Journal of Rail Transport Planning & Management 6 (2016) 99–115

Contents lists available at ScienceDirect

Journal of Rail Transport Planning & Management

journal homepage: www.elsevier.com/locate/jrtpm

Capacity analysis of railway lines in Germany – A rigorous discussion of the queueing based approach

Norman Weik*,[1], Nora Niebel, Nils Nießen

Institute of Transport Science, RWTH Aachen University, 52062 Aachen, Germany

A R T I C L E I N F O

Article history:
Received 11 April 2016
Received in revised form 16 June 2016
Accepted 17 June 2016
Available online 28 June 2016

Keywords:
Railway capacity analysis
Knock-on delays
Delay propagation
Queueing
Railway infrastructure planning

A B S T R A C T

The operation of railway systems requires detailed information on infrastructure capacity. A major challenge, especially in long-term planning, is assessing the quality of operations given very limited information on schedules is available. To this end, analytical models based on a stochastic description of railway systems have found widespread application. We discuss a model for the capacity analysis of railway lines relying on single channel queueing systems. By identifying knock-on delays with waiting times delays can be estimated using methods from stochastics and queueing theory. Mean knock-on delays are used as a quality-dependent indicator of capacity allowing to determine the admissible number of trains for a prescribed level of service.

Though being widely used in Germany the model has not been made fully available to the research community. In this paper two main contributions are made: A new, mathematically rigorous derivation of the pivotal "Strele Formula" for the estimation of knock-on delays, which is based on the convolution of delay distribution functions, is provided. Unlike existing discussions our approach is valid for general independent buffer times. Additionally, we critically review the model assumptions and investigate the "triangular gap problem", an overestimation of capacity resulting from the model's limitation to pairwise correlations.

1. Introduction and literature overview

Concise knowledge of the capacity of railway infrastructure is vitally important for the planning, management and operation of railway systems. Depending on the process stage capacity analysis aims to investigate the ability to satisfy the demands of railway operators in scheduling (Wendler, 2007) or the stability and robustness of schedules in operations (Goverde, 2005). Input data ranges from fragmentary information about prospective operations to fully constructed schedules. Time criticality enters as a major challenge if dispatching and rescheduling are to be optimized with respect to capacity (cf. Cacchiani et al., 2014; Törnquist, 2006; Narayanaswami and Rangaraj, 2011; Weymann and Nießen, 2015). The vastness and complexity of the task – as well as country-specific layout and operations of railway networks – have therefore evoked a multitude of different approaches to determine the feasible number of trains in railway systems.

* Corresponding author.
E-mail address: weik@via.rwth-aachen.de (N. Weik).
[1] URL: http://www.via.rwth-aachen.de.

http://dx.doi.org/10.1016/j.jrtpm.2016.06.001

Arguably, the most common approach is to use track-occupation as an indicator of capacity usage. An important representative of this class of models is schedule compression documented in UIC Code 406 (UIC, 2013). It provides a way to determine idle line capacity based on the schedule's degree of utilization. However, results are hard to compare since track utilization is highly dependent on the structure of the schedule. In addition, the method provides only limited insight into the quality of operations and the schedule's stability in case of disruptions.

In order to determine capacity in relation to a prescribed level of service that operations are required to meet, different indicators such as punctuality or the magnitude of delays have to be considered. This can be done both from the network manager's point of view aiming to minimize overall irregularities (Schwanhäußer, 1974) and from a passenger's perspective maximizing travel utility (Fosgerau and Karlström, 2010). Suggested methods cover microscopic (Huerlimann, 2001; Radtke and Bendfeldt, 2001; Janecek and Weymann, 2010) and macroscopic simulations (Büker and Seybold, 2012) as well as linear and mixed integer optimization (Caimi, 2009) aiming to minimize delay-related objective functions (Oliveira and Smith, 2000; D'Ariano et al., 2007, 2008; Corman, 2010) or to maximize travel utility (Norio et al., 2005). The applicability of both approaches is limited to applications where detailed knowledge about scheduled operations can be used. If input data is scarce and only coarse information about planned operations is provided a large number of problem instances has to be simulated or solved in order to determine the expected capacity. This usually makes these approaches very time-consuming and hence ineffective for early planning stages.

Analytical stochastic or queueing based approaches are heavily used in this area. In the first case, propagation of delays is determined by convoluting probability distribution functions of delays (Büker and Seybold, 2012; Yuan and Hansen, 2007). In the queueing based approach railway lines and junctions are represented as queueing systems, trains corresponding to requests for the usage of some infrastructure-related service channels (cf. Nießen, 2014; Fischer and Hertel, 1990 for an overview). Service times correspond to headway times between train runs. While the stochastic nature of these models prohibits the exact individual representation of schedules they are well suited to quickly derive the overall behavior and stability of operations. Moreover, they only require very limited data input: Knowledge of the probability distributions of primary delays, minimum headway times and the projected number of trains and train types is required, but no planned schedule has to be known. In particular, an infrastructure-centered capacity indicator can be derived which allows to compare different system designs regardless of the precise operation concept. Capacity can either be assessed based on the absolute waiting times (delays) (Schwanhäußer, 1974; Nießen, 2014) or − in addition − by considering their sensitivity with respect to the number of trains (Hertel, 1992).

We subsequently focus on an analytical queueing based approach to determine the capacity of railway lines and nodes based on the height of knock-on delays. Knock-on delays refer to delays resulting from signalling and safety constraints, dispatchers' resolution of track occupation conflicts, connection conflicts or fleet rostering conflicts caused by trains operating behind schedule. The presented framework has originally been proposed by Schwanhäußer more than 40 years ago (Schwanhäußer, 1974), yet software tools based on the approach still build the backbone of capacity analysis in Germany. Although related queueing based approaches have been discussed in the context of timetable quality estimation (Wendler, 2007; Wakob, 1985; Hertel, 1984) Schwanhäußer's pioneering work on the stability of operations has not been made fully accessible beyond the German-speaking region, so far. While Nießen (Nießen, 2014) gives a short overview of the approach and the calculation of knock-on delays in general, his focus is on determining a quality measure based on the amount of knock-on delays. An in-depth discussion of the model assumptions and the derivation of the pivotal Strele Formula for the calculation of knock-on delays is missing.

The goal of this paper is to close this gap by presenting a generalized, mathematically rigorous derivation of the Strele Formula. Our approach, which is based on the convolution of delay distribution functions, provides an efficient and elegant way to determine knock-on delays and avoids a time-consuming case by case analysis of delay propagation effects between trains. Instead, it only requires the calculation of certain conditional expectations which can be easily standardized and executed numerically. It also goes beyond previous presentations (Schwanhäußer, 1974, 1994) as it is applicable for general independent buffer times, such that the method can easily be adapted to the needs of specific railway system and timetable characteristics.

We critically discuss modelling assumptions and limitations of the approach. In particular, a lack of exactness on line segments with intersections resulting from the restriction to pairwise correlations between trains and giving rise to the so-called "triangular gap problem" is investigated in Section 3.

2. Method

2.1. Basic concept

The fundamental characteristic of the method is that it aims to put railway lines to the test individually (Schwanhäußer, 1974). Interactions between different railway lines and junctions arising from their embedding in a larger network of lines and junctions are disregarded. While this course of action limits the precision of the analysis it facilitates comparability between different infrastructure layouts and radically reduces computational effort. The original work's focus was on railway lines (Schwanhäußer, 1974), yet it has been extended to application in the dimensioning of station threads (Schwanhäußer, 1978; Nießen, 2008, 2013). In our discussion we focus on the modelling of railway lines. Extensions such as the application to railway junctions are briefly sketched in Section 2.6.

The fundamental concept of the approach is to model railway lines as queueing systems with a single service channel (cf. Fig. 1). This assumption does not mean the model's applicability is restricted to single track railway lines. It extends to lines