Journal of Rail Transport Planning & Management xxx (2017) 1–13

Contents lists available at ScienceDirect

Journal of Rail Transport Planning & Management

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

A quasi-birth-and-death process approach for integrated capacity and reliability modeling of railway systems

Norman Weik*, 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 12 June 2017
Accepted 24 June 2017
Available online xxx

Keywords:
Railways
Infrastructure planning
Capacity
Reliability
Performance modelling
Queueing
Quasi-birth-and-death-processes

A B S T R A C T

A railway system's capacity is an important performance indicator allowing to assess different infrastructure variants and to devise market-compliant schedules. Existing approaches in capacity analysis assume the unrestricted availability and peak performance of all system components. Disruptions leading to infrastructure unavailability and reduced system performance are not considered in long- and medium term tactical planning of capacity.

We present a quasi-birth-and-death process approach for the integrated modelling of capacity and reliability. By allowing for phase-type distributed arrival, service and repair processes the model permits to describe a wide range of schedule and operational characteristics. At the same time, the solvability of Markovian processes and the information on the queue length distribution are preserved. The model is solved using a Krylov-subspace method, which allows to effectively deal with large state spaces and transition matrices.

The approach is compatible to existing queueing-based models in the capacity analysis of railway lines and junctions. The functionality of the method is demonstrated in a case study of a mixed service railway line with infrastructure unavailability.

1. Introduction and literature review

A railway system's capacity is generally viewed as the maximum number of trains which can be operated concurrently with market-compliant quality in a predefined time frame (UIC 406, 2013). While specific requirements for the level of service vary, capacity remains an important performance indicator allowing to assess different infrastructure variants and to devise market-compliant schedules. The most widespread approaches to measure capacity are either based on infrastructure utilization, e.g. UIC schedule compression method (UIC 406, 2013), or on quality-related criteria such as train punctuality (Graffagnino and Labermeier, 2016) or delays (e.g. Huisman and Boucherie, 2001; Schwanhäußer, 1974; Hertel, 1992; DB Netz AG, 2008a). Depending on the planning stage and input data availability various approaches to determine the feasible number of trains have been described in the literature.

Our focus is on long and medium term tactical planning where infrastructure topology and basic operational characteristics are fixed, but the exact schedule is not known or may still change. This is where analytic stochastic and queueing based approaches have found widespread application in modelling the expected load and waiting time.

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

http://dx.doi.org/10.1016/j.jrtpm.2017.06.001
2210-9706/

Potthoff introduced a technique to determine the number of tracks required in stations based on the loss probability in $M/D/n/0$-queueing systems (Potthoff, 1962). Hertel (1984) later proposed a method based on $GI/GI/n/\infty$-models to estimate waiting probabilities in stations for the same task.

For capacity analysis of railway lines Schwanhäußer (1974) developed an approach to approximate the average knock-on delays based on the probability distributions of primary delays and buffer times. The method is based on pairwise correlations between trains and the identification of railway lines with $M/D/1/\infty$-queueing systems. The method has recently been revisited in Weik et al. (2016), where model assumptions and limitations are discussed and a mathematically more rigorous derivation is given. An extension of Schwanhäußer's method to station thresholds has been discussed in Nießen (2013). Effective single-channel queueing systems are constructed where service times are reduced according to the share of mutually non-exclusive train runs on the infrastructure (Nießen, 2013). It is also applicable to bottleneck analysis based on a deconstruction of station thresholds into sectional route notes, where all train runs are mutually exclusive (Schwanhäußer, 1994).

Huisman and Boucherie (2001) also establish a relation between train delays and the utilization of railway lines. Here, railway lines are modelled as infinite server resequencing queues. Delays are due to different running times of different train types, which can be taken to be either deterministic or random, hence allowing to consider primary delays within track segments. The model additionally allows for correlations between interarrival times or between interarrival and service times.

Wendler (2007) proposes a queueing model exhibiting more general, Semi-Markovian service processes, which can be used to determine the scheduled waiting times in capacity allocation. The model is closely related to models used to assess runway utilization in view of different aircraft types in aviation (Bäuerle et al., 2007). For even more general $GI/GI/1/\infty$-queueing networks Wakob (1985) developed an approximation approach to determine the average scheduled waiting times in station threads. The method is built on a statistical regression model for the waiting times based on the first two moments of the arrival and service process. It has been investigated in detail and validated against empirical data by De Kort et al. (1999).

In Huisman et al. (2002), a Jackson queueing network model for the joint analysis of multiple railway lines has been proposed. The railway network is decomposed into station entries, station exits, and line segments, each being modelled as $M/M/1/\infty$-queueing systems, which ensures the factorization of the stationary probability distribution. Unfortunately, the independence property of different queues in Jackson networks is quickly destroyed once correlations between different queues enter.

Analytical models are particularly suited to cope with uncertain or fragmentary input data as they operate on (empirical) probability distributions or moments of probability distributions. Alternative approaches such as traffic simulations or MIP-based approaches require solving a large number of system realizations in order to obtain statistically reliable results. However, existing analytical approaches generally lack precision for two main reasons: First, in order to obtain solvable models, the exactness of either input or output data is reduced. In the first case, arrival and service processes are approximated by more easily tractable processes, e.g. Markovian ones (Schwanhäußer, 1974; Potthoff, 1962; Huisman et al., 2002). In the latter case, more general arrival and service times can be considered, but the output is limited to mean value data and information on the distribution of waiting times is lost (Hertel, 1984; Wendler, 2007). Second, the unrestricted availability of all system components is assumed. Disruptions such as train malfunctions or infrastructure breakdowns can only be considered implicitly as far as they affect the distribution of arrival or service times. Still, they are very important for practical capacity investigations as complex systems such as railway networks are constantly subject to failure and maintenance processes limiting the effectively achievable capacity. A promising line of research to include failure and repair processes in Markovian railway capacity analysis models has been discussed in Bär et al. (1988), but has not been further pursued.

The topic of integrated reliability and performance analysis of railway systems has recently received new attention in reliability analysis and asset-management. In Fecarotti et al. (2013), the resilience of operations on a generic railway line with varying track switching possibilities is studied and optimized. To this end, a systematic failure mode analysis using FMECA is performed and a discrete event simulation describing train operations is employed (Fecarotti et al., 2013). More recently, the train operation simulation model has been exchanged for a Petri-net based approach in Fecarotti et al. (2015). Still, reliability and performance modelling are separated. A list of timed failure events is generated by a reliability subroutine which is then fed to the simulation of train operations. This can be seen as a performance simulation in a random system environment, yet it prohibits the modeling of load-dependent failures such as train malfunctions.

Our present work picks up on Bär et al. (1988) and complements Fecarotti et al. (2013, 2015) in aiming to provide a more realistic representation of railway systems by an integrated modelling of system availability and performance. Unlike in the models described in Bär et al. (1988), which only allow to consider exponential holding times, a Quasi-Birth-and-Death (QBD) process approach is adopted. By allowing for phase-type distributed arrival and service times, which can be fitted to the moments of given empirical data, this concept provides the flexibility to adequately describe a wide range of schedule and operational characteristics. At the same time, the tractability of Markovian models is preserved and the stationary distribution can be obtained. This not only allows to consider the average capacity of the given railway infrastructure, it also allows to determine the probability that the system performs at a given capacity. This information is highly relevant for infrastructure operators which have to ensure contractually defined standards are held.

In the context of railway operations science the utility of phase-type distributions has been demonstrated before. In Meester and Muns (2007) it has been shown that they are well-suited to model delay propagation in railway networks as they provide a good fit of delay distributions and are closed under the mathematical operations governing delay propagation. Büker and Seybold (2012) build on a similar modelling of delay distributions. In addition, an activity-based framework to formalize delay propagation processes is introduced, which allows to analyze large networks like, for instance, Switzerland.