a feasible timetable with event flexibility for maintenance intervals. During the respective maintenance interval, the scheduled trains in the temporary timetable should be as close to those in the reference timetable, that it is possible to communicate only one 'commercial' timetable to the customers. This is positive from a customer perspective, but also from an operator's perspective since restoring the reference timetable after the maintenance interval has finished is easier in this case. In addition, the free capacity in the network can be used for additional services (e.g. freight trains) during the whole planning horizon.

Aim: Try to generate a commercial timetable with event flexibility feasible for no, one or several (n) construction intervals (i.e. feasible for all scenarios with resource restrictions).

Input: (see iteration scheme 1) In addition:

- Infrastructure restrictions for all n maintenance intervals
- Maximal time tolerance between event times of timetables with event flexibility of the single maintenance intervals. These event flexibility values represent the commercially tolerable variation of departure and arrival times during the planning horizon in contrast to operational event flexibility, which facilitates timetable feasibility at the microscopic level.

Iteration scheme 2

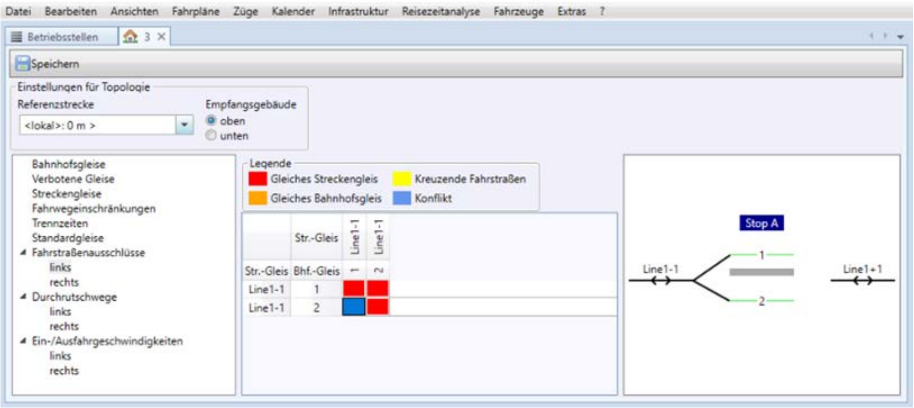
1. Start with a first construction interval: Compute a timetable with event flexibility for this construction interval with the help iteration scheme 1.
2. For each line take the passing times at the station nodes of the disaggregated network (see section 2.3) and add them to the SI with the expected tolerance time from the input. The remaining construction intervals will be computed with this adapted SI.
3. Compute the timetable with event flexibility for all construction intervals with the adaptations of the SI from step 2 and with the help iteration scheme 1.
4. Release timetables with event flexibility as interval timetables.

Iteration scheme 2 is again a heuristic scheme. As in iteration scheme 1, potential backward loops are not part of this paper (see comments in section 2.4.1).

In our case study (see chapter 3) we construct a timetable with event flexibility for each construction interval in the given timetable period. In practice, at every station and for every line the earliest departure and the latest arrival (with respect to all construction intervals) should be communicated as ‘commercial’ timetable to the customers.

2.5 Generation of timetables based on a standard planning tool

One of the main goals of the applied research project with SBB was to make the algorithmic timetable generation based on the proposed TCFPESP-method available to practitioners. Therefore, the generic configuration of whatsoever timetabling scenario should be possible, using a standard timetabling system such as “Viriato”, which is in use at SBB for service planning (see Viriato Info Folder, 2018). All kinds of relevant timetabling information like line and infrastructure data attributes can be entered easily in the appropriate masks (e.g. track connectivity data such as route exclusions between section and station tracks). For an example of the track configuration of an operation point and its neighbouring sections. See the Viriato mask in Figure 7. For more detailed information we refer to the Viriato User Manual (Viriato, 2016).



Einstellungen für Topologie

Referenzstrecke
<lokal>: 0 m >

Empfangsgebäude
● oben
○ unten

Legende
- 🟥 Gleiches Streckengleis
- 🟧 Gleiches Bahnhofsgleis
- 🟨 Kreuzende Fahrstraßen
- 🟦 Konflikt

	Str.-Gleis	Line1-1	Line1-1
Str.-Gleis	Bhf.-Gleis	1	2
Line1-1	1	🟥	🟥
Line1-1	2	🟦	🟥

Figure 7: Viriato editor mask for entering timetabling configuration data attributes. The mask shows an example of track connectivity on one side of an operation point. Connectivity of station tracks and neighbouring section tracks, as well as potential conflicts can be entered and configured using appropriate data masks.

3 Case study ‘Kerenzerberg’

We start with the description of mesoscopic infrastructure and SI on our test sector. According to the IP business requirements (see section 2.1), the SI is the result of the planning steps ‘line planning, line concept’ and ‘traffic planning’ (see Figure 1 in section 2 for details) and is maintained in Viriato and the ZHAW planning tool. For the purpose of our case study, we adapted the existing SI for the timetable of 2018 in such a way, that we are able to proof that we can handle the basic IP requirements with the pro-posed IP use case and the iteration schemes 1 and 2 for computer aided timetable generation.

3.1 Description of the infrastructure

The infrastructure between Ziegelbrücke and Sargans under normal operations is summarized in the following table. The infrastructure table is maintained in Viriato (see chapter 2).

Station/ Track ID	Ziegelbrücke (ZGB)	ZGB-WN	Weesen (WN)	WN-MH	Mühlehor (MH)	MH-TIEF	Tiefenwinkel (TIE)	TIE-MG	Murg (MG)	MG-UNT	Unterterzen (UNT)	UNT-MOL	Mols (MOL)	MOL-WAL	Walenstadt (WAL)	WAL-FMS	Flums (FMS)	FMS-MEL	Mels (MEL)	MEL-SA	Sargans (SA)
Number of tracks	12	2	2	2	2	1	2	2	2	2	2	2	2	2	3	2	2	2	2	2	4
Minimum travel time (Tracks)	1.7		2.8		1.3		1		1.9		1		1.6		1.8		3.3		1.5		

Table 1: Infrastructure data of the sector ZGB-SA. The table row ‘Number of tracks’ indicates the number of tracks at stations (station name abbreviations in brackets) and in sections between stations (pair of neighbouring station abbreviations). The table row ‘Minimum trip time’ indicates the maximum of the train and track specific technical trip times between station coordinates in minutes.

In the first row, we describe the stations (e.g. ZGB) and tracks (e.g. ZGB-MH). We see in Table 1 that there are always at least two tracks available, except between Tiefenwinkel and Mühlehorn, where only one track is available. Minimum trip times are derived from technical restrictions of the tracks.

3.2 Network Segmentation

In order to generate a traffic plan with capacity time bands, we have to segment the railway network into the relevant perimeter as explained in chapter 2.4. The SI in the next section is also adapted to the segmented network. We illustrate the network related to our case study ‘Kerenzerberg’ in Figure 8.

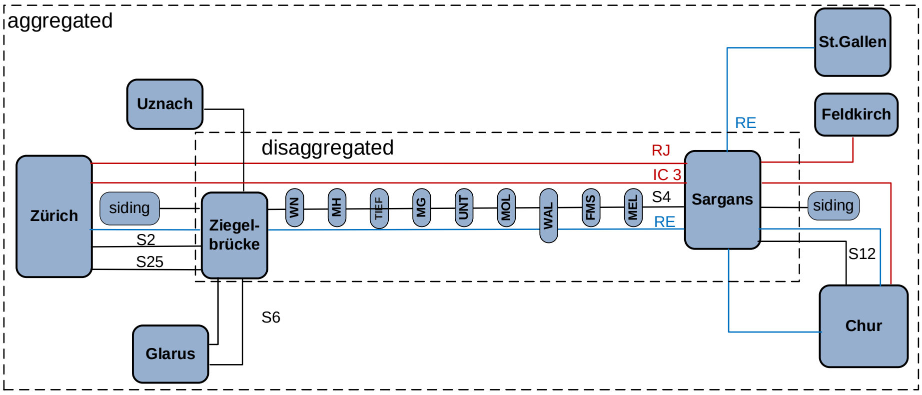


Figure 8: Network of the case study Kerenzerberg. In order to divide the relevant infrastructure for the IP timetabling scenario into a network partition with the relevant level of detail and a peripheral part with more coarse information, the railway network is divided into subnetworks. A disaggregated subnetwork containing the relevant infrastructure segments at mesoscopic level and an aggregated subnetwork, representing simplified infrastructure on the macroscopic level.

As the planned construction or maintenance work for our test scenario is located on the network section between Tiefenwinkel and Mels, we decided to use the corridor Ziegelbrücke-Sargans as the disaggregated partition of the test network, where we will generate a detailed timetable (see section 2.4). The western part of Ziegelbrücke is aggregated, i.e. we introduced the nodes Uznach, Zürich, Glarus and a siding of Ziegelbrücke and connecting tracks. The aggregated network will be used to maintain vehicle circulation (e.g. turnarounds) aspects of lines and to model connections to tangent lines (see the description of SI in the next section). The eastern part of Sargans is also aggregated. We introduced the nodes St.Gallen, Feldkirch, Chur and a siding of Sargans. In the aggregated network we assume to have enough track capacity. Ziegelbrücke and Sargans can be considered as local hubs and represent the boundary nodes of the disaggregated network partition (see section 2.3). At these stations the timetable has to account for passenger transfers between lines. Technically spoken, these transfer requirements result in connections constraints in our TCFPESP-model.

3.3 Description of Service Intention

In chapter 2 we explained that the SI is our main data structure and is maintained in Viriato and the ZHAW planning tool. The SI contains all the information needed to configure the EAN and the TCFPESP model (see sections 2.2.2 and 2.2.3). We start with the lines considered. As mentioned before, our SI-lines represent an adaption of the lines in the corresponding timetable 2018. To demonstrate the turnaround operations, we decided that the line S4 makes a turnaround in a siding next to Ziegelbrücke and Sargans, respectively.

S 12	S 11	S 10	S 9	S 8	S 7	S 6	S 5	S 4	Stops / Line-segments
		[2, 58]	[2, 58]						Glarus (GL)
		[0.5, 0.8]	[0.5, 0.8]						GL-ZGB
	[2, 58]								St.Gallen (SG)
	[0.5, 0.8]								SA-SG
[2, 58]			[2, 58]		[2, 58]	[2, 58]	[2, 58]		Zürich (ZUE)
[0.5, 0.8]			[0.5, 0.8]		[0.5, 0.8]	[0.5, 0.8]	[0.5, 0.8]		ZUE-ZGB
		[2, 58]							Uznach (UZ)
		[0.5, 0.8]							UZ-ZGB
								[2, 3.2]	Siding (SZGB)
								[0.5, 0.8]	SZGB-ZGB
[2, 58]		[2, 3]	[2, 3]		[1, 1.5]	[0, 1]	[0, 1]	[2, 3]	Ziegelbrücke (ZGB)
					[1.7, 2.6]	[1.7, 2.6]	[1.7, 2.6]	[1.7, 2.6]	ZGB-WN
					[0, 1]	[0, 1]	[0, 1]	[0, 1]	Weesen (WN)
					[2.8, 4.2]	[2.8, 4.2]	[2.8, 4.2]	[2.8, 4.2]	WN-MH
					[0, 1]	[0, 1]	[0, 1]	[1, 1.5]	Mühlehorn (MH)
					[1.3, 2]	[1.3, 2]	[1.3, 2]	[1.3, 2.00]	MH-TIE
					[0, 1]	[0, 1]	[0, 1]	[0, 1]	Tiefenwinkel (TIE)
					[1, 1.5]	[1, 1.5]	[1, 1.5]	[1, 1.5]	TIE-MG
					[0, 1]	[0, 1]	[0, 1]	[0.5, 0.8]	Murg (MG)
					[1.9, 2.9]	[1.9, 2.9]	[1.9, 2.9]	[1.9, 2.9]	MG-UNT
					[0, 1]	[0, 1]	[0, 1]	[0.5, 0.8]	Unterterzen (UNT)
					[1, 1.5]	[1, 1.5]	[1, 1.5]	[1, 1.5]	UNT-MOL
					[0, 1]	[0, 1]	[0, 1]	[0.5, 0.8]	Mols (MOL)
					[1.6, 2.4]	[1.6, 2.4]	[1.6, 2.4]	[1.6, 2.4]	MOL-WAL
					[1, 1.5]	[0, 1]	[0, 1]	[0.5, 0.8]	Walenstadt (WAL)
					[1.8, 2.7]	[1.8, 2.7]	[1.8, 2.7]	[1.8, 2.7]	WAL-FMS
					[0, 1]	[0, 1]	[0, 1]	[1, 1.5]	Flums (FMS)
					[3.3, 5]	[3.3, 5.0]	[3.3, 5]	[3.3, 5]	FMS-MEL
					[0, 1]	[0, 1]	[0, 1]	[0.5, 0.8]	Mels (MEL)
					[1.5, 2.3]	[1.5, 2.3]	[1.5, 2.3]	[1.5, 2.3]	MEL-SA
	[2, 3]			[2, 58]	[2, 3]	[1, 1.5]	[2, 3]	[2, 3]	Sargans (SA)
								[0.5, 0.8]	SA-SSA
								[2, 3.2]	Siding SA (SSA)
	[0.5, 0.8]			[0.5, 0.8]		[0.5, 0.8]			SA-CH
	[2, 58]			[2, 58]	[2, 58]	[2, 58]			Chur (CH)
							[0.5, 0.8]		SA-FE
							[2,58]		Feldkirch (FE)

Table 2: Lines in the case study Kerenzerberg.

In Table 2 we summarized the upper and lower bound for dwell at every station ([D_{lo}, D^{up}]) and trip time for every track ([TT_{lo}, TT^{up}]). The routing can be derived from the entries in the table. A line visits all the stations and tracks from top to bottom and vice versa, where an upper and lower bound is given. Stations and tracks, which are not on the routing of a line, have no entry in corresponding field. In the first and the last station the lines perform a turnaround in the given interval ([TU_{lo},TU^{up}]).

The minimum dwell D_{lo} and the minimum trip time TT_{lo} are given technical lower bounds. To compute the upper bounds D^{up} and TT^{up}, we multiplied the lower bounds with 1.5. This reserve will be used to derive flexible plans with the TCFPESP model.

The turnaround times are computed according to the approach of Liebchen and Möhring (2007). The turnaround intervals are computed in such a way, that a service with a minimal number of rolling stock is possible. In our case study, line S4 is operating with one rolling stock. The other lines operate with more than one rolling stock due to longer round-trip times. These bounds are not computed according to Liebchen and Möhring (2007), they are set manually. These lines can cross themselves in opposite directions (as it is in the real-world timetable). Whereas the line information in Table 2 is mainly maintained in Viriato, the turnaround and connection times (see Table 3) are entered in the planning tool ZHAW.

The SI contains the following connections between the given lines:

Connection **[C_{lo}, C^{up}] From/To** **at station**	**S4 (ZGB-SA)**	**IC 3 (ZGB-SA)**	**IC 3 (SA-ZGB)**	**RE1 (ZGB-SA)**	**RE 1 (SA-ZGB)**	**RJ (ZGB-SA)**	**RJ (SA-ZGB)**	**RE 2 (CH-SG)**	**S 12 (SA-CH)**	**S 25 (GL-ZUE)**	**S 25 (ZUE-GL)**	**S 2 (ZGB-ZUE)**
S 4 (ZGB-SA)									[1,15] SA	[1,15] SA		[1,15] ZGB
S 4 (SA-ZGB)										[1,15] ZGB		
IC 3 (SA-ZGB)					[1,15] SA	[1,15] SA						
RE 1 (ZGB-SA)		[1,15] SA						[1,15] SA				
RE 1 (SA-ZGB)						[1,15] SA						
RE 2 (CH-SG)					[1, 15] SA	[1,15] SA						
RE 2 (SG-CH)			[1,15] SA		[1, 15] SA	[1,15] SA						
S 6 (GL- UZ)				[1,15] ZGB								
S 6 (UZ-GL)				[1,15] ZGB								
S 12 (CH-SA)			[1,15] SA		[1, 15] SA	[1,15] SA		[1,15] SA				
S 25 (ZUE-GL)	[1,15] ZGB											
S 2 (ZUE-ZGB)	[1,15] ZGB										[1,15] ZGB	

Table 3: Connections in the case study Kerenzerberg

In Table 3 we find the implemented connections. The connections belong to the SI and are part of the output of the planning step ‘line planning, line concept’. The connections should take place in the time interval [C_{lo},C^{up}] from the line in the first column to line in the corresponding column, e.g. there should be a connection from the line IC 3 (direction SA-ZGB) to line RE 1 (direction SA-ZGB) in Sargans with a minimum and maximum time of 1 and 15 minutes, respectively. The connection Table 3 is maintained in the ZHAW planning tool.

Furthermore, the SI contains:

- A time separation of the lines S4 (ZGB-SA) and RE 1 (ZGB-SA) of [20, 40] minutes in Ziegelbrücke. This should guarantee a frequent service for passengers travelling from Ziegelbrücke to Sargans.
- Trip time restrictions for the lines S4, IC 3, RE 1 and RJ between Sargans and Ziegelbrücke, i.e. trip times should be between 17 and 21 minutes for the IC 3, RE 1 and RJ. Line S4 is restricted to be between 20 and 29 minutes.

The time separation and the trip time restrictions are part of the output of the line planning step.

3.4 Construction of a timetable with event flexibility for periods with normal operations at Kerenzerberg

The models MINTRAVEL and CONTRAVEL are implemented in the algebraic modeling system GAMS (24.7.4). We use CPLEX (12.6.3) to solve these two MIP's. The relative gap was set to 10% and the absolute gap to 15 minutes as stopping criterion. The computations were performed on Dell Latitude E6430 with an Intel 2.4 GHz quad core processor with 8 GB RAM.

In the case study Kerenzerberg we use the SI described in section 3.3. We have set the maximal flexibility δ_{max} to 10 seconds and parameter ϵ to 0.5. These values are based on experience of planning experts. The line S4 should work with one rolling stock.

In step 1 of iteration scheme 1 the MINTRAVEL model results in a MIP with 23072 constraints and total 25629 variables (11401 integers). We could solve the MINTRAVEL model within 1272 seconds. The rolling stock of line S4 is 1. The CONTRAVEL model in step 3 results in a MIP with the same size as the MINTRAVEL model. We could solve the CONTRAVEL model within 1075 seconds. The rolling stock of line S4 is again 1. In total we have to solve the models MINTRAVEL and CONTRAVEL only one time. A reduction of ϵ in step 4 or an adjustment of SI in step 1 and 2 was not necessary.

We get the following reference timetable with event flexibility and resulting track allocations at the end of iteration scheme 1.

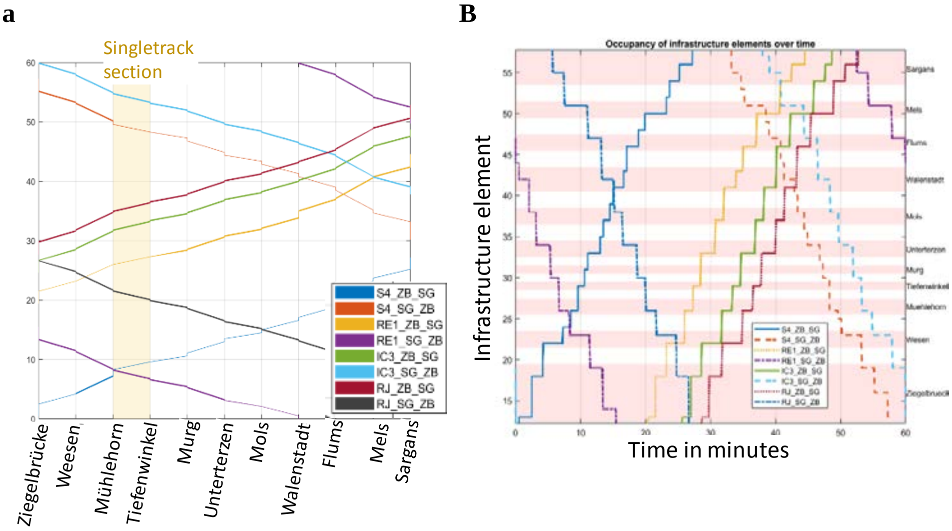


Figure 9: Reference timetable with event flexibility (a) and track assignment diagram (b) under normal operations. The track assignment diagram indicates which infrastructure element (enumerated track) is assigned to which train line during what time interval. Further explanations are given in the text.

Figure 9 illustrates the resulting reference timetable with the corresponding track allocation. As one can see, the SI is fulfilled in general. Especially we can see that

- line S4 operates with one rolling stock as requested.
- the service of S4 and RE 1 is separated in Ziegelbrücke to guarantee smooth services to Sargans.
- the track choice method TCFPESP is able to generate a feasible track allocation on the mesoscopic infrastructure.

In Figure 9 one should note the fact that this reference timetable has crossings between Flums and Mels. It will not be feasible for the considered construction intervals in the next section 3.5.

3.5 Construction of a commercial timetable for periods with maintenance intervals at Kerenzerberg

In this section we want to demonstrate the application iteration scheme 2 from section 2.4.2. We consider two construction sites. The construction sites are between Tiefenwinkel and Unterterzen (construction site 1) resp. Flums and Mels (construction site 2). The construction intervals take place during our planning horizon but in different time windows. Only one track is available during the construction intervals on the affected corridors.

The computer infrastructure, software and parameters of iterations scheme 1 are the same as in section 3.4. Furthermore, we assume a maximal time tolerance of 6 minutes between the computed interval timetables (see section 2.4.2).

In step 1 of iteration scheme 2 we started with construction interval 2 between Flums and Mels, since the single-track section is shorter than the one in construction interval 1 (see Figure 10). We could generate an interval timetable for construction site 2 without a reduction of ϵ or an adjustment of SI. In total we have to solve the models MINTRAVEL and CONTRAVEL only one time. The MINTRAVEL and the CONTRAVEL model results in a MIP with 23072 constraints and total 25581 variables (11353 integers). We could solve the MINTRAVEL resp. CONTRAVEL model within 1564 seconds resp. 1378 seconds. The models are little bit smaller in comparison to section 3.4 due to the reduced number of tracks during construction interval 2.

In step 2 of iteration scheme 2 we took the passing times from all lines in every station between Ziegelbrücke and Sargans from the interval timetable for construction site 2 and add them to the SI with the maximal tolerance of 6 minutes. For the construction interval 1 we allow therefore the lines to pass +/- 3 minutes with respect to passing times from construction interval 2.

In step 3 of iteration scheme 2 we compute a timetable for construction interval 1 with the adapted SI from step 2. We could generate an interval timetable for construction site 1 without a reduction of ϵ or an adjustment of SI. In total we have to solve the models MINTRAVEL and CONTRAVEL only one time. The MINTRAVEL and the CONTRAVEL model results in a MIP with 23072 constraints and total 25586 variables (11385 integers). We could solve the MINTRAVEL resp. CONTRAVEL model within 19 seconds resp. 13 seconds. The faster running times are due to the adapted SI from step 2 (resp. the fixed passing times with a tolerance of +/- 3 minutes).

In step 4 of iteration scheme 2 we could release the interval timetables for construction site 1 and 2 and convert it into a ‘commercial timetable’.

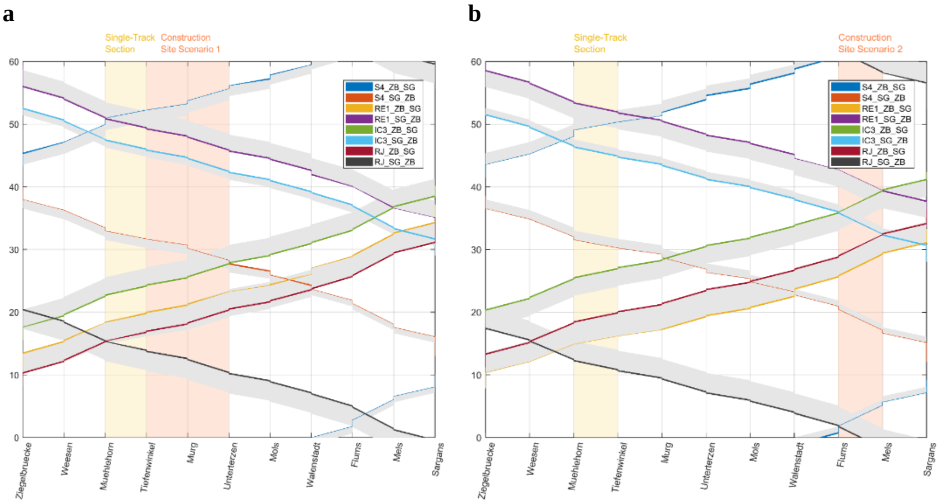


Figure 10: a) Timetable with event flexibility for construction interval 1, b) Timetable with event flexibility for construction interval 2. Both timetable scenarios (coloured lines are construction interval specific) are consolidated within one commercial Timetable (the grey bands contain the coloured lines). They are identical for both scenarios.

In Figure 10 we see the interval timetables with event flexibility for both construction intervals. Due to iteration scheme 2 the timetables for the lines are at the lower or the upper boundary of the grey band. The grey band corresponds to the ‘commercial’ timetable. The timetable for construction interval 1 (Figure 10a) is not feasible for construction interval 2 (Figure 10b) and vice versa, e.g. line RJ and line S4 have a crossing between Flums and Mels during construction interval 1. It is worth mentioning that the order of line RJ and line RE1 from Ziegelbrücke to Sargans changes from construction interval 1 to 2.

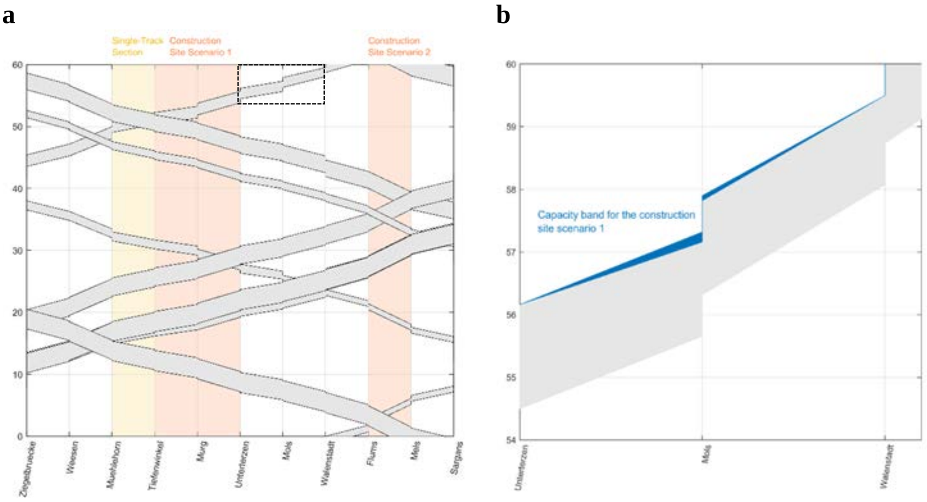


Figure 11: a) ‘Commercial timetable’, with boundaries of the grey time bands indicating earliest departure and latest arrival times b) Event flexibility of line S4 illustrated in blue colour for the train slot between Unterterzen and Walenstadt. This example corresponds to the zoomed in area marked by the stipulated rectangle in a)

Figure 11a shows the ‘commercial timetable’ for the entire planning horizon covering both construction intervals. The departure times correspond to the lower boundary of the grey band and the arrival times to the upper boundary. During the planning horizon, we therefore always find a feasible timetable for all construction intervals. On the right (b) we see a detailed view of the timetable of the line S4 between Unterterzen and Walenstadt. The blue bands represent the event flexibility, e.g. during construction interval 1 we have around 10 seconds flexibility for the arrival and the departure in Mels.

Applying the iteration scheme 2 of section 2.4.2, we were thus able to generate one single commercial timetable with two similar but different capacity plans (‘timetables with event flexibility’), which fulfils the SI during the entire planning horizon. That means that in practice the railway operator would have to communicate the commercial timetable to the passengers only once.

Hence, with the case study ‘Kerenzerberg’ we could show, that based on the SI and our TCFPESP-model, we were able to integrate operational stability (generating two different capacity plans) and passenger travel time aspects (finding one single commercial timetable with robust travel times as all transfers are guaranteed) in the proposed interval planning use case.

4 Discussion and outlook

4.1 Summary

We introduced and successfully applied the new timetabling model TCFPESP, which can be used to support timetable planners for generating train and vehicle schedules with track assignment. This model is based on an extension of the well-known FPESP model and can be configured by using a standard schedule editor.

The use case and the TCFPESP model that we describe in chapter 2 are tested in a small-scale test and a real-world case study for IP in chapter 3. The generation of the commercial timetable is achieved by an iterative execution of the IP use case for finding timetables with event flexibility for two different maintenance planning scenarios. We show how the concept of SI can be used to develop a customer timetable, which is valid during the complete timetable period. At the same time, it is now possible, that two different construction or maintenance intervals with different locations can be planned during one single timetable period. This is of considerable practical relevance, especially with regard to the increasing number of intervals to be planned and executed under conditions of continued production of railway services.

4.2 Outlook and future research

If timetabling requirements turn out to be infeasible to be solved by TCFPESP, because, e.g., the given SI is not realizable on the respective railway infrastructure (a typical situation during construction intervals), this situation must be solved by a relaxation of the SI. This is indicated by the grey backward arrows in the planning process of the IP business requirements in section 2.1, showing that in this case one has to go back to previous planning steps and relax the SI. In a next research step, we want to find out, how the SI can be generated using standard line planning methods similar to those described by, e.g., Schöbel and Scholl (2006) or Friedrich et al. (2017). Our preliminary investigations show that these methods can generate SI configurations that take reduced resource availability (due to the fact, that e.g. tracks are temporary out of service) into consideration. This research will help to make detailed specifications of data interfaces and service levels between TOC and IOC in case of IP and operational disruptions in real-time conditions.

Another aim of future research concerns the method for the utilization of timetable stability measures, such as cumulative delay impacts and cumulative delay sensitivity, obtained from timetable performance measurement for assigning event flexibility to improve timetable robustness (see Wüst et al. (2019) for preliminary results). With the outcome of this future research, we will be able to provide a detailed use case description of the iteration between the IP use case presented here and use cases for the assessment of timetable robustness. In that way, we expect to further improve the quality of TCFPESP results and contribute for speeding up and facilitating practical railway timetabling.

Acknowledgements

We highly acknowledge the whole SBB-team of SmartRail 4.0, Programm TMS-PAS, for their encouraging support during the whole course of the project. We like to thank especially Thomas Wieland and Thomas Künzi for their inspiring ideas and the nice atmosphere, they knew to create.

This work was financially supported by the SBB Research Fund, c/o SBB Lab, Universität St.Gallen, Dufourstrasse 40a, 9000 St.Gallen.

References

Bešinović, N., Goverde, R.M.P., Quaglietta, E., Roberti, R., 2016. An integrated micro-macro approach to robust railway timetabling. Transportation Research Part B: Methodological, vol. 87, pp. 14-32

Bütikofer, S., Frick, K., Köchli, J., Weber, C., 2019. Automatisierte Linienplanung im öffentlichen Verkehr, ETR Swiss / Wissen, accepted, Eurailpress.de

Caimi, G., 2009. Algorithmic decision support for train scheduling in a large and highly utilised railway network. Diss. ETH Zürich Nr. 18581, ETH Zürich, Zürich, Switzerland

Caimi, G., Laumanns, M., Schüpbach, K., Wörner, S., Fuchsberger, M., 2011a. The Periodic Service Intention as a Conceptual Frame for Generating Timetables with Partial Periodicity. Transportation Planning and Technology, vol. 34, issue 4, pp. 323-339

Caimi, G., Fuchsberger, M., Laumanns, M., Schüpbach, K., 2011b. Periodic railway timetabling with event flexibility. Networks, vol. 57, no. 1, pp. 3-18

de Fabris, S., Longo, G., Medeossi, G., Pesenti, R., 2014. Automatic generation of railway timetables based on a mesoscopic infrastructure model. Journal of Rail Transport Planning & Management, vol. 4, pp. 2–13

Friedrich, M., Hartl, M., Schiewe, A., Schöbel, A., 2017. Integrating passengers' assignment in cost-optimal line planning. Technical Report 2017-5, Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Georg-August Universität Göttingen. URL: `http://num.math.uni-goettingen.de/preprints/files/2017-5.pdf`

GAMS, 2018. General Algebraic Modeling System GAMS, GAMS Software GmbH, Frechen, Germany, URL: `https://www.gams.com/` (last accessed: 19 September 2018)

Hansen, I.A., Pachl., J. (eds.), 2008. Railway Timetable & Traffic. Analysis – Modelling – Simulation. Eurail Press, Hamburg, Germany

Herrigel, S., 2015. Algorithmic decision support for the construction of periodic railway timetables, Diss. ETH Zürich Nr. 22548, Department Bau, Umwelt und Geomatik, ETH Zürich, Zürich, Switzerland

Howald, P., Künzi, Th., Wild, P., Wieland, Th., 2017. Grobkonzept «Linienplanung» Version 0.7, 19.10.2017. SBB Smart Rail 4.0 TMS-PAS, project document repository

Liebchen, C., Möhring, R.H., 2007. The modeling power of the periodic event scheduling problem: Railway timetables – and beyond. In: Geraets, F., Kroon, L., Schöbel, A., Wagner, D., Zaroliagis, C. (eds.), Algorithmic Methods for Railway Optimization, Lecture Notes in Computer Science, vol. 4359, Springer, Berlin/Heidelberg, Germany, pp. 3–40

Schöbel, A., Scholl, S., 2006 Line planning with minimal travel time, 5th workshop on algorithmic methods and models for optimization of railways, Dagstuhl Seminar Proceedings 06901

SMA, 2018. Viriato - software for railways. Info Folder: URL: `https://www.railsystemsengineering.com/RSE/documents/SMA%20Viriato%20Brochure.pdf` (last accessed: 7 September 2018)

SMA, 2016. Viriato - software for railways. User Manual (German): Viriato 8.9 Standard, Benutzerhandbuch

Wüst, R. M., Laube, F., Roos, S., Caimi, G., 2008. Sustainable Global Service Intention as objective for Controlling Railway Network Operations in Real Time, In: Proceedings of the WCRR 2008. Seoul

Wüst, R., Steiner, A., Gomez. C., 2017. Kundenfreundliche und robuste Ersatzfahrpläne während Bau- und Unterhaltsintervallen, ETR Swiss / Wissen, vol. 10, pp. 57-61, Eurailpress.de

Wüst, R., Bütikofer, S., Ess, Gomez, C., Steiner, A., Laumanns, M., Szabo, J., 2018. Periodic timetabling with 'Track Choice'-PESP based on given line concepts and mesoscopic infrastructure. Operations Research 2018, Bruxelles, Conference proceedings

Wüst, R., Bütikofer, S., Ess, S., Gomez, C., Steiner, A., Laumanns, M., Szabo, J., 2019. Improvement of maintenance timetable stability based on iteratively assigning event flexibility in FPESP. 8th International Conference on Railway Operations Modelling and Analysis, Norrköping, Sweden (accepted)

K Conference paper, 2nd International Railway Symposium Aachen (IRSA), 26 to 28 November 2019, Aachen, Germany (to be submitted)

Title: Generation of the transport service offer with application to timetable planning considering constraints due to maintenance work

Authors: R. Wüst, S. Bütikofer, J. Köchli, S. Ess, Institute for Data analysis and Process design (IDP), ZHAW

Key Words: Timetable planning, Maintenance intervals, Transport service intention, Constraint relaxation, Line planning

1. Introduction

1.1. Generation of the transport service intention

Automation and digitization in the development of transport chains is a big challenge for public transport in the future. The goal is to achieve optimal and consistent planning across all process steps and time horizons to increase the degree of automation and the service quality. The two major steps in transport service planning are: (a) line planning and (b) timetable generation. These steps are carried out in several iteration loops involving coordinated activities across different companies, such as railway operators and infrastructure managers. In both steps, mathematical models can support the planning decisions. A detailed overview of mathematical models for step (a) line generation was provided by Schöbel (2012). Models considering both steps (a) and (b) are described in several publications (see e.g. Barber et al. (2008) and Liebchen et Möhring (2007)). However, there is still little literature on the coordination of these models for practical applications (see Schöbel (2017). This article is about the generation of timetable scenarios based on origin-destination (OD) demand-matrices and the transport service offer in public transport serving the OD-demand. We call this offer service intention (SI). The SI is represented by a set of public transport lines including technical and commercial parameters. Basically, technical line properties are represented by line categories and trip times. Commercial properties include dwell and transfer times and thus represent customer relevant service levels for each network OD-relation. A formal description of the SI was first presented by Caimi (2009). Caimi also demonstrates that the SI is a suitable input for the timetabling step (b) especially if the result of (b) is used as input for generating track slots for the configuration of a traffic management system at a microscopic level (see Caimi (2009)). Like the approach of Caimi et al. (2011), we make use of a timetabling model which is based on an extension of the PESP model. Our model refers to a mesoscopic level of infrastructure detail. We call this PESP extension TCFPESP (Track Choice Flexible Periodic Event Scheduling Problem) (see Wüst et al. (2018) and (2019) for technical details). In this article we want to demonstrate, that the SI can be generated automatically in step (a). To emphasize the relevance of our results, we demonstrate how the proposed method can be used as input for automatically generating the timetable considering constraints due to maintenance work in a real-world scenario.

1.2. Steps of the planning process

The method for step (a), which is presented here consists of the determination of line routes together with their frequencies in a specific rail network such that a given passenger transport demand can be satisfied. Lines typically connect two endpoints by a sequence of intermediary stations. All lines in a given rail network are hierarchically organized by line categories. Passenger demand for each line category is estimated based on a passenger assignment method which is described in detail by Oltrogge (1994). We will refer to it as ‘system split’. Each line category has a maximum passenger capacity Cap which is determined by the seat capacity of the specific rolling stock unit of the line. The capacity of a line is then calculated from Cap and the operating frequency. If several lines are operated on the same track edge, the respective edge capacity in terms of a maximum number of train slots (e.g. per hour) has to be respected additionally. There are two conflicting objectives in line planning. On the one hand, the operating costs and on the other a weighted combination of passenger travel times and the number of transfers is to be minimized. We will show, that the result of (a) is the SI. Once the SI is given, we create a timetable in (b) which can be tested for feasibility at a mesoscopic level of detail (see Wüst et al. (2019)). This timetable in combination with passenger flow is the basis for customer information as well as the subsequent steps of operational planning.

If there is a reduction of transport capacity, for instance due to track maintenance work, it may happen, that no feasible timetable can be created which satisfies the SI. In this case, we propose to go back to the line planning step and create a new (relaxed) SI. The challenge is to find a feasible timetable which on one side has a low impact on the passenger service level in terms of total travel time and on the other side takes temporary reduced resource availability and operational restrictions into account. The innovation here is the fact, that the use of the SI allows to go back to the previous planning step, i.e. the line planning step, in order to create a revised timetable input that takes into account the new restrictions.

1.3. Structure of this article

In chapter 2 we first describe a suitable line planning model that we selected for the application in maintenance timetable planning (see section 2.1). In section 2.2 we describe how the line plan is used to configure the SI and in section 2.3, we describe how the SI is constructed, based on the results from the line planning step. The aim is to provide sufficient detail in order to be used for configuring the timetabling model TCFPESP.

Finally, in chapter 3 we present the results of a case study, where we applied the methods described in chapter 2 to a real-world scenario. We apply the methods for generating an (adapted) transport service offer and applying it to timetable planning twice, once for generating a reference timetable for a given network scenario and once considering constraints due to maintenance work on a section in the given network. In chapter 4 we summarize the findings of our proposed approach and make some conclusion that we also use to motivate future research for further elaborating the iterative planning process.

2. From line planning to timetabling

In this section we give a short description of the proposed line planning (section 2.1) and the timetabling model (section 2.3). The generation of SI (section 2.2) is the main contribution of this paper.

2.1. Generating the line plan in planning step (a)

The customer demand in the line planning step (a) is given by an OD-matrix, in which the coefficients represent the demand between pairs of nodes in a given time period T (e.g. 60 minutes). The basic requirement is to cover the demand for transportation according to the OD-matrix, the customer-friendliness is based on the shortest possible journey times and the cost-efficiency is given by achieving these goals at the lowest possible operating costs.

The line planning model presented here builds on the solution approach of Friedrich et al. (2017). They describe in detail how lines and the appropriate line frequencies from a given line pool are selected. In the strategic planning process, a line pool represents all lines belonging to a given line category, such as intercity lines or commuter train lines. Basically, each line l in the line pool $\mathcal{L}_0$ is specified by its route $(v_0^l \ldots, v_K^l)$, i.e. its sequence of station or stop locations $v_k^l,\ k = 0, \ldots, K,$ out of the set of stations N, its vehicle type, its trip times from stop to stop and an OD demand for the corresponding line category

The application in chapter 3 refers to tactical planning requirements as we want to determine a timetable, which takes time intervals with reduced track capacity into consideration. There, we assume the set of given line pools (for the different line categories) to be fixed in the case of normal operations. For the time interval with maintenance work we will adapt the line pool due to reduced track capacity (see section 3.4). We show, how the frequencies of different lines in the line pool and the corresponding timetable can be derived from the underlying OD-demand once for the case of normal operations and once for the case of the maintenance work on a certain track segment.

The line planning model is based on two different network types, the ‘Public Transport Network’ (PTN) and the ‘Change & Go Network’ (CGN) (see Schöbel, 2012 for an introduction). The PTN is an undirected graph with station nodes $v \in N$ connected by direct track edges $e \in E$. In the PTN only existing track connections are considered.

As there are three line categories given in our application example we also have three categories of station or stop locations: intercity (IC), interregio (IR) and commuter (S-Bahn, see Table 1). This follows from to the system split which is described in section 1.2. Hence, every node $v \in N$ belongs to one of these three categories. To be able to define line connections for the timetable planning, not only the frequencies of the lines are needed, but also the information at which stations and how many passengers change to connecting lines. For this reason and to avoid frequent line transfers of passengers at transfer stations, the CGN is used in addition. The CGN will be built based on the given line pool and the PTN.

Like the PTN, the CGN is an undirected graph based on connected vertices represented by network stations. In addition, at each station and for each line serving that station, an interchange node is inserted and connected to the station nodes by a connecting edge. Further, these transfer nodes are then connected by driving edges to the transfer nodes of the nearest station served by the lines (see Table 1 for an example). These driving edges between consecutive nodes v_{k-1}^{l}, and v_{k}^{l}, of line l are weighted by the minimal travel time t_{trip}^{k-} plus the minimal dwell time $t_{dwell}^{(k-1)-}$ at node v_{k-1}^{l}, of line category corresponding to line l. Both times can be computed from infrastructure data (e.g. by using the length of a track-section and the technical speed restriction). Transfer edges receive weights θ^{+} , which represent passenger transfer times between two lines serving node v_k.

Based on the OD demand and the CGN, the shortest routes can now be determined from all nodes v_{k_1} to all nodes v_{k_2} in the network N, thereby calculating the number of passengers on each edge $e \in E$ of the PTN and also on the transfer edges in the CGN. Furthermore, with this edge demand w_e and the capacity Cap of the trains, the minimum frequency per edge $e \in E$ can be calculated from $f_e^{min} := \left\lceil \frac{w_e}{Cap} \right\rceil$. On the other side, f_e^{max} is given as the maximum slot capacity of edge $e \in E$, which depends on properties of the track infrastructure, safety restrictions (mainly headways) and planned rolling stock of the different line categories. f_e^{min} and f_e^{max} are used as input for determining the lines and their frequencies.

The problem $LineP$ is the basic cost model for line planning:

$$\min \sum_{l \in \mathcal{L}_0} cost_l f_l, \tag{1}$$

$$s.t.\ f_e^{max} \geq \sum_{l \in \mathcal{L}_0; e \in l} f_l \geq f_e^{min}, \qquad \forall e \in E,\ f_l \in \mathbb{N}, \forall\, l \in \mathcal{L}_0$$

and $cost_l$ being the fixed cost of operating line l.

$LineP$ results in the minimum and hence most cost-effective number of vehicle trips per line, which satisfies the given demand. The selected lines together with its frequencies, minimum trip and dwell times, start and end locations (turnaround), and transfer conditions represent the SI. In the proposed approach the connection and time dependency conditions in the SI can be derived from the passenger assignment step and the resulting travel chains (i.e. the transfer edges).

2.2. Generation of the SI

In the first part of this section we give a short introduction the notation of the SI. For details we refer to Caimi (2009). According to Caimi, the periodic SI for a given railway network is defined as $\mathcal{G} = (T, Z, C, D)$, where $T \epsilon \mathbb{R}^+$ is the considered time period (equal to the time period we used in section 2.1), Z the set of all periodic train runs, C the set of all connections and D the set of all dependencies. A train run $z \in Z$ is defined as the run over $K + 1$ nodes in the topology, repeated R times with periodicity ρ minutes:

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

where $v_k \epsilon N$ is the node visited in the k-th step of the train run. We associate an arrival event $arr_z(v)$ and a departure event $dep_z(v)$ to each node $v \epsilon N$ on the train run z. $t_{dwell}^{k-/+}$ is the minimal and maximal dwell time of the train between the arrival $arr_z(v_k)$ and $dep_z(v_k)$ (a value of zero means that the train passes the node without stopping), $\left[t_{trip}^{k-}, t_{trip}^{k+}\right]$ defines the allowed interval of the trip time between $dep_z(v_{k-1})$ and $arr_z(v_k)$, and $[\omega_k^-, \omega_k^+]$ is the (optional) time slot for the departure event of the first train recurrence.

A connection $c = (z_1, z_2, v, r_1, r_2,, \theta^-, \theta^+)$, $c \in C$, is defined as the possibility for the passenger to change from train run z_1 to train run z_2 in station $v \in N$. The connection takes place for the first time during the r_1-th repetition of train run r_1 and the r_2-th repetition of train run z_2 between θ^- and θ^+minutes, i.e. the arrival event $arr_{z_1}(v)$ should take place at least θ^- resp. at most θ^+ minutes before the departure event $dep_{z_2}(v)$.

A time dependency $d \in D$ is defined as a time constraint between two nodes of the periodic service intention, where $d = \left(z_1, z_2, e_z(v_{k_1}), e_z(v_{k_2}), r_1, r_2, \theta^-, \theta^+\right)$ and $e_z(v)$ is an arrival or a departure event associated to a node $v \in N$ on train run z. Again, the event $e_z(v_{k_1})$ of the k_1-th node of train run z_1 should occur between θ^- and θ^+minutes before the event $e_z(v_{k_2})$

k_2-th node of train run z_2. The dependency takes place for the first time during the r_1-th repetition of train run z_1 and the r_2-th repetition of train run z_2.

At next, we explain how the SI is generated. We state that we get from the line planning step (a) in section 2.1 all information required for constructing the SI. To illustrate our approach, we consider two lines l_1 and $l_2 \in \mathcal{L}_0$ from the output of the line planning. The lines $l_i = (v_0^{l_i}, \ldots, v_K^{l_i})$ are operated with frequency f_i for $i = 1, 2$. Without loss of generality there should be a transfer from line l_1 in direction of the end node $v_K^{l_1}$ to l_2 in direction of the end node $v_0^{l_2}$ at the common node $v_K^{l_1} = v_K^{l_2}$, which should take place in at most θ^+ minutes.

First the lines l_1 and l_2 generate four train runs (two in each direction)

$$z_i = \left(\left(v_k^{l_i}, t_{dwell}^{k-}, t_{dwell}^{k+}, t_{trip}^{k-}, t_{trip}^{k+} \right)_{k=0}^{K}, \frac{T}{f_i}, f_i \right) \tag{2}$$

$$z_{2+i} = \left(\left(v_k^{l_i}, t_{dwell}^{k-}, t_{dwell}^{k+}, t_{trip}^{k-}, t_{trip}^{k+} \right)_{k=K}^{0}, \frac{T}{f_i}, f_i \right) \tag{3}$$

for $i = 1, 2$. Train run z_{2+i} runs in the opposite direction of train run z_i, both being part of line l_i. The upper bounds of the trip time t_{trip}^{k+} and the dwell time t_{dwell}^{k+} are computed from the lower bounds by multiplying them with a (individual) constant factor. These time intervals will be used to compute flexible and stable plans (see section 2.3 and section 3 for details). The range of these intervals have to be adapted to the given track topology.

We use time dependencies to separate the departure events in the repetitions of train run z_i, $i = 1, 2$, during time period T by exactly $\frac{T}{f_i}$ minutes, namely

$$d = \left(z_i, z_i, dep_{z_i}(v_{k_j}^{l_i}), dep_{z_i}(v_{k_j}^{l_i}), r_m, r_{m+1}, \frac{T}{f_i}, \frac{T}{f_i} \right)$$

at each node of $v_{k_j}^{l_i}$, $0 \leq k_j \leq K$ on the rain run z_i and for each repetition r_m with $1 \leq r_m \leq (f_i - 1)$. Of course, we can also add some time tolerance in the departure times between repetitions.

We also introduce a certain service level for the length of travel times. This is achieved by defining time dependencies that force the travel times between the first and the last node of each line to be not longer than α (≥ 1) times the minimum travel time, i.e. for each train run z_i, $i = 1, 2$, we have

$$d = \left(z_i, z_i, arr_{z_i}(v_0^{l_i}), arr_{z_i}(v_K^{l_i}), r_m, r_m, \sum_{k=0}^{K} \left(t_{dwell}^{k-} + t_{trip}^{k-} \right), \alpha \left(\sum_{k=0}^{K} \left(t_{dwell}^{k-} + t_{trip}^{k-} \right) \right) \right) \tag{4}$$

and the same for the train runs in the opposite direction z_i , $i = 3, 4$,

$$d = \left(z_i, z_i, arr_{z_i}(v_K^{l_i}), arr_{z_i}(v_0^{l_i}), r_m, r_m, \sum_{k=0}^{K} (t_{dwell}^{k-} + t_{trip}^{k-}), \alpha(\sum_{k=0}^{K} (t_{dwell}^{k-} + t_{trip}^{k-})) \right) \tag{5}$$

for each repetition r_m with $1 \leq r_m \leq f_i$.

Our timetabling model TCFPESP (see section 2.3) therefore is flexible to adjust travel times between two consecutive nodes, but it must respect this restriction of the total travel time. This property can be used to make the timetable more robust (see Wüst et al. 2019).

Turnaround conditions at both ends of each line can also be implemented using time dependencies. We get each for train run z_i , $i = 1, 2$,

$$d = \left(z_i, z_{2+i}, dep_{z_i}(v_K^{l_i}), arr_{z_{2+i}}(v_0^{l_i}), r_m, r_m, \theta^-, \theta^+ \right)$$

$$d = \left(z_{2+i}, z_i, dep_{z_{2+i}}(v_0^{l_i}), arr_{z_i}(v_K^{l_i}), r_m, r_m, \theta^-, \theta^+ \right)$$

and each repetition r_m with $1 \leq r_m \leq f_i$. θ^- represents the (technical) minimum turnaround time needed at the end node $v_K^{l_i}$ resp $v_0^{l_i}$ of line l_i. θ^+ and can be used to control the minimum number of rolling stock needed to execute the timetable. In this case $\theta^+$is depending on the total trip times between the starting and the end node of the line and can be computed according to the approach described in Liebchen (2007).

Transfer conditions from the line planning step are natural candidates for connections, since they represent the travel chains of the passengers. In the line planning step we get a transfer possibility from train run z_1 to train run z_4 at common node $v_K^{l_1}$ (see assumption at the beginning of this section), but we don't fix explicitly at which concrete repetition of the lines these transfer should take place. The repetition is obvious, if the frequency f_i is equal to 1 for $i = 1, 2$ or if for some higher reasons the repetitions are fixed. We therefore distinguish two cases:

a) The repetitions of the train runs for the transfers are **known**:
   The transfer from train run z_1 at repetition r_m, $1 \leq r_m \leq f_1$, to train run z_4 at repetition r_n $1 \leq r_n \leq f_2$, takes place at node $v_K^{l_1}$ in at most θ^+minutes. With this assumption we generate the following SI-connection: $c = \left(z_1, z_4, v_K^{l_1}, r_m, r_n, , \theta^-, \theta^+ \right)$,
   The lower bound θ^- minutes is a minimum time needed for the transfer. This bound is given by the walking distance of the platforms of the connecting lines at the transfer node.
b) The repetitions of the train runs for the transfers are **not known**:

In this case we generate time dependencies in the SI, which configure the model in such a way that for a certain combination of repetitions of train runs z_1 and z_4 a feasible solution with connections can be found. This approach is described in Peeters (2003). We demonstrate this approach for the case of $f_1 = 1$ and $f_2 = 2$. The transfer therefore should take place between

repetition r_1 of train run z_1 and repetition r_1 of train run z_4 or
repetition r_1 of train run z_1 and repetition r_2 of train run z_4

This or-condition can be transformed into two time dependencies. This is possible mainly due to the consideration of a periodic timetable. The transfer should take place in the time interval $[\theta^-, \theta^+]$, where θ^- again corresponds to the minimum time needed for the transfer at the considered node.

In Peeters (2003) they give the following proposition. We assume that $(\theta^+ - \theta^-) \leq \frac{T}{f_2}$ holds. If the four time dependencies

$$d = \left(z_1,\, z_4,\, arr_{z_1}(v_K^{l_1}),\, dep_{z_4}(v_K^{l_2}),\, r_1,\, r_m,\, \theta^-,\, \theta^+ + \frac{T}{f_2}\right) \tag{6}$$

$$d = \left(z_1,\, z_4,\, arr_{z_1}(v_K^{l_1}),\, dep_{z_4}(v_0^{l_2}),\, r_1,\, r_m,\, \theta^- + \frac{T}{f_2},\, \theta^+ + T\right) \tag{7}$$

for $m = 1, 2$ are fulfilled, then exactly one of the repetitions r_1 and r_2 of train run z_4 allow a transfer from train run z_1 in the time interval $[\theta^-, \theta^+]$.

The assumption $(\theta^+ - \theta^-) \leq \frac{T}{f_2}$ is not too strong. If this assumption is not fulfilled we just could wait for the next repetition of the train instead of forcing a connection.

Approach b) above prevents the combination of the “wrong” repetitions a priori, which could lead to an infeasible SI (see also section 3).

2.3. Generation of the timetable in planning step (b)

For generating the timetable, we use a model for generating flexible periodic timetables based on a mesoscopic resolution of track infrastructure and safety configurations. This model is called TCFPESP and has been described in detail recently (see Wüst et al. 2018 and Wüst et al. 2019). The input to this model is given by the SI. If the SI is logically consistent and feasible with respect to the model configuration a timetable can be generated. For detecting conflicts we make use of the approach of Polinder et al. (2018). They can identify possible relaxations in an infeasible SI. In the case of maintenance work the result of the line planning step leads to an inconsistent SI. We did relax some of the computed transfers to get a feasible SI with the approach of Polinder (see section 3 for details). Besides the timetable itself the output also contains the train-track assignment. The output is described in more detail in Wüst et al. (2019). In this article we focus on the description of the input configuration for generating the SI and how this input configuration is generated from the result of the line planning step.

3. Real World Case Study

3.1. Delineation of the case study

To illustrate the line planning algorithm on a real-world example, we have selected a railway corridor in eastern Switzerland. Referring to the geographical location of the corridor we call the case study "Kerenzerberg" (see Figure 1). The infrastructure (i.e. the PTN), the minimum travel time and the line pools of the line categories are read out from the timetable valid in the year 2018. The OD-demand between the considered stations is constructed manually in such a way that the lines operate with the frequencies of the actual timetable. A total of 23836 passengers have to be transported in the considered hour. With this case study we want to demonstrate how we iteratively adapt a reference timetable to a timetable with maintenance work on a track section in the network corridor by re-executing line planning step (a) considering the restricted resource conditions. In this case, one of the two tracks of the section between the nodes Flums and Mels is completely blocked. With this constraint, a feasible temporary timetable in planning step (b) can only be constructed if the SI is relaxed in terms of the number or kind of operated lines. A change in the resulting line plan induces changes of the passenger flows. Based on this line plan as input a temporary timetable for the maintenance interval is calculated with the objective e.g. to minimally reduce the overall passenger travel time and at the same time respect the arrival and departure times of the reference timetable as much as possible.

3.2. Network segmentation, station and line categories

To avoid making timetable changes at locations that have no or negligible influence on the solution, it is important to identify which part of the entire rail network needs to be adapted and which part can be assumed to remain as specified in the reference timetable. Therefore, in a first step, the relevant lines of the subnetwork that will be directly affected by the construction sites must be identified. In a second step, all lines which are coupled (e.g., by transfers or technical dependencies) to the directly affected lines have to be identified.

The relevant rail network is subdivided into two subnetworks as shown in figure 1. First one identifies the subnet nodes that isolate the relevant infrastructure partitions from the periphery with fixed timetable times. In this way we separate the disaggregated subnetwork with the relevant infrastructure segments from an aggregated subnetwork with fixed timetable and infinite capacity. All train movements are planned in detail on the disaggregated subnetwork.

For each line coming from or going beyond the boundary nodes of the disaggregated subnet, we create a virtual end station node. In order to ensure that the different line categories stop at the right stations also in case of creating new lines that take reduced track capacity into consideration, stations are associated to line categories as shown in table 1. Furthermore we need this categorization to perform the 'system split' according to Oltrogge (1994).

Stops / Line categories		Zürich HB (ZHB)	Richterswil (RW)	Glarus (GL)	Uznach (UZ)	Ziegelbrücke (ZGB)	Mühlehorn (MH)	Tiefenwinkel (TIE)	Murg (MG)	Unterterzen (UNT)	Mols (MOL)	Walenstadt (WAL)	Flums (FMS)	Mels (MEL)	Sargans (SA)	Sewelen (SE)	Buchs (BU)	Maienfeld (MF)	Chur (CH)	**Line rotation time**
IC	**IC3**	x													x				x	178
	RJ	x													x		x			162
IR	**RE1**	x				x						x			x				x	182
	RE2														x		x		x	62
S-Bahn	**S2**	x	x			x														106
	S4				x	x	x	x	x	x	x	x	x	x	x	x	x			98
	S6			x	x	x									x			x	x	50
	S12																			42
	S25	x	x	x		x														134

Table 1: Stations, lines with their categories (IC (intercity), IR (interregio) and S-Bahn (commuter)and their turnaround times under normal operations.

The sum of all trip times along the line sections in both directions, the dwell times at the stations and the required turnaround times at the final stations results in the line rotation time which is indicated in the last column of Table 1. Because the planned maintenance work is located on the network section between stations Flums and Mels between the transfer nodes Ziegelbrücke and Sargans, we decided to use this corridor as a disaggregated partition of the test network. The western part of Ziegelbrücke and the eastern part of Sargans are aggregated. See Wüst et al. (2018a) for more details on the partitioning of the network.

1.1. Generation of a reference line plan and the corresponding timetable

We apply the line planning model $LineP$ (equation 1) to each line pool (IC, IR and S-Bahn) separately. We assume costs per line are equal, i.e. we minimize the sum of all frequencies of the lines selected by the line planning model.

The maximal transfer time is 10 minutes between the lines. The computed frequencies of the model $LineP$ for all lines are the same as in the timetable valid in the year 2018. The frequencies of the lines passing the corridor between Ziegelbrücke and Sargans are shown in figure 4a.

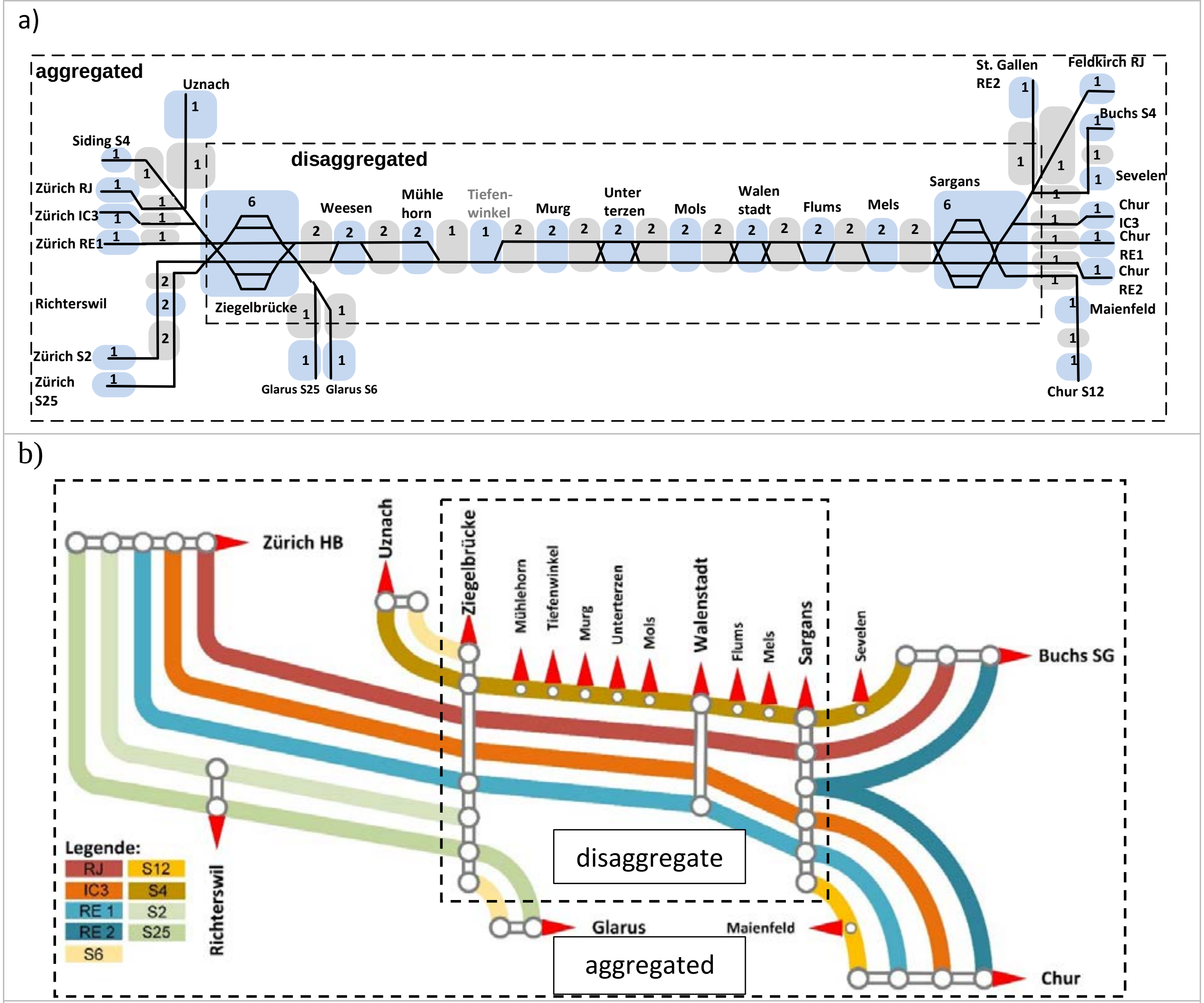


Figure 1: Network partition in case study 'Kerenzerberg'. a) PTN and track capacities. b) total set of line pools on the PTN in a). RJ, and IC3 are intercity lines, RE1, RE2, S2 and S25 are IR lines, S6, S4 and S12 are commuter lines

We combined all selected lines of all line categories from the line pools in one CGN (see figure 2). The thickness of the lines in figure 2 corresponds to the passenger loads. The white nodes represent the individual stations of the network partition. These are connected to the transfer nodes of the lines that connect these stations by the transfer edges. F1 to F30 represent the driving edges. 40 transfer edges between lines at different stations and a total of 2520 transfer passengers resulted from the line planning step (see figure 4 and table 2 for important transfer edges in Walenstadt).

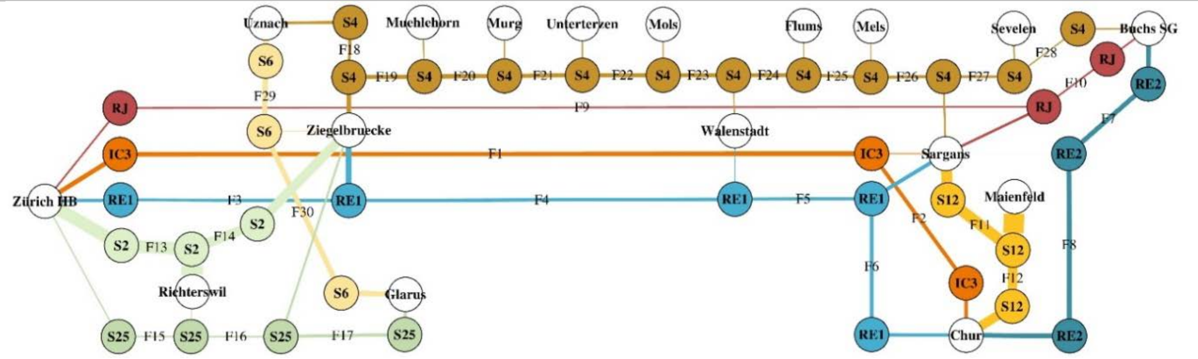


Figure 2: CGN for case study 'Kerenzerberg'. Line colours correspond to line types in Figure 1. Thin lines connecting station nodes with line transfer vertices indicate boarding and deboarding edges. The thickness of the line edges is weighted with the passenger volume on the edge.

By using the approach described in section 2.2 we generated the SI. For the upper bounds of travel and dwell times in (2) and (3) we multiplied the minimum times with 1.5. For the time dependencies in (4) and (5) we set α equal to 1.2. We parametrized the TCFPESP with this SI and could generate a feasible, reference timetable (see figure 5). We could therefore guarantee all (40) connections of the SI within a maximum of 10 minutes.

1.2. Generation of line plan and a timetable for a time period with maintenance work

Restricting the section between Flums and Mels due to a site-specific track blocking results in only one track available. We want to generate a timetable for this timetable period with maintenance work, but we only admit a time tolerance of +/- 3 minutes with respect to the departure times of the reference timetable. With this restriction and the reduced infrastructure the TCFPESP becomes infeasible. In figure 5 one can see that there are two crossings of lines between Flums and Mels, such that the reference timetable itself is not feasible any more. Next we tried to relax the constraints with the approach of Polinder et al. (2018). The effect on the quality of the timetable with respect to passenger travel times was not acceptable. Therefore we decided to go back to planning step (a, line planning). We had to change the line pool in order to reduce the number of lines crossing the section between Flums and Mels.

One solution, for example, is to delete the edge F25 of the lowest category S4 between Flums and Mels (see figure 2) and introduce two new independent line fragments S4.1 and S4.2. These new lines now operate between Uznach and Flums in the western part and between Mels and Buchs in the eastern part. Hence, line S4 no longer crosses the affected section. We used the given OD-matrix and the new (reduced) line pool of the commuter lines to resolve the line planning model $LineP$ (with the same costs and transfer times as in the reference case). Except for line S4 all other lines operate with the same frequencies as before. The new fragment lines S4.1 and S4.2 operate with frequency 1 (see Figure 4). We compare the line planning output of the reference case for the normal operation and the operation with the maintenance interval.

- **Passenger flows**: Since the S4 in the maintenance interval no longer runs between Flums and Mels, some passengers are forced to change in Sargans or Walenstadt. This can be seen from the line widths of the individual edges in figure 4, which are scaled with the number of passengers. Those passengers, who have used the S4 for transport between Flums and Mels in normal operation, change to the line RE1 in construction site operation. This is illustrated in Figure 4 for the reference line plan (a) and can be compared to the line plan for the maintenance interval (b).

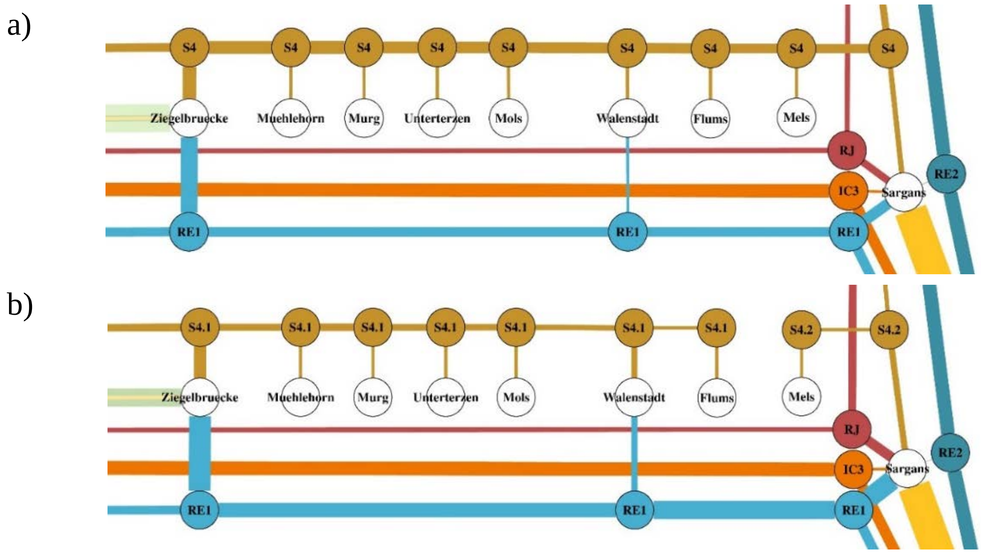


Figure 3: Passenger flow in the case of the reference line plan (a) and in case of period with maintenance work (b)

- **Transfer connections**: In figure 4 the passenger flows are illustrated across lines and nodes. Additionally, the transfer edges can be identified. 40 transfer edges between lines at different stations and a total of 2520 transfer passengers resulted from the line planning step under normal operations. For the maintenance line plan we get a total of 51 transfer edges with 3806 transfer passengers. The transfer edges are an important input to timetabling to specify connections between train runs (see section 2.2). In table 2 these are shown by the example of Walenstadt station. In parentheses, the predecessor or successor stations are indicated. The increase from a total of 132 transfers in normal operation to a total of 444 transfers in the maintenance interval represent the impact of the construction site on the transfers in Walenstadt to the RE1.

Walenstadt (normal operations)		
From	**To**	**Number of passengers**
RE1 (Ziegelbrücke)	S4 (Flums)	40
S4 (Flums)	RE1 (Ziegelbrücke)	40
RE1 (Sargans)	S4 (Mols)	26
S4 (Mols)	RE1 (Sargans)	26
Total:		**132**

Walenstadt (operation period with maintenance work)		
From	**To**	**Number of passengers**
RE1 (Sargans)	S4 (Mols)	134
S4 (Mols)	RE1 (Sargans)	134
RE1 (Sargans)	S4 (Flums)	68
S4 (Flums)	RE1 (Sargans)	68
RE1 (Ziegelbrücke)	S4 (Flums)	20
S4 (Flums)	RE1 (Ziegelbrücke)	20
Total:		**444**

Table 2: transfer connections for line plan with normal operations (left) and operation with maintenance work (right)

- **Operation costs**: Using line rotation times and line frequencies and the fixed costs of the line, the operating costs could be calculated. For the duration of the maintenance time window thus any additional costs could be determined. As mentioned above, in our calculations, we assumed cost rates to be equal. We therefore refrain from presenting the costs.

From the output of the line planning step in the maintenance interval we generate again the SI according to section 2.2. The factors in (2)-(5) are the same as in the reference case. As described before we only admit a time tolerance of +/- 3.75 minutes with respect to the departure times of the reference timetable. We added these constraints to our TCFPESP model. In a first run the TCFPESP model became infeasible. We used again the approach of Polinder (2018) to determine how much we have to loosen the constraints to become feasible. We admitted only the connection constraints related to (6) and (7) to be relaxed. Furthermore we used the transfer passengers as weights in the objective of the model of Polinder. The result was to relax 6 of the 51 transfers times related to constraints (6) and (7) (one to 17, 4 to 33 and one to 55 minutes). Since this only affected 226 of 3806 transfer passengers, we accepted the timetable (see figure 5 b)).

In figure 5 we can compare the two timetables. The conflicts between Flums and Mels disappeared due to the new lines S 4.1 and S 4.2. Furthermore all the lines without line S4 are in a time band of +/- 3.75 minutes compared to the reference case.

To assess the convenience impact for passengers one can either calculate the increase in travel time for all passengers or the travel time of those who are affected by the construction site otherwise. Passengers concerned are those who travel in normal operation with the S4 between Flums and Mels (edge F25 in figure 2). For all passengers the weighted travel times with the maintenance timetable is only 1% higher than with the reference timetable. For the affected passengers the overall increase in the weighted travel time amounts to 24%. In Table 3 we illustrate travel times for some selected origin destination combinations.

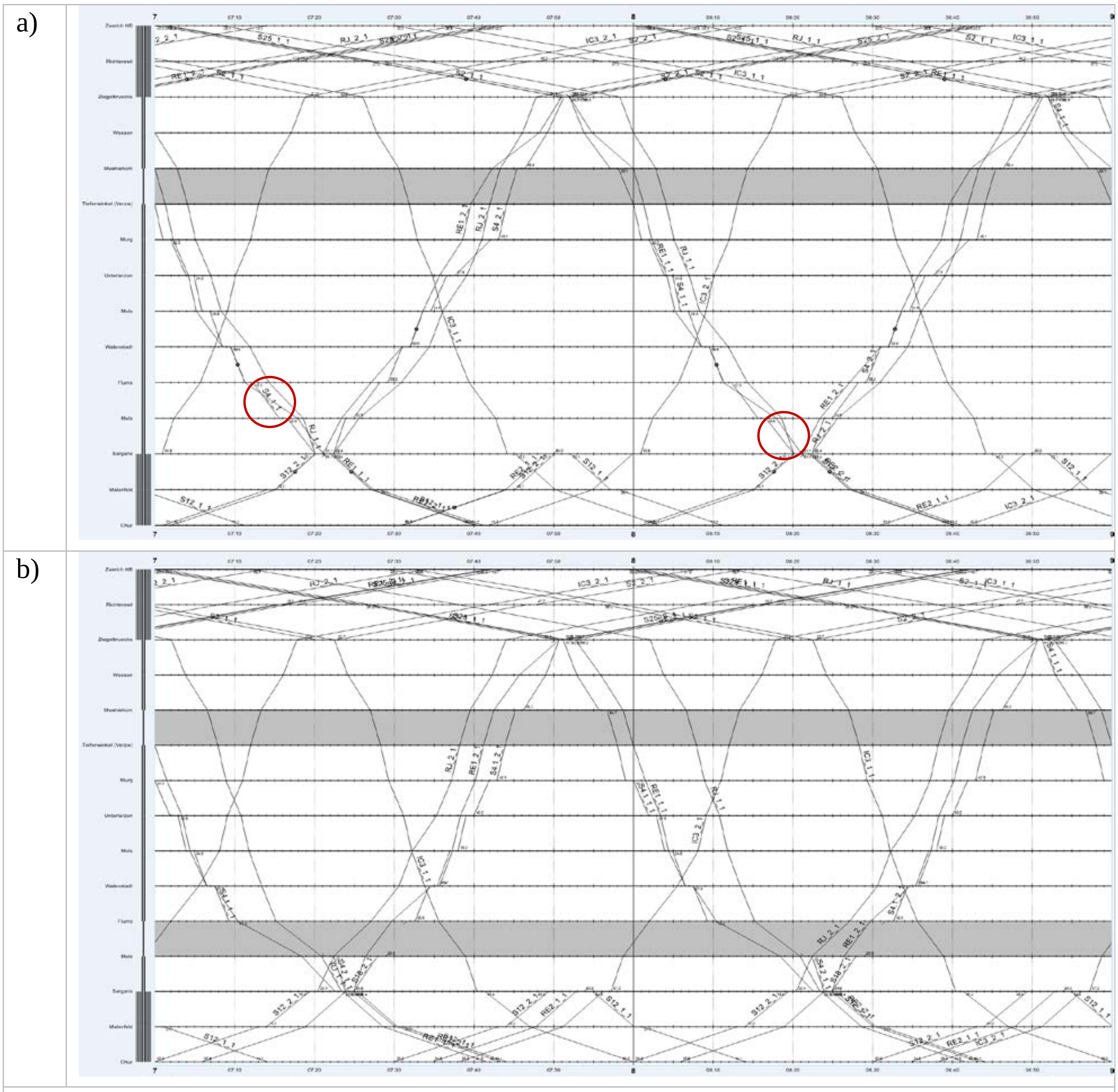


Figure 4: Graphical timetable for Korridor 'Kerenzerberg'. a) normal operations, red circles indicate crossing trains between Flums and Mels b) operations with track maintenance work between stations Flums and Mels. No train crossings occur on this section any more

Travel times				
From	**To**	**Reference timetable**	**Timetable with maintenancemaintenance interval**	**Extension**
Flums	Mels	5	57	1020%
Ziegelbrücke	Mels	30	37	23%
Chur	Flums	28	64	156%
Zürich HB	Chur	89	92	3%

Table 3: Selected travel times for both line planning scenarios

2. Summary and Conclusions

Line planning is a fundamental step in the creation of a transport service. By automating this planning step, different scenarios can be compared with each other within short time. From these scenarios cost-effective timetables can be derived. Above all, schedule deviations due to construction sites and disruptions are virtually unmanageable in the multitude of operations. In cooperation between SBB and ZHAW, a process for automated line planning was developed, which provides fast solutions to such capacity limitations. The deletion of the edge F25 (i.e., the division of the S4 into two sub-lines) was still done manually and justified with the small priority of the S4. Currently, the project team is also working on a method for Line pool generation, which will also automate this step in the future (see Gattermann et al., 2017).

The case study "Kerenzerberg" illustrates how the affected network partition can be divided into an aggregated and disaggregated subnetwork and how the use of Change & Go network resp. the line planning model produces the SI for the timetabling model. The results are highly dependent on the quality of the OD-matrix, which should be continually improved by surveying customer movements.

The results of the line planning are not limited to the line frequencies and their associated travel times and operating costs. By supplementing the results with the passenger flows and transfer connections, an important added value for ensuring line transfers in the timetable planning step could be generated.

Literature

Barber F, Ingolotti L, Lova A, Marin A,Mesa J, Ortega F, Perea F, Tormos P (2008). Integrating timetabling, network and line design. Technical report, ARRIVAL project

Caimi, G., (2009). Algorithmic decision support for train scheduling in a large and highly utilised railway network. Diss. ETH Zürich Nr. 18581

Caimi, G., Fuchsberger, M., Laumanns, M., Schüpbach, K. (2011). Periodic railway timetabling with event flexibility. Networks, vol. 57, no. 1, pp. 3-18

Friedrich, M., Hartl, M., Schiewe, A., Schöbel, A. (2017). Integrating passengers' assignment in cost-optimal line planning. 17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017), no. 5, pp. 1-16, (DOI: 10.4230/OASIcs.ATMOS.2017.5),

URL: http://drops.dagstuhl.de/opus/volltexte/2017/7901

Gattermann, Ph., Harbering J., Schöbel, A. (2017). Line pool generation. Public Transport vol. 9: pp. 7–32, DOI 10.1007/s12469-016-0127-x

Liebchen, C., Möhring, R. (2007) The modelling power of the periodic event scheduling problem: railway timetables—and beyond. In: Algorithmic methods for railway optimization. Lecture notes in computer science, vol 4359, pp 3–40. Springer, Berlin

Oltrogge, C. (1994) Linienplanung für mehrstufige Bedienungssysteme im öffentlichen Personenverkehr. PhD thesis, TU Braunschweig (in German)

Peeters, L. (2003). Cyclic railway timetable optimization (PhD thesis). Erasmus University Rotterdam. Accessed via: https://www.researchgate.net/publication/254804709_Cyclic_Railway_Timetable_Optimization

Polinder, G.-J., Kroon, L.G., Aardal, K., Schmidt, M.E., Molinaro, M., (201. Resolving infeasibilities in railway timetabling instances, ERIM Report Series Research in Management ERS-2018-002-LIS, Erasmus Research Institute of Management (ERIM)

Schöbel, A. (2012). Line planning in public transportation: models and methods. OR Spectrum (2012) 34:491–510, DOI 10.1007/s00291-011-0251-6

Schöbel, A. (2017). An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. Transportation Research Part C: Emerging Technologies, vol. 74, pp. 348-365.

Wüst, R., Laube, F., Roos, S., Caimi G. (2008). Sustainable global service intention as objective for controlling railway network operations in real time. Proceedings of the 8th World Congress of Railway Research (WCRR), Seoul, Korea.

Wüst, R. M., Bütikofer, S., Ess, S., Gomez, C, Steiner A., Laumanns, M., Szabo, J., (2018). Periodic timetabling with 'Track Choice'-PESP based on given line concepts and mesoscopic infrastructure. In: Operations Research Proceedings 2018, In press, Springer, Berlin/Heidelberg, Germany

Wüst, R. M., Bütikofer, S., Ess, S., Gomez, C, Steiner A., 2019. Improvement of maintenance timetable stability based on iteratively assigning event flexibility in FPESP. In: Planning, Simulation and Optimisation Approaches, 8th International Conference on Railway Operations Modelling and Analysis Norrköping, Sweden, Conference Proceedings (accepted submission), www.railnorrkoping2019.org