ELSEVIER

Discrete Optimization

A solvable queueing network model for railway networks and its validation and applications for the Netherlands

Tijs Huisman [a,*], Richard J. Boucherie [b], Nico M. van Dijk [a]

[a] *Faculty of Economics and Econometrics, Department of Operations Research, University of Amsterdam, Roetersstraat 11,
1018 WB Amsterdam, Netherlands*
[b] *Faculty of Mathematical Sciences, Department of Discrete Mathematics, Operations Research and Statistics, University of Twente,
P.O. Box 217 7500, AE Enschede, Netherlands*

Abstract

The performance of new railway networks cannot be measured or simulated, as no detailed train schedules are available. Railway infrastructure and capacities are to be determined long before the actual traffic is known. This paper therefore proposes a solvable queueing network model to compute performance measures of interest without requiring train schedules (timetables). Closed form expressions for mean delays are obtained. New network designs, traffic scenarios, and capacity expansions can so be evaluated. A comparison with real delay data for the Netherlands supports the practical value of the model. A special Dutch cargo-line application is included.
© 2002 Published by Elsevier Science B.V.

Keywords: Transportation; Railways; Queueing networks

1. Introduction

1.1. Motivation

In Western Europe, the demand for transportation is rapidly growing. Stimulated by governments, in response to the increasing congestion on the roads, which is undesirable for both economical and environmental reasons, railway companies are working hard to participate in this growth. Train frequencies are increased and new train services are introduced, like high speed and light rail trains [6].

* Corresponding author. Present address: Railned B.V., Department of Innovation, P.O. Box 2101, 3500 GC Utrecht, Netherlands.
Tel.: +31-30-2354333.
E-mail address: t.huisman@railned.nl (T. Huisman).

The increased rail traffic will require extensions of the existing networks in order to keep the operationable planning feasible and most importantly to keep delays within acceptable limits. These extensions have to be planned long in advance, as the design and construction of new railway infrastructure is a most time consuming process, which may take up to 20 years. Moreover, as railway infrastructure is extremely expensive, it should be capable of handling railway traffic during several decades.

For such a planning horizon of one or two decades only rough estimates can be provided for the traffic demand and corresponding train frequencies. Certainly, detailed timetables are not yet available. It would thus be desirable to have an analytical tool for railway networks not requiring timetables but just train frequencies as input. Such a tool could be most powerful in the first stages of design, to identify bottlenecks in the network, to compare alternative designs in a global way, or to analyse several traffic scenarios.

Some stochastic models along this line have been developed recently (see [8,12,13], also see [7] for an overview of applications of performance models for railway networks). These however all concern specific components. The results in [12] are also approximative.

In [12] a railway station is considered as a circuit switch loss model, so that closed formulas for delay probabilities can be derived; [8] provides running time distributions and delays on railway sections by translating the model to a resequencing queue, while [13] is concerned with calibration of standard single node queueing models to fit railway characteristics. Thus far, however, only parts of a railway network have been considered in isolation.

1.2. Objective

The main objective of this paper in contrast is to develop an analytically tractable queueing *network* model for total railway networks, taking into account dependencies and interaction between the individual components.

1.3. Results

The starting point for this model is the usual partition of a railway network into stations, junctions, and sections. By a careful definition of these components, the railway network will be transformed into a so-called product form queueing network. This product form result justifies a decomposition of the network in its components, which in turn justifies a more detailed analysis of components in isolation.

Clearly, to obtain this result, some simplifying and modifying assumptions have to be made. The components of the network are modeled in a rather global way, although not very different from existing railway models, as in [10]. Despite these practical concessions, however, a comparison of computed mean delays and practical delay data for 1997/1998 even at component level appears to be highly supportive for its practical usefulness.

Summarizing, the contribution of this paper is twofold:

- a queueing model for long term performance computations of railway networks,
- a justification for and to which extent a decomposition of the railway network in its detailed components.

1.4. Outline

The outline is as follows. Section 2 describes the lay-out of a railway network and transforms the railway network into a queueing network. The steady-state distribution of this network is obtained in

Section 3, as well as closed form expressions for various performance measures. Section 4 validates the model by a comparison with practical delay data. Next, the implications of a specific new cargo-line application that is still of great practical interest in the Netherlands are studied as application. Section 5, finally, discusses the assumptions of the model, and provides suggestions for further research.

2. The model

A railway network consists of three types of components: *stations*, *junctions*, and *section tracks*, that connect the stations and junctions (Fig. 1). We denote by

$\mathscr{S}$: the set of stations in the network,
$\mathscr{J}$: the set of junctions,
$\mathscr{T}$: the set of section tracks,
$\mathscr{I}$: the set of all infrastructural components in the network; that is $\mathscr{I} = \mathscr{S} \cup \mathscr{J} \cup \mathscr{T}$.

The set of train classes in the network is denoted by $\mathscr{C}$. A train class $c \in \mathscr{C}$ is characterised by its route through the network, denoted by the sequence of all components it passes:

$$\mathbf{r}_c = (i_c(1), \ldots, i_c(s_c)),$$

with $i_c(j) \in \mathscr{I}$ the component of the network that a class c train passes on the jth stage of its route, and s_c the number of stages of its route. We will now describe the components in more detail.
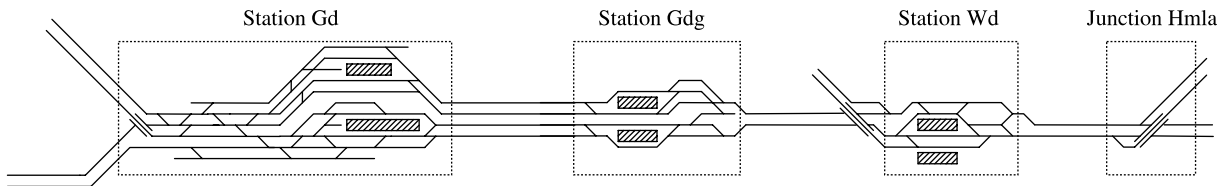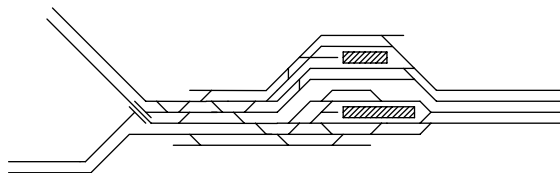

Fig. 1. A part of the Dutch railway network.

Fig. 2. The lay-out of a station.

2.1. Stations

The general lay-out of a station consists of a set of tracks, that are connected by switches and crossings (Fig. 2). We distinguish *halting tracks*, where trains stop to enable passengers to board and alight, and *storing tracks*, where trains wait after completion of their service until their next service starts.

To enter the station a train requires simultaneously a halting track, and its route to that track (Fig. 3(a)). If no track is available, or part of this route is occupied by other trains, a train has to wait on the section track before the station until both a halting track and its route to that track are free. At the moment a train has arrived at its halting track, its route to that track is released and only the halting track is occupied during its halting time (Fig. 3(b)). After its stop, a train either continues its route through the network (Fig. 3(c)), or moves to a storing track if its service is completed (Fig. 3(d)). In both cases the corresponding route on the station should be free, otherwise a train will have to wait on its halting track. Furthermore, at a station trains may start their route through the network. Then they are moved from a storing track to a

(a) Entering

(b) Halting

(c) Departure

(d) Ending a route
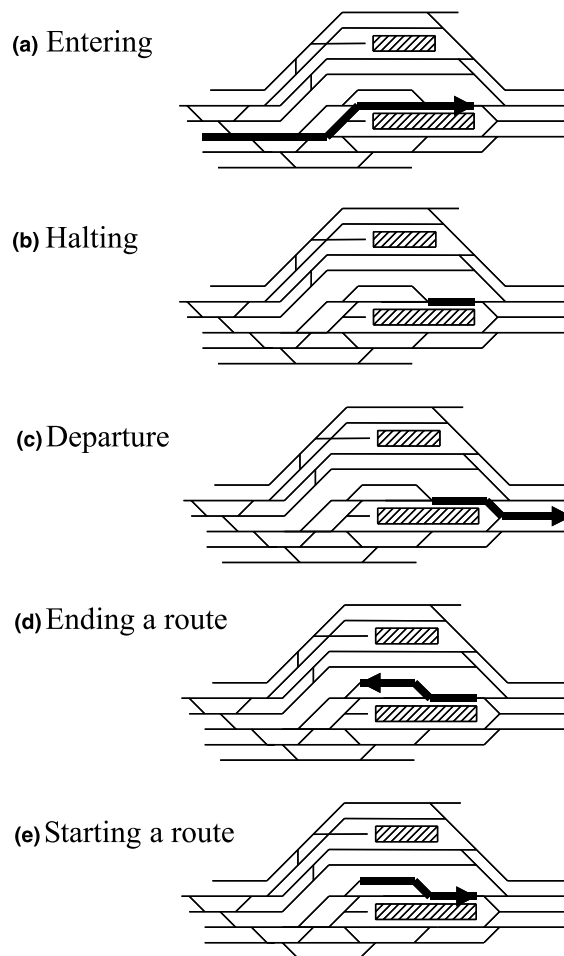
(e) Starting a route

Fig. 3. Train movements at a station.

halting track, where passengers can board (Fig. 3(e)). Again, if the route from the storing track to the halting track, or the halting track itself is occupied, a train has to wait until both are free.

On the long term, however, these train movements are generally not known in detail. For example, which train will use which platform track may be unknown. We therefore propose a simplified model: as in [10] we incorporate the halting tracks only, and assume that each train can use each platform track. Storing tracks are modeled as the "outside" of the model, i.e., a train ending its route leaves the model, and a train starting its route enters the model. Furthermore, the waiting of trains is shifted from the last part of the section tracks to an (imaginary) queue immediately before the station. A train that finds upon arrival at the station, either from a section track, or from a storing track (the "outside"), all halting tracks occupied, joins the tail of this queue, and waits until all trains in the queue have entered the station, and a halting track becomes available.

These assumptions permit modeling a station by means of a first-come-first-served multiserver queueing system, in which each server represents a halting track and in which the service time corresponds to the occupation time of a platform track.

For station $s \in \mathscr{S}$, let k_s be the number of halting tracks, and assume that occupation times are exponentially distributed with mean μ_s^{-1}, equal for all train classes. The contents $\mathbf{c}_s$ of station $s \in \mathscr{S}$ is denoted by the sequence of trains present in s, in order of their arrival, i.e.

$$\mathbf{c}_s = (c_s(1), \ldots, c_s(n_s)),$$

with n_s denoting the number of trains at the station, and $c_s(p)$ the class of the train at position p ($p = 1, \ldots, n_s$). We assume that trains of class c with $r_c(1) = s$ start their route according to a Poisson process with rate λ_c, and define $\lambda_s = \sum_{\{c \in \mathscr{C} | s \in \mathbf{r}_c\}} \lambda_c$, the total number of trains running through station s per time unit.

2.2. Junctions

The lay-out of a junction is given in Fig. 4. It consists of two switches (s_1 and s_2) and a crossing (c). Crossing c can be an equal level crossing (Fig. 4(a)), or a free level crossing ("fly-over") (Fig. 4(b)). Over the junction, four routes are defined, labeled 1, 2, 3 and 4. A train can enter the junction only if its entire route
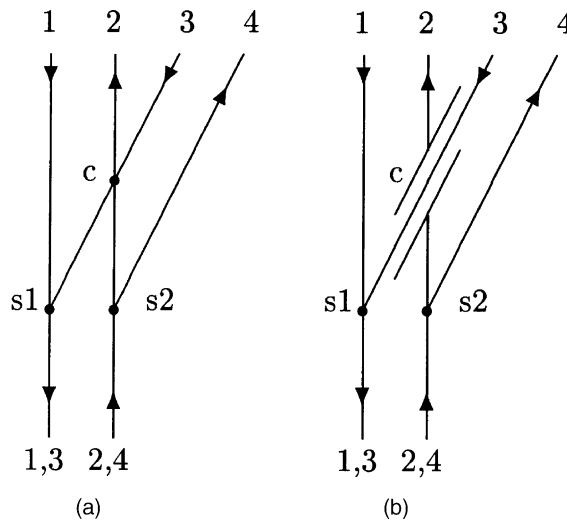

Fig. 4. The lay-out of a junction.

over the junction is free. If this is the case it occupies its route during the time needed to cross the junction; if (part of) its route is occupied by an other train, it has to wait on the section track leading to the junction until all elements of its route are released.

For modeling purposes, waiting of trains will be shifted again from the section tracks to imaginary queues immediately before the junction, and occupation times are assumed to be exponentially distributed, with mean μ_j^{-1} for junction $j \in \mathscr{J}$.

In case of a free level crossing, the junction can be modeled by two first-come-first-served single server queueing systems, one for each switch. The contents of the system representing switch s_i ($i = 1, 2$) of junction $j \in \mathscr{J}$ will be denoted by

$$\mathbf{c}_j^{(i)} = (c_j^{(i)}(1), \ldots, c_j^{(i)}(n_j^{(i)})),$$

where $n_j^{(i)}$ is the number of trains present at switch s_i, and $c_j^{(i)}(p)$ denotes the class of the train at position p in the system ($p = 1, \ldots, n_j^{(i)}$).

In case of an equal level crossing, we model junction $j \in \mathscr{J}$ by a single queueing system, whose contents is described by

$$\mathbf{c}_j = (c_j(1), \ldots, c_j(n_j)),$$

with n_j denoting the number of trains at the junction, and $c_j(p)$ the class of the train the pth position ($p = 1, \ldots, n_j$). The service rate provided at position p when the contents of the junction is $\mathbf{c}_j$ will be denoted by $\mu_j^{(p)}(\mathbf{c}_j)$, and the total service rate by $\mu_j(\mathbf{c}_j) = \sum_{p=1}^{n_j} \mu_j^{(p)}(\mathbf{c}_j)$. Defining $\mu_j^{(p)}(\mathbf{c}_j)$ according to the rules described above unfortunately does not lead to an analytically tractable queueing model. We therefore slightly modify the behaviour of the junction. A class c train at position p in the queue is allowed to enter the junction, if either its route is free (as before), or if there are two trains at positions $1, \ldots, p - 1$, whose routes differ from each other and from the route of class c, and only one of these two trains is crossing the junction. The last, rather artificial, condition permits modeling a junction by the so-called order-independent queue (see [2]), as is shown in Lemma B.1 in Appendix B. It is easily seen that this condition applies very rarely: three trains have to be simultaneously at the junction, and even then it mostly coincides with the first one. As an example, consider the case where trains following route 1, 2 and 3 are simultaneously present at the junction If they arrived in order (1,2,3) or (2,1,3), the first two trains will be crossing the junction, and the artificial condition does not apply. If the order is (1,3,2) or (2,3,1) the condition does apply. The result is, however, that the last train will cross the junction, which corresponds to the actual behaviour. Only for the orders (3,1,2) and (3,2,1), the last train will enter the junction, when it is not allowed. It can therefore be expected that this condition will not seriously effect the behaviour of the junction.

For convenience, we define $\lambda_j^{(r)}$ for junction $j \in \mathscr{J}$ and $r = 1, \ldots, 4$ as the sum of all λ_c, for which $j \in \mathbf{r}_c$ and class c trains follow route r over junction j.

2.3. Sections

A section consists of a number of parallel tracks that are not connected to each other. We assume that section tracks are used in one direction only. This excludes single track sections, that are extremely rare in the Dutch railway network.

In contrast to halting tracks at a station, section tracks can be used by more trains simultaneously. There are, however, two important capacity restrictions. First, overtaking is physically impossible, which may play an important role when trains running at different speed share the same track: a fast train running immediate after a slow train, has to slow down until it can pass at a station. Second, the signaling system divides each track into a number of blocks, that can be used by only one train at a time. This prohibits consecutive trains to be too near: they are at least separated by the running time on a block, called *minimal headway time*.

Similar to [8,10,11], headway times are considered at the beginning and end of a track only. This is done by modeling the first and last block by two first-come-first-served single server queues, one at the beginning and one at the end of the track, with service times equal to the minimal headway time at the beginning resp. end. These headway times are assumed to be exponentially distributed.

The remaining part of the track is modeled by a tandem of $M|M|1$-queues. The total service time of these queues represents the *free running time* on this part of the track, which is defined as the time it takes a train to run through this part, when it is not delayed by other trains. Choosing the number of queues equal to the number of blocks seems natural, but allows fitting the mean free running time only, and thus neglects the important limiting factor of speed differences. We therefore propose an alternative choice, which neglects, like [8,10,11], the number of blocks, but allows to fit both mean and variance of the free running times.

To explain our model, we first introduce some notation. Let k_t be the total number of queues modeling track t, including the two queues for the minimal headway, and assume that these queues are labeled $1, \ldots, k_t$, with queue 1 and k_t the first, resp. last queue. For $i = 1, \ldots, k_t$, define $\mu_t^{(i)}$ as the service rate of queue i of track t. The mean of the total service time of queues $2, \ldots, k_t - 1$ is given by: $\sum_{i=2}^{k_t-1} (\mu_t^{(i)})^{-1}$, and its variance by $\sum_{i=2}^{k_t-1} (\mu_t^{(i)})^{-2}$. Thus, we can fit mean and variance of the free running time on this part of the track by choosing k_t and $\mu_t^{(i)}$ such that

$$\sum_{i=2}^{k_t-1} (\mu_t^{(i)})^{-1} = m_t, \tag{1}$$

with m_t the mean free running time, and

$$\sum_{i=2}^{k_t-1} (\mu_t^{(i)})^{-2} = v_t, \tag{2}$$

with v_t the variance of the free running times on track t. It is easily seen that this system has a solution, provided that $v_t \leqslant m_t^2$. This condition in practice is always fulfilled.

Short tracks, containing one or two blocks only, can only be modeled with one or two queues, and thus only the mean free running time can be fitted. On such short tracks, however, speed differences will never play a great role.

For later convenience, we define $\lambda_t = \sum_{\{c \in \mathscr{C} | t \in \mathbf{r}_c\}} \lambda_c$, the total number of trains running through track t per time unit, and $\mathbf{c}_t^{(i)}$ as the contents of queue i of track t:

$$\mathbf{c}_t^{(i)} = (c_t^{(i)}(1), \ldots, c_t^{(i)}(n_t^{(i)})),$$

with $n_t^{(i)}$ denoting the number of trains at queue i of track t, and $c_j^{(i)}(p)$ the class of the train at position p of this queue ($p = 1, \ldots, n_j$). Moreover, let $\mathbf{c}_t = (\mathbf{c}_t^{(1)}, \ldots, \mathbf{c}_t^{(k_t)})$ denote the contents of all queues of track t.

3. Analysis

We will analyse the network using a Markovian description. Since all queueing systems used in our model have exponential service times, and arrivals at the network occur according to a Poisson process, a sufficient state-description $\mathbf{C}$ of the network is given by the contents of all queues in the model:

$$\mathbf{C} = \left((\mathbf{c}_s)_{s \in \mathscr{S}}, (\mathbf{c}_j)_{j \in \mathscr{J}}, (\mathbf{c}_t)_{t \in \mathscr{T}} \right).$$

Hence, the network can be modeled by a continuous-time Markov chain $(\mathbf{C}(\tau), \tau > 0)$, where $\mathbf{C}(\tau)$ denotes the state of the network at time τ. The steady-state distribution $\pi(\mathbf{C}) \equiv \lim_{\tau \to \infty} \Pr\{\mathbf{C}(\tau) = \mathbf{C}\}$ of this Markov chain is found observing that all queues in the model are quasi-reversible (see [2,9]). Since the

network is open, $\pi(\cdot)$ is given by the product of the steady-state distributions of the queues in isolation, as if they were fed by independent Poisson streams [9, p. 70]. Using the results of [1,2] and Appendix B for the distributions of the states of the queues in isolation, the following theorem is obtained.

Theorem 3.1. *The steady-state distribution of the network is given by*

$$\pi(\mathbf{C}) = \prod_{s \in \mathscr{S}} \pi_s(\mathbf{c}_s) \prod_{j \in \mathscr{J}} \pi_j(\mathbf{c}_j) \prod_{t \in \mathscr{T}} \pi_t(\mathbf{c}_t), \tag{3}$$

where

$$\pi_s(\mathbf{c}_s) = \pi_s(0) \prod_{p=1}^{n_s} \frac{\lambda_{c_s(p)}}{\mu_s \min(p, k_s)}, \tag{4}$$

$$\pi_j(\mathbf{c}_j) = \begin{cases} \pi_j(0) \prod_{p=1}^{n_j} \frac{\lambda_{c_j(p)}}{\mu_j(c_j(1)\ldots c_j(p))} & (\text{equal level crossing}), \\ \left(\pi_j^{(1)}(0) \prod_{p=1}^{n_j^{(1)}} \frac{\lambda_{c_j^{(1)}(p)}}{\mu_j} \right) \left(\pi_j^{(2)}(0) \prod_{p=1}^{n_j^{(2)}} \frac{\lambda_{c_j^{(2)}(p)}}{\mu_j} \right) & (\text{free level crossing}), \end{cases} \tag{5}$$

$$\pi_t(\mathbf{c}_t) = \prod_{i=1}^{k_t} \left(\pi_t^{(i)}(0) \prod_{p=1}^{n_t^{(i)}} \frac{\lambda_{c_t^{(i)}(p)}}{\mu_t^{(i)}} \right), \tag{6}$$

with

$$\pi_s(0) = \left(\sum_{i=0}^{k_s-1} \frac{\rho_s^i}{i!} + \frac{\rho_s^{k_s}}{(k_s - 1)!(k_s - \rho_s)} \right)^{-1}, \tag{7}$$

$$\pi_j(0) = \left\{ \frac{1}{2 - \sum_{r=1}^{4} \rho_j^{(r)}} \left(\frac{1}{1 - \rho_j^{(1)} - \rho_j^{(3)}} + \frac{1}{1 - \rho_j^{(2)} - \rho_j^{(3)}} + \frac{1}{1 - \rho_j^{(2)} - \rho_j^{(4)}} - \frac{1}{1 - \rho_j^{(2)}} - \frac{1}{1 - \rho_j^{(3)}} + 1 \right) \right\}^{-1}, \tag{8}$$

$$\pi_j^{(1)}(0) = 1 - \rho_j^{(1)} - \rho_j^{(3)}, \tag{9}$$

$$\pi_j^{(2)}(0) = 1 - \rho_j^{(2)} - \rho_j^{(4)}, \tag{10}$$

$$\pi_t^{(i)}(0) = 1 - \rho_t^{(i)} \tag{11}$$

and $\rho_s = \lambda_s/\mu_s$, $\rho_j^{(r)} = \lambda_j^{(r)}/\mu_j$, $\rho_t^{(i)} = \lambda_t/\mu_t^{(i)}$.

Form (3) of the steady-state distribution implies that – in steady state – the random variables describing the contents of the queues in the components of the network are stochastically independent. Moreover, from formulas (4)–(6), it is seen that the distribution π_i of the state of component i can be found using parameters of that component only: the physical lay-out of component i, occupation times for component i, and the intensities of all train classes passing component i suffice to compute π_i. This observation has important practical consequences: it is both justified and practically possible to consider only those components of the network that one is interested in. We emphasize that the independence is purely stochastic: physically, the components are strongly dependent, due to trains leaving a component, and simultaneously entering another component.

For practical computation, however, the form of the steady-state distributions for the components (4)–(6) is too detailed. In the following sections, therefore, an aggregation of the state-space is presented. This aggregation is motivated by the so-called arrival theorem [9, p. 70], which states that in an open network of quasi-reversible queues, the distribution of the state of a queue as seen by an arriving customer equals the steady-state distribution of that queue at an arbitrary instant. This aggregation results in explicit expressions for waiting times.

3.1. Stations

By the arrival theorem and the assumption that a train entering station s can use each platform track, a train's waiting time in station s is independent of its class, and depends on the distribution of the total number of trains in this station only. Consequently, the waiting time W_s is given by the waiting time in an ordinary $M|M|k_s|\infty$-queue with arrival rate λ_s. Using the well-known results for this type of queue (e.g. [5]), the expected waiting time EW_s equals

$$\mathrm{EW}_s = \pi_s(0) \frac{\rho_s^{k_s}}{\mu_s(k_s - \rho_s)^2 (k_s - 1)!}, \tag{12}$$

and the probability that a train has to wait more than x time units by

$$\Pr\{W_s > x\} = \pi_s(0) \frac{\rho_s^{k_s}}{(k_s - 1)!(k_s - \rho_s)} e^{-\mu_s(k_s - \rho_s)x}. \tag{13}$$

3.2. Junctions

The arrival theorem implies that the waiting time at a junction depends on the route of the train over the junction only. Therefore, all train classes following the same route over a junction can be aggregated to compute waiting times.

For junctions with equal level crossing, Lemma B.2 in Appendix B derives the joint probability function of the number of trains of each type, from which moments of the number of trains can be computed. By differentiation and the use of Little's formula, the following formulas are obtained for the mean waiting time $\mathrm{EW}_j^{(r)}$ of a train following route r over junction $j \in \mathscr{J}$ ($r = 1, \ldots, 4$).

$$\mathrm{EW}_j^{(1)} = \frac{1}{2\mu_j - \sum \lambda_j^{(r)}} \left(1 + \frac{\pi_j(0)}{(1 - \rho_j^{(1)} - \rho_j^{(3)})^2} \right) - \frac{1}{\mu_j}, \tag{14}$$

$$\mathrm{EW}_j^{(2)} = \frac{1}{2\mu_j - \sum \lambda_j^{(r)}} \left(1 + \frac{\pi_j(0)}{(1 - \rho_j^{(2)} - \rho_j^{(3)})^2} + \frac{\pi_j(0)}{(1 - \rho_j^{(2)} - \rho_j^{(4)})^2} - \frac{\pi_j(0)}{(1 - \rho_j^{(2)})^2} \right) - \frac{1}{\mu_j}, \tag{15}$$

$$\mathrm{EW}_j^{(3)} = \frac{1}{2\mu_j - \sum \lambda_j^{(r)}} \left(1 + \frac{\pi_j(0)}{(1 - \rho_j^{(1)} - \rho_j^{(3)})^2} + \frac{\pi_j(0)}{(1 - \rho_j^{(2)} - \rho_j^{(3)})^2} - \frac{\pi_j(0)}{(1 - \rho_j^{(3)})^2} \right) - \frac{1}{\mu_j}, \tag{16}$$

$$\mathrm{EW}_j^{(4)} = \frac{1}{2\mu_j - \sum \lambda_j^{(r)}} \left(1 + \frac{\pi_j(0)}{(1 - \rho_j^{(2)} - \rho_j^{(4)})^2} \right) - \frac{1}{\mu_j}. \tag{17}$$

Although these formulas look quite complicated, they do reveal the relationship between waiting times and traffic intensities. For instance, the mean waiting of trains on route 2 will increase when the traffic intensities

of trains on route 3 or 4 increase; an observation following intuition, since Fig. 4 (a) indicates that these routes are conflicting with route 2.

For junctions with free level crossing, another aggregation step is suitable: the arrival theorem implies that waiting times for route 1 and 3 are equally distributed, and that waiting times for route 2 and 4 are equally distributed. Hence, the waiting time for route 1 and 3 (2 and 4) equals the waiting time in an $M|M|1$-queue with arrival rate $\lambda_j^{(1)} + \lambda_j^{(3)}$ $(\lambda_j^{(2)} + \lambda_j^{(4)})$, resulting in the following expressions for mean waiting times:

$$EW_j^{(1)} = EW_j^{(3)} = \frac{\rho_j^{(1)} + \rho_j^{(3)}}{\mu_j - \lambda_j^{(1)} - \lambda_j^{(3)}}, \tag{18}$$

$$EW_j^{(2)} = EW_j^{(4)} = \frac{\rho_j^{(2)} + \rho_j^{(4)}}{\mu_j - \lambda_j^{(2)} - \lambda_j^{(4)}}. \tag{19}$$

3.3. Sections

The waiting time that a train experiences in each queue of a section track is, by the arrival theorem, independent of its class, and determined by the distribution of the total number of trains in the queue only. Aggregating the train classes in each queue results in an ordinary $M|M|1$-queue, and thus the mean waiting time EW_t on track t is given by

$$EW_t = \sum_{i=1}^{k_t} \frac{\rho_t^{(i)}}{\mu_t^{(i)} - \lambda_t}. \tag{20}$$

Moreover, a track is overtake-free, and thus the consecutive times that a train spends in queue $1, \ldots, k_t + 2$ of track t are stochastically independent (see [4]). Thus, the Laplace–Stieltjes transform $E(e^{-xR_t})$ of the actual running time R_t on track t is given by

$$\tau_t(x) = \prod_{i=1}^{k_t} \frac{\mu_t^{(i)} - \lambda_t}{\mu_t^{(i)} - \lambda_t + x}.$$

By differentiation of this formula with respect to x, and substitution of $x = 0$, all moments of the actual running time distribution can be obtained.

3.4. Network

The mean waiting time formulas (12), (14)–(20) indicate that the waiting time in component i depends on the routing of trains through λ_i only. Thus, if one is interested in waiting times per component only – for example to identify bottlenecks in the network – the model can be applied using forecasts of the number of trains per time unit only. As forecasts of train numbers will be more reliable than forecasts of train routes and numbers of trains on each route, this property will yield more robust results.

This independence for the routing of trains clearly holds for waiting times per component only. Computation of the total delay a train has accumulated upon entrance into a component, requires the route of that train to be taken into account. For $c \in \mathscr{C}$ and $n = 1, \ldots, s_c$, let $EW_c(n)$ denote the expected total waiting time of a train upon entering component $i_c(n)$, including the waiting time at $i_c(n)$. Since the expectation of a sum of random variables equals the sum of the expectations, $EW_c(n)$ is given by:

$$EW_c(n) = \sum_{s \in \mathscr{S} \cap \{i_c(1),\ldots,i_c(n)\}} EW_s + \sum_{j \in \mathscr{J} \cap \{i_c(1),\ldots,i_c(n)\}} EW_j^{r_{j,c}} + \sum_{t \in \mathscr{T} \cap \{i_c(1),\ldots,i_c(n)\}} EW_t, \tag{21}$$

where $r_{j,c} \in \{1, 2, 3, 4\}$ is the route that class c trains follow over junction j. Adding the corresponding occupation times for stations and junctions, and free running times on the section tracks to this waiting times, provides the sojourn time for each train class upon entrance at the nth stage of the route. This sojourn time corresponds to the actual running time of a train in the network from the start of its route until the nth stage.

For parts of a route that are overtake-free, with possible exception of the first and last component, also the *distribution* of the sojourn time can be characterised via its Laplace–Stieltjes transform [4]. This allows calculation of all moments of the running time on the corresponding part of the route. For instance, the Laplace–Stieltjes transform $\sigma(x) = E(e^{-xR(s_1, t, s_2)})$ of the actual running time $R(s_1, t, s_2)$ on a sequence consisting of station $s_1 \in \mathcal{S}$, section track $t \in \mathcal{T}$, and station $s_2 \in \mathcal{S}$ is given by

$$\sigma(x) = \prod_{i=1}^{2} \frac{\mu_{s_i}}{\mu_{s_i} + x} \left(1 - \pi_{s_i}(0) \frac{x \rho_{s_i}^{k_{s_i}}}{(k_{s_i}\mu_{s_i} - \lambda_{s_i} + x)(k_{s_i} - \rho_{s_i})(k_{s_i} - 1)!} \right) \prod_{i=1}^{k_t} \frac{\mu_t^{(i)} - \lambda_t}{\mu_t^{(i)} - \lambda_t + x}.$$

Note that, by the arrival theorem, this expressions holds for all train classes passing s_1, t and s_2 in this order.

4. Application and validation

In this section, we apply the model to two parts of the Dutch railway network. In the first application, an existing part of the network is analysed, and the forecasts of the model are compared with practical data. The second application is concerned with a part of a future railway line, and shows how our model can be used to investigate bottlenecks in new railway designs.

4.1. Application 1

The first application considers the network depicted in Fig. 5. This part of the Dutch railway network is heavily used: there are two major lines, Rotterdam (Rtd)–Utrecht (Ut) and Den Haag (Gvc)–Utrecht, sharing the same infrastructure between Gouda (Gd) and Utrecht, and two minor lines, Alphen aan de Rijn (Apn)–Utrecht and Rotterdam–Amsterdam, entering and leaving at station Woerden (Wd), and junction Harmelen (Hmla), respectively. Moreover, especially during the rush hours, there is frequent regional traffic, stopping also at the smaller stations, like Woerden and Gouda Goverwelle (Gdg).

Between 1997 and 1998, the Dutch Railways have made a significant change in the train classes (Table 1): more interregional trains on the lines Rotterdam–Utrecht and Den Haag–Utrecht have been added, the number of regional trains is reduced, and the routes of the regional trains have been changed. To investigate
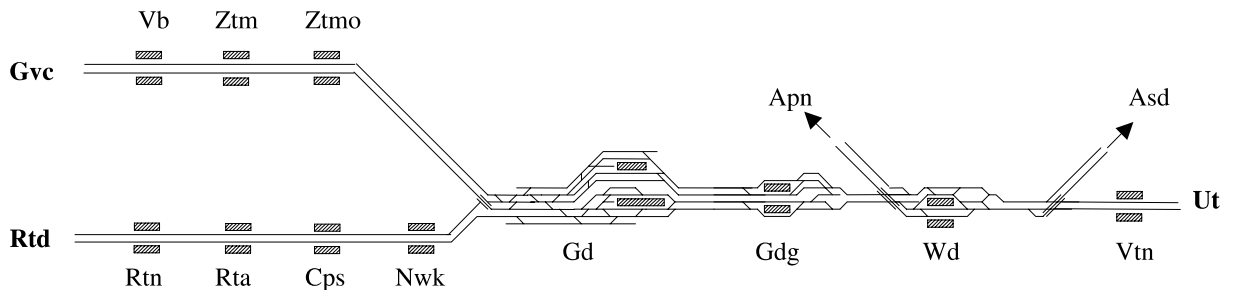

Fig. 5. The network between the stations Rotterdam, Den Haag and Utrecht.

Table 1
Train classes and daily train numbers

Class		# trains per day	
Service	Route	1998	1997
IC (Intercity)	Rtd–Ut	36	36
IC	Gvc–Ut	36	36
IR (Interregional)	Rtd–Ut	10	5
IR	Rtd–Gd	26	–
IR	Gvc–Ut	36	5
R (Regional)	Rtd–Asd	15	18
R	Rtd–Wd	21	–
R	Rtd–Gdg	10	44
R	Rtd–Gd	–	10
R	Gvc–Ut	–	36
R	Gvc–Gdg	36	–
R	Apn–Ut	36	36
F (Freight)	Rtd–Ut	10	10
F	Rtd–Asd	6	6

whether our model captures these changes according to reality, we will apply our model to the situation of 1997, as well as that of 1998.

Application of our model first requires a suitable partitioning in this network. This partitioning is given in Fig. 6. We have incorporated the small stations where no overtaking is possible into the corresponding section tracks (Rtd–Gd, Gvc–Gd, Hmla–Ut). The stations Gd, Gdg and Wd are all modeled by two stations, one for each direction with half of the platform tracks. This corresponds to the actual use of the platform tracks in these stations.

The necessary parameters for the components are provided in Table 2. The occupation times and headway times have been obtained by averaging the corresponding times per train class, which explains the difference between 1998 and 1997. The short section tracks between Gd and Gdg, and between Wd and Hmla have been modeled by one, resp. two $M|M|1$-queues.

From Table 2, the mean waiting times upon entering the stations Rtd, Gvc, Gd and Ut were computed for each train class, using formula (21). The results are given in Tables 3 and 4, together with the
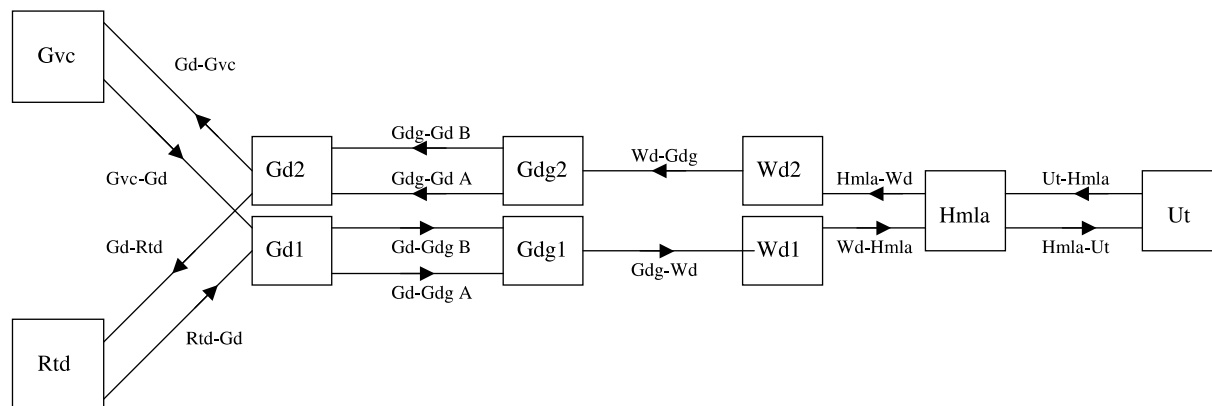

Fig. 6. Partitioning of the network into components.

Table 2
Input

Station	# platform tracks	Occupation time			# trains per day	
		1998	1997		1998	1997
Rtd	4	5.5	5.3		268	258
Gvc	3	5.7	5.5		216	154
Gd1, Gd2	3	4.4	4.3		242	206
Gdg1, Gdg2	2	3.5	3.6		216	196
Wd1, Wd2	2	4.0	4.2		206	188
Ut	7	5.7	5.5		490	490

Junction	Free level crossing	Occupation time			# trains per day	
		1998	1997		1998	1997
Hmla	Yes	1.7	1.8		185	188

Section	Min. headway		m_t		v_t		# trains per day	
	1998	1997	1998	1997	1998	1997	1998	1997
Rtd–Gd, Gd–Rtd	1.7	1.8	15.2	16.2	7.2	7.4	134	129
Gvc–Gd, Gd–Gvc	1.7	1.7	16.3	16.7	5.4	5.5	108	77
Gd–Gdg A	1.7	1.6	–	–	–		180	116
Gd–Gdg B	2.0	2.0	–	–	–		36	80
Gdg–Gd A	1.6	1.7	–	–	–		165	98
Gdg–Gd B	2.0	1.9	–	–	–		51	98
Gdg–Wd, Wd–Gdg	1.7	1.7	3.7	4.0	0.7	0.9	170	152
Wd–Hmla, Hmla–Wd	1.6	1.6	–	–	–	–	185	188
Hmla–Ut, Ut–Hmla	1.6	1.8	4.6	5.0	0.8	1.0	164	164

Table 3
Mean waiting times in 1998

Class	Rtd		Gvc		Gd		Ut	
	Model	Pract	Model	Pract	Model	Pract	Model	Pract
IC Rtd–Ut	7.6	6.0	–	–	3.8	3.4	7.6	5.5
IC Gvc–Ut	–	–	7.2	6.4	3.5	4.0	6.9	7.3
IR Rtd–Ut	7.6	6.9	–	–	3.8	3.4	7.6	6.2
IR Gvc–Ut	–	–	7.2	5.8	3.5	4.2	6.9	6.1
R Rtd–Asd	6.1	6.5	–	–	3.1	3.4	–	–
R Rtd–Wd	4.2	4.8	2.1	2.7	–	–		
R Rtd–Gdg	2.6	3.7	–	–	1.3	2.7	–	–
R Gvc–Gdg	–	–	1.8	2.8	0.8	1.8	–	–
R Apn–Ut	–	–	–	–	–	–	3.4	1.8
Average	6.3	5.9	5.4	5.0	2.9	3.2	6.3	5.2

corresponding waiting times in practice. The latter are the sum of the waiting times added in the timetable (to make it conflict free, and to handle small delays), and the delays that were observed in practice. The last data were only available in aggregated form (eastern and western direction together); therefore the waiting times by our model are aggregated in the same way.

The results of our model appear to be rather close to the practical observations (approximately 20%), and appear to capture the differences between 1998 and 1997: an increase in waiting times in Rotterdam,

Table 4
Mean waiting times in 1997

Class	Rtd		Gvc		Gd		Ut	
	Model	Pract	Model	Pract	Model	Pract	Model	Pract
IC Rtd–Ut	7.5	6.0	–	–	3.8	3.7	7.5	6.8
IC Gvc–Ut	–	–	6.7	6.6	3.4	3.2	6.6	5.3
IR Rtd–Ut	7.5	7.1	–	–	3.8	3.0	7.5	6.0
IR Gvc–Ut	–	–	6.7	6.1	3.3	2.9	6.6	5.8
R Rtd–Asd	6.2	5.6	–	–	3.2	3.0	–	–
R Rtd–Gdg	2.4	4.5	–	–	1.2	1.7	–	–
R Rtd–Gd	1.9	3.9	–	–	1.0	2.3	–	–
R Gvc–Ut	–	–	6.8	5.0	3.4	1.9	6.6	3.6
R Apn–Ut	–	–	–	–	–	–	3.7	3.4
Average	4.8	5.2	6.7	5.8	2.2	2.5	6.2	4.9

Gouda and Utrecht, and a decrease in Den Haag. Note that these results are in fact rather surprising, since our model is not based on a timetable, but on traffic intensities only. The model thus provides aggregated performance measures over all possible departure times of trains, while the data from practice are obviously the result of the specific departure times in the timetable.

4.2. Application 2

The second application is concerned with a part of the Betuwe route, a new freight line that will connect the port of Rotterdam to Germany. We consider the part from Maasvlakte (Mvt) to Pernis (Ps), (see Fig. 7), which is used by six train classes (Table 5).

The required parameters for the components are given in Table 6. Note that the variance of the free running time distributions is very small, as all trains are freight trains of approximately the same weight and length.

Substitution of these parameters into formulas (14)–(20) and application of (21) provides the expected waiting times per train class, presented in Figs. 8 and 9. The contribution of the section tracks to the total delays is relatively small. This agrees with intuition: the headway times are small, and trains run almost with the same speed. Only the track Ps–Bot plays a significant role, due to the large headway time at Ps. The main bottleneck, however, appears to be the fly-over at Bot in Eastern direction, due to its large occupation time.

When all traffic intensities are increased, the congestive influence of the Botlek bottleneck becomes more and more apparent. This is seen in Fig. 10, where traffic intensities of all train classes are varied by multiplying them simultaneously with a factor ranging from 0 to 2.
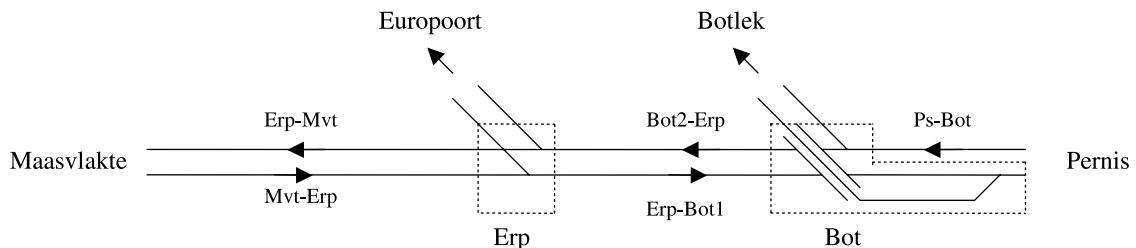

Fig. 7. The Betuwe route between Maasvlakte and Pernis.

Table 5
Train classes of the Betuwe route

Class	Number per hour	Route
1	4	Mvt–Erp, Erp, Erp–Bot, Bot
2	1	Erp, Erp–Bot, Bot
3	1	Bot
4	4	Ps–Bot, Bot, Bot–Erp, Erp, Erp–Mvt
5	1	Ps–Bot, Bot, Bot–Erp, Erp
6	1	Ps–Bot, Bot

Table 6
Input for the Betuwe route

Junction	Fly-over	Mean occupation time (min)	
Erp	No	2.53	
Bot	Yes	2.88	(Western direction)
		3.63	(Eastern direction)
Section track	Minimal headway (min)	m_t (min)	v_t
Mvt–Erp	1.0	7.44	0.043
Erp–Bot	1.0	4.23	0.019
Ps–Bot	2.5 (Ps)	1.49	0.006
	1.0 (Bot)		
Bot–Erp	1.0	5.08	0.024
Erp–Mvt	1.0	6.87	0.038
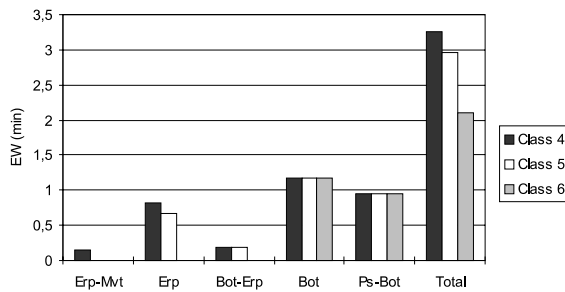
Fig. 8. Waiting times per component in Western direction.
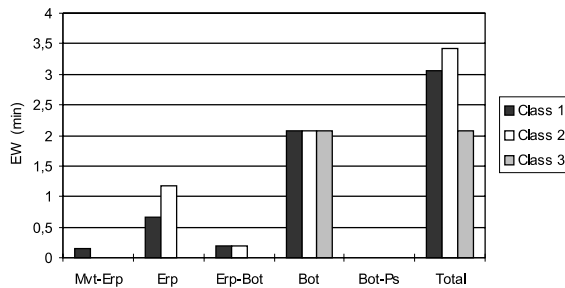

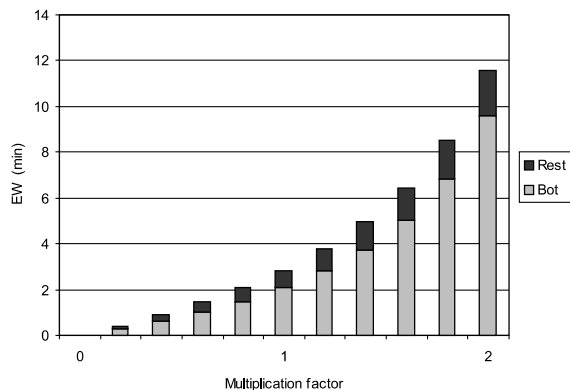Fig. 9. Waiting times per component in Eastern direction.

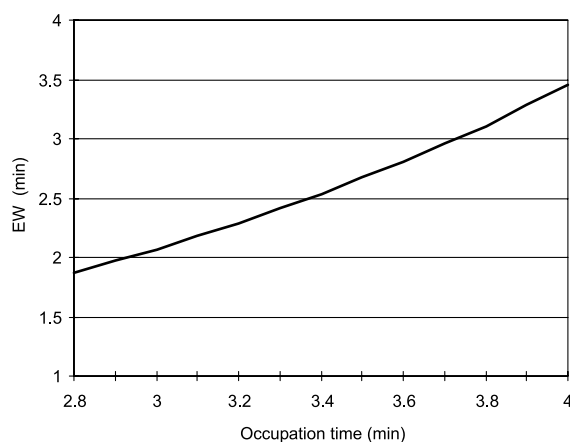Fig. 10. Waiting times in Eastern direction when traffic intensities vary.

Fig. 11. The waiting time in Eastern direction as function of the occupation time at Bot.

Thus, it would be desirable to reduce the occupation time of *Bot* in Eastern direction, for example by a different positioning of the signals around the junction. In Fig. 11 the influence of such a reduction on the waiting time is investigated. The occupation time is varied from 2.8 to 4 minutes, and the original traffic intensities from Table 5 are used. Reducing the occupation time to 2.9 minutes, equal to the occupation time in Western direction, will reduce the total waiting time from Maasvlakte to Pernis with about one-third.

5. Discussion

A product form queueing network model for railway networks is provided. Application to the Dutch railway network shows that the model is accurate enough for long term global performance evaluation of railway networks. However, a number of aspects need to be considered, as will be discussed below.

- Stations are modeled in a rather simple way, with the assumption that every train can use every platform track. In reality, in contrast, each train class can use only one or two of the platform tracks. Accordingly, one should model stations by means of an MSCCC-queue [3], instead of an ordinary multiserver queue.

In this type of queue, extra conditions can be imposed on the number of trains of each class that can be in service. Most likely, this modeling will involve some approximations, like the junction model in Section 2.2. Further research to this topic is suggested.
- While the simple station model underestimates waiting times, the exponentiality assumption for occupation times and minimal headway times overestimates waiting times. Exponentiality assumptions, are crucial in product form networks with first-come-first-served discipline, and cannot be relaxed to general distributions, without losing the analytic solution for the stationary distribution of the network. Approximation methods to deal with this problem are thus of interest.
- Another crucial assumption is the incorporation of artificial intermediate queues between the components of the network with unlimited capacities. This generally underestimates waiting times. However, as railway networks are lightly loaded (even in the heavily used part of the Dutch railway network in Application 1, loads are generally below 30%), this assumption results in small errors.
- The dependence between free running times, and the route of a train through the network has been neglected. As can be seen from the numerical validation of the model, this may affect waiting times per train class, but the aggregated waiting times over all train classes remain sufficiently accurate.
- Priority rules are not incorporated. Only first-come-first-served strategies are considered. The validation of the model yields the same conclusion as above: this assumption influences waiting times per train class, but not on an aggregate level.
- Finally, our results are based on the assumption that the network is in steady state. If the network contains no trains at time zero, this assumption overestimate waiting times. Product form networks, however, are known to converge fairly rapidly to steady state.

Acknowledgements

Appendix A. Notation

Due to the complexity of a railway network, the notation has become rather complicated. This appendix therefore summarises the notation introduced in Section 2.

Network

$\mathscr{S}$	the set of stations
$\mathscr{J}$	the set of junctions
$\mathscr{T}$	the set of tracks
$\mathscr{I}$	the set of all infrastructural components $\mathscr{S} \cup \mathscr{J} \cup \mathscr{T}$
$\mathscr{C}$	the set of train classes
$\mathbf{r}_c$	the route of class c trains ($c \in \mathscr{C}$), denoted by $(i_c(1), \ldots, i_c(s_c))$ with $i_c(j)$ the component of the network a class c train passes on the jth stage of his route $i_c(j) \in \mathscr{I} (j = 1, \ldots, s_c)$ s_c the number of stages of the route
λ_c	the arrival rate of trains of class $c \in \mathscr{C}$ at their starting station $i_c(1)$

Stations

λ_s	the mean number of trains per time unit at station $s \in \mathscr{S}$
k_s	the number of halting tracks in station s

μ_s^{-1}　　the mean halting time at station s

n_s　　the number of trains present in station s

$\mathbf{c}_s$　　the contents of station s, given by $(c_s(1), \ldots, c_s(n_s))$, with $c_s(p)$ the class of the train in position p ($p = 1, \ldots, n_s$)

Junctions

$\lambda_j^{(r)}$　　the mean number of trains per time unit following route r over junction $j \in \mathscr{J}$ ($r = 1, 2, 3, 4$)

μ_j^{-1}　　the mean occupation time of junction j

Free level crossing

$n_j^{(i)}$　　the number of trains at switch i of junction j ($i = 1, 2$)

$\mathbf{c}_j^{(i)}$　　the contents of switch i of junction j, given by $(c_j^{(i)}(1), \ldots, c_j^{(i)}(n_j))$ with $c_j^{(i)}(p)$ the class of the train in position p ($p = 1, \ldots, n_j^{(i)}$)

Equal level crossing

n_j　　the number of trains at junction j ($i = 1, 2$)

$\mathbf{c}_j$　　the contents of junction j, given by $(c_j(1), \ldots, c_j(n_j))$, with $c_j(p)$ the class of the train in position p ($p = 1, \ldots, n_j$)

$\mu_j(p; \mathbf{c}_j)$　the service rate provided at position p of the queue of junction j when the contents of the queue is $\mathbf{c}_j$, given by

$$\mu_j(p; \mathbf{c}_j) = \begin{cases} \mu_j & \text{if the train at position } p \text{ can enter the junction, or} \\ & \exists q, r < p \text{ such that } q \neq r, \mu_j(q; \mathbf{c}_j) + \mu_j(r; \mathbf{c}_j) = \mu_j \\ & \text{and } c_j(p), c_j(q) \text{ and } c_j(r) \text{ follow different} \\ & \text{routes over the junction} \\ 0 & \text{otherwise} \end{cases}$$

$\mu_j(\mathbf{c}_j)$　$= \sum_{p=1}^{n_j} \mu_j(p; \mathbf{c}_j)$, the total service rate when the contents of the queue is $\mathbf{c}_j$

Section tracks

λ_t　　the mean number of trains per time unit at track $t \in \mathscr{T}$

k_t　　the number of queues modelling track t

$\mu_t^{(i)}$　　the service rate of queue i of track t ($i = 1, \ldots, k_t$)

$n_j^{(i)}$　　the number of trains present in queue i of track t

$\mathbf{c}_t^{(i)}$　　the contents of queue i of track t, given by $(c_j^{(i)}(1), \ldots, c_j^{(i)}(n_j))$, with $c_j^{(i)}(p)$ the class of the train in position p of the queue ($p = 1, \ldots, n_j^{(i)}$)

$\mathbf{c}_t$　　$= (\mathbf{c}_t^{(1)}, \ldots, \mathbf{c}_t^{(k_t)})$

Appendix B. Junctions

In Section 2.2 a junction with equal level crossing is modelled by a multiclass queue with exponential service times. A train of class c at position p is served if either its route is free, or if there are two trains at positions $1, \ldots, p - 1$, whose routes differ from each other and from the route of class c, and only one of these two trains is crossing the junction.

Lemma B.1. *This model satisfies conditions of an order-independent queue.*

Proof. According to [2], a queue with exponential service times is an order independent queue, if the following three conditions hold.

- The service rate is positive if the queue is non-empty.
- The service rate provided at a customer in position p in the queue is not influenced by customers on positions $p + 1, p + 2, \ldots$.
- The total service rate provided at the queue is independent of the order in which customers arrived.

By the form of the service-discipline, it is easily seen that the first two conditions are satisfied. To prove the third condition we distinguish between the number of different routes of all trains in the queue, say d. If $d = 1$, only the first train will be served and the service rate equals μ_j. In case $d = 2$ and the two routes are conflicting, only one train will be served and the total service rate equals μ_j; if the two routes are not conflicting, the first trains of each route will be served, and the service rate will be $2\mu_j$. Finally, for $d = 3$ and $d = 4$ the second (approximate) condition of the service discipline applies, and always two trains will be served. Hence, in all cases, the total service rate is independent of the order in which trains arrived, and also the last condition is satisfied. $\quad\square$

Lemma B.2. *Let $n_j^{(r)}$ be the number of trains at junction $j \in \mathscr{J}$ that follow route r ($r = 1, \ldots, 4$). The joint probability generating function $g_j(s_1, \ldots, s_4)$ of $n_j^{(1)}, \ldots, n_j^{(4)}$ is given by:*

$$g_j(s_1, \ldots, s_4) = \frac{\pi_j(0)}{2 - \sum_{r=1}^{4} s_r \rho_j^{(r)}} \left(\frac{1}{1 - s_1 \rho_j^{(1)} - s_3 \rho_j^{(3)}} + \frac{1}{1 - s_2 \rho_j^{(2)} - s_3 \rho_j^{(3)}} + \frac{1}{1 - s_2 \rho_j^{(2)} - s_4 \rho_j^{(4)}} \right.$$
$$\left. - \frac{1}{1 - s_2 \rho_j^{(2)}} - \frac{1}{1 - s_3 \rho_j^{(3)}} + 1 \right), \tag{B.1}$$

with

$$\pi_j(0) = \left\{ \frac{1}{2 - \sum_{r=1}^{4} \rho_j^{(r)}} \left(\frac{1}{1 - \rho_j^{(1)} - \rho_j^{(3)}} + \frac{1}{1 - \rho_j^{(2)} - \rho_j^{(3)}} + \frac{1}{1 - \rho_j^{(2)} - \rho_j^{(4)}} - \frac{1}{1 - \rho_j^{(2)}} - \frac{1}{1 - \rho_j^{(3)}} + 1 \right) \right\}^{-1}. \tag{B.2}$$

Proof. Let $p_j(\mathbf{n}) = \mathrm{Pr}\{n_j^{(1)} = n_1, \ldots n_j^{(4)} = n_4\}$, and define $p_j(\mathbf{n}) = 0$ if $n_i < 0$ for some i. For non-negative $\mathbf{n}$ with at least one positive n_i, $p_j(\mathbf{n})$ satisfy the following recursive relation (see [2]):

$$p_j(\mathbf{n}) = \frac{1}{\mu_j(\mathbf{n})} \sum_{r=1}^{4} \lambda_j^{(r)} p_j(\mathbf{n} - \mathbf{e}_r), \tag{B.3}$$

where $\mu_j(\mathbf{n})$ is the total service-rate at which customers are served whenever $\mathbf{n}$ customers are present (note that, by the definition of an order-independent queue, this rate is independent of the order of these customers). Using this recursion and the fact that

$$\mu_j(\mathbf{n}) = \begin{cases} \mu_j & \text{if } \mathbf{n} = n_r \mathbf{e}_r, (n_1, 0, n_3, 0), (0, n_2, n_3, 0) \text{ or } (0, n_2, 0, n_4), \quad (n_i > 0, i = 1, \ldots, 4), \\ 2\mu_j & \text{otherwise,} \end{cases}$$

we get

$$g_j(s_1, \ldots, s_4) \stackrel{\mathrm{def}}{=} \sum_{n_1, \ldots, n_4 = 0}^{\infty} s_1^{n_1} \ldots s_4^{n_4} p_j(\mathbf{n})$$
$$= p_j(\mathbf{0}) + \sum_{r=1}^{4} \frac{1}{2} \rho_j^{(r)} \left\{ \sum_{n_1, \ldots, n_4 = 0}^{\infty} s_1^{n_1} \ldots s_4^{n_4} p_j(\mathbf{n} - \mathbf{e}_r) \right.$$
$$\left. + \sum_{n_1 = 1}^{\infty} s_1^{n_1} p_j((n_1, 0, 0, 0) - \mathbf{e}_r) + \sum_{n_2 = 1}^{\infty} s_2^{n_2} p_j((0, n_2, 0, 0) - \mathbf{e}_r) \right.$$

$$+ \sum_{n_3=1}^{\infty} s_3^{n_3} p_j((0,0,n_3,0) - \mathbf{e}_r) + \sum_{n_4=1}^{\infty} s_4^{n_4} p_j((0,0,0,n_4) - \mathbf{e}_r)$$

$$+ \sum_{n_1=1,n_3=1}^{\infty} s_1^{n_1} s_3^{n_3} p_j((n_1,0,n_3,0) - \mathbf{e}_r)$$

$$+ \sum_{n_2=1,n_3=1}^{\infty} s_2^{n_2} s_3^{n_3} p_j((0,n_2,n_3,0) - \mathbf{e}_r)$$

$$+ \sum_{n_2=1,n_4=1}^{\infty} s_2^{n_2} s_4^{n_4} p_j((0,n_2,0,n_4) - \mathbf{e}_r) \Bigg\}.$$

Excluding terms that are equal to zero and rewriting the summations leads to:

$$g_j(s_1,\ldots,s_4) = p_j(\mathbf{0}) + \sum_{r=1}^{4} \frac{1}{2} s_r \rho_j^{(r)} g_j(s_1,\ldots,s_4) + \frac{1}{2} s_1 \rho_j^{(1)} \sum_{n_1=0}^{\infty} s_1^{n_1} p_j(n_1,0,0,0)$$

$$+ \frac{1}{2} s_2 \rho_j^{(2)} \sum_{n_2=0}^{\infty} s_2^{n_2} p_j(0,n_2,0,0) + \frac{1}{2} s_3 \rho_j^{(3)} \sum_{n_3=0}^{\infty} s_3^{n_3} p_j(0,0,n_3,0) + \frac{1}{2} s_4 \rho_j^{(4)} \sum_{n_4=0}^{\infty} s_4^{n_4} p_j(0,0,0,n_4)$$

$$+ \frac{1}{2} s_1 \rho_j^{(1)} \sum_{n_1=0,n_3=1}^{\infty} s_1^{n_1} s_3^{n_3} p_j(n_1,0,n_3,0) + \frac{1}{2} s_3 \rho_j^{(3)} \sum_{n_1=1,n_3=0}^{\infty} s_1^{n_1} s_3^{n_3} p_j(n_1,0,n_3,0)$$

$$+ \frac{1}{2} s_2 \rho_j^{(2)} \sum_{n_2=0,n_3=1}^{\infty} s_2^{n_2} s_3^{n_3} p_j(0,n_2,n_3,0) + \frac{1}{2} s_3 \rho_j^{(3)} \sum_{n_2=1,n_3=0}^{\infty} s_2^{n_2} s_3^{n_3} p_j(0,n_2,n_3,0)$$

$$+ \frac{1}{2} s_2 \rho_j^{(2)} \sum_{n_2=0,n_4=1}^{\infty} s_2^{n_2} s_4^{n_4} p_j(0,n_2,0,n_4) + \frac{1}{2} s_4 \rho_j^{(4)} \sum_{n_2=1,n_4=0}^{\infty} s_2^{n_2} s_4^{n_4} p_j(0,n_2,0,n_4).$$

Noting that $p_j(\mathbf{0}) = \pi_j(0)$ and regrouping summations gives

$$g_j(s_1,\ldots,s_4) = \pi_j(0) + \sum_{i=1}^{4} \frac{1}{2} s_i \rho_j^{(i)} g_j(s_1,\ldots,s_4) + \left(\frac{1}{2} s_1 \rho_j^{(1)} + \frac{1}{2} s_3 \rho_j^{(3)} \right) \sum_{n_1=0,n_3=0}^{\infty} s_1^{n_1} s_3^{n_3} p_j(n_1,0,n_3,0)$$

$$+ \left(\frac{1}{2} s_2 \rho_j^{(2)} + \frac{1}{2} s_4 \rho_j^{(4)} \right) \sum_{n_2=0,n_4=0}^{\infty} s_2^{n_2} s_4^{n_4} p_j(0,n_2,0,n_4)$$

$$+ \left(\frac{1}{2} s_2 \rho_j^{(2)} + \frac{1}{2} s_3 \rho_j^{(3)} \right) \sum_{n_2=0,n_3=0}^{\infty} s_2^{n_2} s_3^{n_3} p_j(0,n_2,n_3,0) - \frac{1}{2} s_2 \rho_j^{(2)} \sum_{n_2=0}^{\infty} s_2^{n_2} p_j(0,n_2,0,0)$$

$$- \frac{1}{2} s_3 \rho_j^{(3)} \sum_{n_3=0}^{\infty} s_3^{n_3} p_j(0,0,n_3,0). \tag{B.4}$$

We first evaluate the infinite summations on the right-hand side. Since, by (B.3),

$$\sum_{n_1=0,n_3=0}^{\infty} s_1^{n_1} s_3^{n_3} p_j(n_1,0,n_3,0) = \pi_j(0) + \rho_j^{(1)} \sum_{n_1=0,n_3=0}^{\infty} s_1^{n_1} s_2^{n_3} p_j(n_1-1,0,n_3,0)$$

$$+ \rho_j^{(3)} \sum_{n_1=0,n_3=0}^{\infty} s_1^{n_1} s_2^{n_3} p_j(n_1,0,n_3-1,0)$$

$$= \pi_j(0) + s_1 \rho_j^{(1)} \sum_{n_1=0,n_3=0}^{\infty} s_1^{n_1} s_3^{n_3} p_j(n_1,0,n_3,0)$$

$$+ s_3 \rho_j^{(3)} \sum_{n_1=0,n_3=0}^{\infty} s_1^{n_1} s_3^{n_3} p_j(n_1,0,n_3,0),$$

it follows that

$$\sum_{n_1=0,n_3=0}^{\infty} s_1^{n_1} s_3^{n_3} p_j(n_1,0,n_3,0) = \frac{\pi_j(0)}{1 - s_1\rho_j^{(1)} - s_3\rho_j^{(3)}}. \tag{B.5}$$

In the same way, one can derive that

$$\sum_{n_2=0,n_4=0}^{\infty} s_2^{n_2} s_4^{n_4} p_j(0,n_2,0,n_4) = \frac{\pi_j(0)}{1 - s_2\rho_j^{(2)} - s_4\rho_j^{(4)}}, \tag{B.6}$$

$$\sum_{n_2=0,n_3=0}^{\infty} s_2^{n_2} s_3^{n_3} p_j(0,n_2,n_3,0) = \frac{\pi_j(0)}{1 - s_2\rho_j^{(2)} - s_3\rho_j^{(3)}} \tag{B.7}$$

and

$$\sum_{n_2=0}^{\infty} s_2^{n_2} p_j(0,n_2,0,0) = \frac{\pi_j(0)}{1 - s_2\rho_j^{(2)}}, \tag{B.8}$$

$$\sum_{n_3=0}^{\infty} s_3^{n_3} p_j(0,0,n_3,0) = \frac{\pi_j(0)}{1 - s_3\rho_j^{(3)}}. \tag{B.9}$$

Substituting (B.5)–(B.9) into (B.4) gives

$$g_j(s_1,\ldots,s_4) = \sum_{r=1}^{4} \frac{1}{2} s_r \rho_j^{(r)} g_j(s_1,\ldots,s_4) + \left(\frac{1}{2} s_1 \rho_j^{(1)} + \frac{1}{2} s_3 \rho_j^{(3)} \right) \frac{\pi_j(0)}{1 - s_1\rho_j^{(1)} - s_3\rho_j^{(3)}}$$
$$+ \left(\frac{1}{2} s_2 \rho_j^{(2)} + \frac{1}{2} s_4 \rho_j^{(4)} \right) \frac{\pi_j(0)}{1 - s_2\rho_j^{(2)} - s_4\rho_j^{(4)}} + \left(\frac{1}{2} s_2 \rho_j^{(2)} + \frac{1}{2} s_3 \rho_j^{(3)} \right) \frac{\pi_j(0)}{1 - s_2\rho_j^{(2)} - s_3\rho_j^{(3)}}$$
$$- \frac{1}{2} s_2 \rho_j^{(2)} \frac{\pi_j(0)}{1 - s_2\rho_j^{(2)}} - \frac{1}{2} s_3 \rho_j^{(3)} \frac{\pi_j(0)}{1 - s_3\rho_j^{(3)}} + \pi_j(0).$$

We simplify this expression using the fact that $a/(1-a)$ equals $(1/(1-a)) - 1$:

$$g_j(s_1,\ldots,s_4) = \sum_{r=1}^{4} \frac{1}{2} s_r \rho_j^{(r)} g_j(s_1,\ldots,s_4) + \frac{1}{2} \pi_j(0) \left\{ \frac{1}{1 - s_1\rho_j^{(1)} - s_3\rho_j^{(3)}} + \frac{1}{1 - s_2\rho_j^{(2)} - s_4\rho_j^{(4)}} \right.$$
$$\left. + \frac{1}{1 - s_2\rho_j^{(2)} - s_3\rho_j^{(3)}} - \frac{1}{1 - s_2\rho_j^{(2)}} - \frac{1}{1 - s_3\rho_j^{(3)}} + 1 \right\}.$$

Solving this equation for $g_j(s_1,\ldots,s_4)$ leads to (B.1) and, since

$$g_j(1,1,1,1) = \sum_{n_1,\ldots,n_4=0}^{\infty} p_j(\mathbf{n}) = 1,$$

(B.2) follows immediately. $\square$

References

[1] F. Baskett, K.M. Chandy, R.R. Muntz, F.G. Palacios, Open closed and mixed networks of queues with different classes of customers, Journal of the ACM 22 (2) (1975) 248–260.

[2] S.A. Berezner, C.F. Kriel, A.E. Krzesinski, Quasi-reversible multiclass queues with order independent departure rates, Queueing Systems 19 (1995) 345–359.

[3] J.Y. Le Boudec, A BCMP extension to multiserver stations with concurrent classes of customers, in: Proceedings of the 1986 ACM Sigmetrics Conference Performance Evaluation Review, vol. 14, 1986, pp. 78–91.

[4] O.J. Boxma, H. Daduna, Sojourn times in queueing networks, in: Stochastic Analysis of Computer and Communication Systems, North-Holland, Amsterdam, 1990, pp. 401–450.

[5] R.B. Cooper, Introduction to Queueing Theory, second ed., North-Holland, Amsterdam, 1981.

[6] C. Gerondeau, Transport in Europe, Artech House Publishers, 1997.

[7] G. Hertel, Leistungsfähigkeit und Leistungsverhalten von Eisenbahnbetriebsanlagen – Modelle und Berechnungsmöglichkeiten, in: 1. Eisenbahnbetriebswissenschaftliches Kolloquium, Technische Universität Dresden, 1995, pp. 62–104 (in German).

[8] T. Huisman, R.J. Boucherie, Running times on railway sections with heterogeneous train traffic, Transportation Research Part B 35 (2001) 271–292.

[9] F.P. Kelly, Reversibility and Stochastic Networks, Wiley, New York, 1979.

[10] M. Noordeen, Stability analysis of cyclic timetables for a highly interconnected rail network, PhD Thesis, Ecole Polytechnique Fédérale de Lausanne, 1996.

[11] A. Schrijver, A. Steenbeek, Dienstregelingontwikkeling voor Railned – Rapport CADANS 1.0. Technical Report, Center for Mathematics and Computer Science, Amsterdam, 1994 (in Dutch).

[12] N.M. van Dijk, Where would we be without trains? A queueing model to determine train delays, in: Operations Research Proceedings 1993, Springer, Berlin, 1993, pp. 523–529.

[13] H. Wakob, Ableitung eines generellen Wartemodells zur Ermittlung der planmäßigen Wartezeit im Eisenbahnbetrieb unter besonderer Berückischtigung der Aspecte Leistungsfähigkeit und Anlagenbelastung. Veröffentlichungen des Verkehrswissenschaftlichen Institutes der RWTH Heft 36, Rheinisch-Westfälische Technische Hochschule Aachen, 1985 (in German, English summary).