Train Movement Analysis at Railway Stations

Procedures & Evaluation of Wakob's Approach

Ir. A.F. de Kort
Dr. B. Heidergott
Ir. R.J. van Egmond
Dr. G. Hooghiemstra
TRAIN MOVEMENT ANALYSIS
AT RAILWAY STATIONS:
Procedures & Evaluation of Wakob's Approach
TRAIN MOVEMENT ANALYSIS
AT RAILWAY STATIONS:

Procedures & Evaluation of Wakob's Approach

TRAIL Research School, Delft

Authors:
ir. Antoine F. de Kort
Transportation Planning and Traffic Engineering Section,
Faculty of Civil Engineering and Geosciences, Delft University of Technology

dr. Bernd Heidergott
ir. Robert J. van Egmond
Applied Analysis Department,
Faculty of Information Technology and Systems, Delft University of Technology

dr. Gerard Hooghiemstra
Statistics, Probability Theory and Operations Research Department,
Faculty of Information Technology and Systems, Delft University of Technology
4 TFK WAITING TIME CALCULATIONS

4.1 Introduction ........................................... 25
4.2 Proper construction of the $E_k(\lambda)/E_l(\mu)/1$ system .... 25
4.3 The TFK occupancy rate ................................. 26
4.4 Waiting time calculations ................................ 27
4.5 Waiting times due to folding and crossing ............... 29
4.6 Procedures for station capacity assessment ............. 29
4.7 Conclusions and remarks regarding the waiting time calculations .... 30

5 CASE STUDY: STATION THE HAGUE HS

5.1 Introduction ............................................. 33
5.2 Priority rule definition .................................. 35
5.3 Definition of unique train series .......................... 35
5.4 Identification of TFK's at The Hague HS ............... 36
5.5 Arrival process specification ............................... 36
5.6 Service process specification ............................... 38
5.7 Numerical results ........................................... 39
5.8 Conclusions ................................................. 41

6 OVERALL CONCLUSIONS AND FURTHER RESEARCH

6.1 Overall conclusions ....................................... 43
6.2 Further research ........................................... 44

BIBLIOGRAPHY

A WAKOB'S WAITING TIME APPROXIMATION

A.1 Why using approximations? ............................... 49
A.2 The "ersatz" systems .................................... 50
  A.2.1 The $g^*$ - system .................................. 50
  A.2.2 The $\tilde{g}$ - system ................................ 51
A.3 An algorithm for Wakob's waiting time approximation .... 52
A.4 The functions $a(k, l)$ and $b(k, l)$ .................. 52
A.5 Concluding remarks on Wakob’s approximation method .... 53
A.6 A comparison of several approximation methods .......... 54
This publication is the result of the research program "Seamless Multimodal Mobility", carried out within The Netherlands TRAIL Research School for Transport, Infrastructure and Logistics, and financed by the Delft University of Technology.

The "Seamless Multimodal Mobility" research program will provide tools for the design and operation of attractive and efficient multimodal personal transport services. In this report we present the results of a study which is part of project 3, called "Dependable Scheduling". The main objective of this SMM-project is to develop models for the analysis and prediction of disturbance propagation in service networks. In this report we restrict ourselves to delay propagation in railway transport systems, which can however be seen as a typical example of a service network.

The objective of this study is to assess once and for all the practical value and empirical validity of Wakob's approach for the capacity assessment of railway stations. Additionally, we examine whether the approach is also effective for analysing delay propagation. The report should be understandable for both researchers and interested railway companies, represented in the Netherlands by task organisations Railned (capacity management), NS Railinfrabeheer (maintenance of railway infrastructure), and NS Verkeersleiding (traffic control), and Holland Railconsult (railway consultancy).

This study has been performed by Antoine de Kort (Faculty of Civil Engineering and Geosciences), Bernd Heidergott, Gerard Hooghiemstra and Robert-Jan van Egmond (Faculty of Information Technology and Systems) under the supervision of Ingo Hansen (Faculty of Civil Engineering and Geosciences) and Geert Jan Olsder (Faculty of Information Technology and Systems). Many thanks to all the above participants for their contribution in this indepth research. Furthermore, we would like to thank Piet Bovy for his editorial suggestions.
SUMMARY

The scope of this study is the analysis of delay propagation at railway stations. Delay are caused by variations in the actual train running times and station dwell times. One way to take account of these uncertainties is to represent the train movements by a queueing network. Wakob has proposed an analytical framework for capacity assessment of railway stations which is based on queueing theory. More precisely, he applies queueing theory to predict the waiting time incurred by the simultaneous arrival and random processing of two trains at isolated parts of the infrastructure.

This report contains an indepth assessment of the practical value and empirical validity of Wakob's approach for both railway capacity planning and delay propagation analysis. In addition, this report can be regarded as a manual for railway staff who want to adopt Wakob's method in practice.

The assumed queueing characteristics appear suitable for capacity planning. That is, Wakob's method can indeed be useful for capacity assessment. However, the sources of randomness involving delay propagation are substantially different. Therefore, the method appears to be inappropriate for analysing delay propagation.

Wakob's method is a "timetable"-free approach. In this report we show that systems running under a given timetable cannot be used to verify or falsify Wakob's approach, nor that they can provide reasonable input data. That is, the results of Wakob's approach cannot be compared to daily observations. Additionally, the waiting times are expected to be generally larger than those obtained via simulations. Therefore, Wakob's method should only be adopted as a first approximation for the capacity assessment of railway stations.

In the course of the assessment study, the method has been applied to station The Hague HS in the Netherlands. The case study indicates that the approach is indeed able to locate the bottlenecks of this particular station. Moreover, it proves that the method is also useful for capacity assessment of railway stations in the Netherlands. However, the method seems to be rather uncomfortable for the practical use by railway staff due to the substantial efforts that are required to implement and to maintain the algorithms of Wakob's approach.

The results of this study allow the definition of future research for developing an accurate model to predict delay propagation at railway stations. Future research should concentrate on the development of a new model which takes account of the processes as they arise in daily railway operation. Additionally, the model should take account of the real sources of randomness (delays) and the interactions between train movements imposed by the signalling system. Validation of the model is only possible if its outcomes can be related to quantities or realisations that are easy to observe in practice.
Chapter 1

INTRODUCTION

1.1 Analysing delay propagation at railway stations

Delays are often considered as one of the main reasons for the poor attractiveness of railway transport in the Netherlands. Although a lot can be done to prevent delays, they will inevitably arise, mainly because of the following reasons.

- The individual driving behaviour of the engine drivers affects the way train movements actually take place.
- A wide variety of rolling stock (with different characteristics) uses the same infrastructure.
- Strong fluctuations are present in the boarding times, alighting times and transfer times of passengers.
- The dispatchers or automatic train regulation systems cannot always set up the required route in time.

In particular, the delay of a single train may affect several other trains, that is, delays are propagated. In fact, propagation of delays is the main source for the delays experienced by the Netherlands Railways (NS). Delay propagation occurs for the following reasons. On the one hand, there is a large number of trains simultaneously using the infrastructure, whereas, on the other hand, the signalling system imposes strong restrictions on the way these trains are processed.

The system's punctuality can be improved substantially by preventing or confining delay propagation. The scope of this study is the analysis and prediction of delay propagation at railway stations. The motivation for our choice is the
following. Railway stations have limited capacity compared to the adjacent block sections due to the large occupation times of particular infra elements (e.g. platform tracks) and the extra safety margins that are required to set up and release routes at level crossings and switches. Consequently, railway stations are considered as the bottlenecks of the system. In fact, also delay propagation is most likely to be expected at railway stations since

- trains arriving (departing) in time expect a red signal if the block section ahead is still occupied by a delayed train (e.g. if two trains make level crossings) and

- trains may be forced to stop at a platform track longer than planned in order to prevent transferring passengers from missing their connection or until the route for departure is cleared.

The extent of delay propagation at stations indicates whether the station capacity is sufficient to execute a given timetable at a predefined level of punctuality. If not, the infrastructure management might decide to adjust or extend the existing railway infrastructure. In other words, this type of analysis provides new design criteria for railway infrastructure. In addition, predictions of delay propagation at stations enable timetable designers to obtain insight in how to optimise a timetable concept by adding buffertimes to the most 'delay sensitive' train successions.

Much progress has already been achieved in delay propagation analysis on railway tracks or routes. In fact, relative simple analytic models are available for this purpose thanks to efforts of (amongst others) Kraft [14], Mühlhans [15], Schwanhäußer [22] and Weigand [31]. However, it is much more complicated to perform delay propagation analysis at railway stations. In fact, it requires the use of either large-scale simulations or advanced mathematical techniques, in particular since we are interested in the impact of uncertainties accompanying the arrivals, dwellings and departures of trains.

In this report, we evaluate Wakob’s approach which offers a mathematical framework for the analysis of railway stations. In [30] Wakob uses queueing theory to predict the blocking times of particular parts of the infrastructure due to uncertainties in the intervals between and the order of successive train movements. The next section presents the main concepts of queueing theory and their railway traffic counterparts. Subsequently it is explained why this report focuses on Wakob’s approach.
1.2 Queueing networks

Timetables are basically constructed using fixed running times, dwell times and minimal headways of trains. During operation however, these quantities exhibit fluctuations due to various causes (see De Kort [5]). Considering the train movements at railways stations as queueing processes is one way to take account of these uncertainties. In this sense, the trains represent the customers while the railway tracks of the station and the signalling system represent the servers of the system (the signalling blocks represent the buffers). The trains arrive at the station area with stochastic interarrival times (actual headways). Subsequently, they claim the release of particular infra elements (block section, level crossing, switch, platform track, etc.) or, equivalently, a given amount of service time during which no other train is allowed to occupy those particular elements. Hence, the minimal headways can be regarded as the service times in railway traffic processes [5].

Queues arise if trains arrive at an occupied element: the arriving trains have to wait then. Note that both early and late arrivals as well as departures can indeed imply deviations from the mean interarrival times and service times, as is shown in Table 1.1. Hence, the waiting times can be interpreted as propagated delays because they reflect detentions caused by disturbances in the arrival and service process of preceding train movements.

<table>
<thead>
<tr>
<th></th>
<th>early departure</th>
<th>late departure</th>
</tr>
</thead>
<tbody>
<tr>
<td>early arrival</td>
<td>smaller actual interarrival time</td>
<td>smaller actual interarrival time</td>
</tr>
<tr>
<td></td>
<td>smaller actual service time</td>
<td>larger actual service time</td>
</tr>
<tr>
<td>late arrival</td>
<td>larger actual interarrival time</td>
<td>larger actual interarrival time</td>
</tr>
<tr>
<td></td>
<td>smaller actual service time</td>
<td>larger actual service time</td>
</tr>
</tbody>
</table>

Table 1.1: The implied actual interarrival times and service times (compared to mean values) if trains arrive and/or depart early or late.

Trains actually move from one server to another as they run along their routes through the station area. In other words: after leaving a queue, they enter the next one. Consequently, a queueing network representation is needed to describe the train movements at stations.

Note that the type of queueing networks mentioned above is characterised by the following three elements which should be carefully taken into account:

- *train type dependent* running times and minimal headways,
- *train type dependent* routes and train orders
  (and thus service disciplines),
• physical constraints on the train movements imposed by the signalling system, and operational constraints imposed by synchronisation control.

Moreover, the interarrival times and minimal headways are expected to have rather typical stochastic properties. The interarrival times during operation will generally exhibit very small fluctuations. On the other hand, rather large variations are expected in the actual occupation times of particular infra elements (e.g. platform tracks). Consequently, the minimal headways are likely to exhibit large fluctuations as well.

1.3 Queueing representations of train movements at railway stations

As indicated by Rubinstein and Melamed [20] it is impossible to fit all of the above system characteristics into a pure analytic model. In other words, one will always end up with numerical procedures and/or simulations if all elements need to be taken into account. Nevertheless, efforts have been put in developing queueing (network) representations for this type of systems. These efforts have resulted in the approaches by Van Dijk [6] and Wakob [30]. These approaches are briefly discussed below.

Van Dijk’s model is used to obtain approximate indications about the sufficiency of the station capacity. Although the model takes account of physical constraints, it is assumed that blocked train movements are lost in order to establish an analytical solution for the described system. In other words, the train disappears instead of incurring a delay. Due to this unrealistic assumption, the model is inappropriate for delay propagation analysis.

In contrast to Van Dijk’s approach, it is emphasised that Wakob’s approach does not give a queueing model of an entire railway station. Instead, he proposes an analytical framework for capacity planning. The main difference with Van Dijk’s model is that it describes the performance of isolated parts of the network instead of the station as a whole. He applies queueing theory to predict the waiting time incurred by the simultaneous arrival of two trains at the isolated parts of the infrastructure.

Sources in literature mentioning Wakob’s approach, state rather contradicting opinions about its appropriateness for capacity assessment. Some authors are quite positive and even refer to applications of the method in practice (Deutsche Bahn [4], De Kort [5] and Sitz et al. [26]). However, the method’s value is questioned by Odijk, see [17]. Presumably, this is why it has only been applied to some stations in Germany up till now. In our opinion, the negative critics are due
to a limited understanding of the procedures involving Wakob's approach rather than due to inconsistencies in the method itself. Still, it is unknown whether the approach is also effective for analysing delay propagation. Additionally, it has not yet been assessed whether the approach gives valid results for railway stations in the Netherlands. These contradictory viewpoints were the motivation to undertake an indepth assessment of Wakob's approach to railway capacity planning and delay propagation.

1.4 Research objectives

This research aims to assess the practical value and empirical validity of Wakob's approach. Therefore, the following objectives are pursued.

1. To describe and explain the method, including its assumptions (that is, to assess whether the approach is behaviourally sound).
2. To assess whether the approach is mathematically sound.
3. To assess its practical value for capacity planning.
4. To assess its empirical validity (that is, does the method yield reasonable results if applied to capacity planning?).

However, main concern is the study of delay propagation. If Wakob's method appears to be sound and effective for capacity assessment of railway stations (even/also for the Dutch railway system), we want to know whether it can also be used for analysing delay propagation.

Therefore, an additional research objective is

4. to assess its practical value for predicting delay propagation at railway stations.

The results of this study will help us to define future research for developing an accurate model for analysing and predicting delay propagation at railway stations.

1.5 Outline of this report

Our strategy to pursue the above objectives, was as follows. First, we have thoroughly examined Wakob's thesis and other publications on this subject in order to fully understand Wakob's approach. By doing so, we obtained insight
in its behavioural and theoretical soundness as well as in its practical usefulness for capacity planning and delay propagation analysis. Subsequently, Wakob's method has been applied to station The Hague HS in the Netherlands, to check whether the approach is indeed able to detect the bottlenecks of this particular railway station.

This report is organised as follows. In chapter 2 we explain Wakob's motivation for using queueing theory as a theoretical framework for his investigation. We show that the assumed model characteristics are suitable for capacity planning. That is, the method may indeed be useful for capacity assessment. Meanwhile however, we indicate that the sources of randomness involving delay propagation are substantially different.

In chapter 3 we assess the theoretical soundness of the method by examining the procedures involving the model specification. In chapter 4 we do the same with respect to the stationary waiting time calculations.

To assess its empirical validity for capacity assessment, we apply Wakob's approach to station The Hague HS, one of the main railway stations of The Hague in the Netherlands. Some additional problems are detected during the case study. In chapter 5 we present the practical implications and the numerical results of this case study.

The overall conclusions are stated in chapter 6. This chapter also indicates how future research on the subject of delay propagation at railway stations should be organised.
INTERMEZZO:
TRAIN SERIES DEFINITION

We assume throughout the sequel that each train visiting a station belongs to a particular train series. In this report, each train series is identified by the following features

- the serial number distinguishing the heading direction (e.g. odd numbers for trains running from origin station X to destination station Y and even numbers for trips from Y to X),
- the train type (intercity/IC, express train/IR, stop train/AR or freight train/FR),
- the route used to traverse the station,
- the number of trains of each series visiting the station during one hour (the frequency).

Distinguishing types of trains is necessary since the required minimal headways heavily depend on the applying dynamic characteristics such as acceleration and braking rates as well as maximum speed. In addition, the dwell times at the platform tracks often also depend on the corresponding types of trains. However, although speaking of types of trains, we actually refer to types of regular line services (IC/IR/AR/FR) instead of the types of rolling stock being used. As a consequence, the above mentioned dynamic characteristics used throughout the sequel represent averages values. For the situation in the Netherlands, the dynamic train characteristics are summarised in e.g. [12, pp. 85 & 89].
Chapter 2

CHARACTERISTICS OF WAKOB’S SINGLE–SERVER APPROACH

2.1 Introduction

In this chapter we discuss the assumptions which allow Wakob to apply queueing theory for quantifying the traffic performance of a railway station. It is emphasised again that the approach does not provide a queueing model of an entire railway station but rather a theoretical framework for capacity assessment. Indeed, Wakob only analyses the train movements at isolated parts of the station area. These specific parts behave like single server elements. A definition of these parts is given in §2.2. We indicate the possible problems involved with the separate analysis of isolated infra parts.

In order to apply queueing theory to the isolated infra parts, we have to specify the uncertainties of the arrival and service processes. In §2.3 we discuss the assumptions underlying the interarrival time specification. In §2.4 the same is done for the service time specification. In §2.5 and §2.6 we indicate the practical value of the method for capacity planning and delay propagation analysis respectively.

2.2 Identifying single–server elements: TFK’s

The station under consideration is partitioned into specific parts of the infrastructure carrying the single server identity. In fact, all basic infra elements (block sections, switches, platform tracks, merging points, level crossings) have this property as only a single train is admitted to occupy each of them during a given time
slot. However, it would not be efficient to consider them all separately since stations generally consist of a very larger number of basic infra elements. Instead, Wakob therefore constructs sets of basic infra elements which behave like single server points as well. By doing so, we restrict the number of queueing systems to be examined.

Let a set $C$ consist of basic infra elements $s_1, s_2, \ldots, s_p$ and assume that $C$ carries the single server identity. If an arbitrary infra element, $s_i, i \in \{1, \ldots, p\}$ of $C$ is occupied by one train, then the single server identity implies that all other infra elements inside $C$ are also blocked. Consequently, the entire set of basic infra elements cannot be used by any other train during the same time slot. Generally, complex junctions like $C$ satisfying this condition correspond with common parts of several routes. That is why they are called *Teilfahrstraßenknoten* in German. Throughout this report, we only use the abbreviation of this German word: TFK. Figure 3.1 on page 17 shows an example of a TFK demarcation inside a station area.

There is also a practical reason for decomposing the station area into single server elements. For dimensioning the infrastructure, one is often interested in locating the bottleneck(s) of the system. Via identifying TFK’s more specific information is obtained compared to a global investigation of waiting time development inside the system as a whole.

The identification of TFK’s depends on the planned routes through a station. In general, it is not known in advance (i.e. before a timetable is available) which train series will dwell at which platform track. Consequently, no proper selection of TFK’s can be defined either. This problem can be tackled by considering the routes most commonly used in recent operation history by the main lines visiting the station. This solution is also used by Wakob. In addition, this allows for taking into account the different routes and platform tracks that might incidentally be used during operation in case of calamities.

Wakob assumes that all TFK’s have infinite queueing space. This prevents trains from getting lost/blockcd if a queue has reached its capacity. As a result, the TFK queueing processes are independent. In fact, the main reason for assuming infinite queueing spaces is that it enables Wakob to analyse the identified TFK’s separately [32]. Of course, it is admitted to split up the network into TFK’s. However, we might not be able to reproduce specific network properties properly when considering the isolated parts of the network separately. In other words, the properties of the simplified representation should match those of the original system. In view of this, we detect two probable aspects that may not be properly accounted for in this way.
Discrepancy 1: Unaccounted coherence between train movements at TFK’s on the same route

The traffic processes at adjacent TFK’s are expected to be behave in a similar way since they are visited by the same trains. Coherence in the arrival patterns becomes immediately apparent for TFK’s lying on the same route because the uncertainties are directly related to the interarrival times of trains belonging to the same series. But there also exist coherence between the service processes at TFK’s along the same route. This can be explained as follows. Generally, the lengths of TFK’s are quite small compared to the lengths of the passing trains. As a consequence, an arbitrary train will actually occupy several TFK’s at the same time. This implies that the service characteristics of adjacent TFK’s are also comparable (at least to some extent).

The coherence between traffic processes at adjacent TFK’s may be lost if the TFK’s are examined separately. However, the proposed queueing model specification ensures that this discrepancy does not occur (see §3.3 and §3.4).

Discrepancy 2: Unmodelled interactions between the TFK queueing processes

Second, a train that occupies a certain TFK may block several other train movements. Therefore, the traffic processes show strong interdependencies. Wakob indeed recognises that the separate analysis of TFK’s may give rise to unmodelled interactions between them. However, he does not deal with this problem explicitly, as we show in §4.5.

The waiting locations do not belong to the TFK’s. In practice, they will often be part of other TFK’s. This complicates the retrieval of observations in practice as we have to look elsewhere in the network to obtain the corresponding waiting times. Consequently, even if we are at all able to relate any observations to the correct TFK’s, we still need to decide which waiting time fraction is contributed by the individual TFK’s. We thus expect that validation by ‘real life’ observations/experiences will be very difficult.

2.3 Assumptions on the TFK arrival process

As Wakob’s approach is meant for capacity assessment, it will be used for strategic issues in which case timetables are generally unavailable. Therefore, Wakob assumes the trains to arrive in random order. Furthermore, he assumes that the observed interarrival process at the identified TFK’s can be approximated by an Erlang distribution [30]. Erlang distributions are in fact Gamma distribution with an integer shape parameter.

Note that the interarrival time fluctuations do not represent probable delays (at arrival). Instead, they indicate the possible intervals that may appear between
consecutive train movements according to the unknown timetable. In other words, the interarrival time variations reflect deviations from the average time interval which is based on the total number of arriving trains and the length of the observation period (see §3.3).

2.4 Assumptions on the TFK service process

Wakob assumes that the service times, expressed in terms of minimal headways, are also Erlang distributed. For each possible train succession, the minimal headway is determined along the common route part through the entire station rather than the length of the TFK under consideration. This is explained by the fact that the virtual occupation of a TFK consists of the actual running time of the first train along the entire distance between the clearing points of the corresponding signalling block(s) and the required braking distance of the following train (including safety margins).

Subsequently, Wakob imposes variations on the minimal headways by considering worst case scenarios in which two trains arrive at a TFK simultaneously. Wakob assumes that the arriving trains are processed in random order, that is either or not according to the priority rules that apply. This results in upper and lower bounds for the minimal headways that will be needed in practice. In §3.4 we explain how the minimal headway bounds are determined. The variations increase with the number of train series and priority levels. In fact, the order in which successive trains are processed is the only stochastic aspect Wakob uses to specify the TFK service process.

Simultaneous arrivals may be useful for capacity planning since they provide insight in the worst case situation. Note that, in practice, an arrival implies either the actual occupation of a specific infra element or just its release (which occurs as soon as a train passes the insulted joint ahead of the main signal). Although several trains may simultaneously claim the setting-up of routes including the same TFK, only one of them will expect a line-clear signal for actually occupying its route until the release of the TFK. Consequently, the assumption of simultaneous arrivals does not completely match with daily operation. Therefore, a validation of Wakob's method via simulations is only possible to a certain extent.

In case of delay propagation analysis, variations in running times and dwell times should be considered as the real source of service time randomness. In fact, this might require the use of different distribution types having substantially bigger variances than those proposed by Wakob.
2.5 The practical value for capacity planning

The approach can be useful for capacity assessment as it takes into account the essential sources of randomness that apply to these situations. However, rather pessimistic results (in terms of waiting times) are to be expected because of the underlying worst case assumptions.

We have indicated that the outcomes can hardly be validated. Still, the approach may be useful to obtain qualitative or approximate traffic performance indications for TFK's and stations as a whole. In fact, this is confirmed by the studies performed in Germany (see DB [4] and Sitz et al. [26]). The model's empirical validity has been assessed for railway stations in the Netherlands by means of a case study. The results are discussed in chapter 5.

2.6 The practical value for delay propagation analysis

From the assumptions made by Wakob, we conclude that his approach appears to be inappropriate for delay propagation analysis for the following reasons.

- Wakob does not take into probable delays (at arrival). Moreover, the service time fluctuations should be based on running time variations and dwell time variations instead of simultaneous arrivals and random order of train processing.

- The mechanism behind delay propagation is explained by interacting train movements at different locations inside the station area. However, the TFK queueing processes are assumed to be independent in Wakob's approach.

- Wakob's method returns upper bounds for the total waiting time instead of its mean value (and its standard deviation) for likely to be expected disturbances/delays.
Chapter 3

TFK QUEUEING PROCESS SPECIFICATION

3.1 Introduction

In this chapter we explain the procedures developed by Wakob to analyse the individual TFK's by means of queueing theory. This is done to assess the method's theoretical soundness. Note however, that he does not really develop a train process model. Instead, he applies queueing theory to predict the detentions (waiting times) incurred by two trains arriving simultaneously and being processed in random order.

A queueing process is characterised by the arrival process, the service process and the number of servers. By definition, TFK's are parts of the infrastructure carrying the single server identity. In §3.2 we indicate how TFK's can be properly identified inside the station area. Section 3.3 describes how the arrival process for a single TFK is specified. The TFK service process specification is discussed in §3.4. In §3.5 we present the conclusions on the method's theoretical soundness.

3.2 Demarcation of TFK's inside the railway station

Recall that each TFK should behave as a single server element. The TFK's can thus be found by examining whether common parts of several routes are fully blocked as soon as they are occupied by a single train. This procedure is illustrated below.

A route is specified for each train movement through the station area. Assume (without loss of generality) that the trains of each train series traverse the station
via the same route. Along these routes, the trains run from one basic infra element to another. Now consider $C$ which, for instance, consists of two basic infra elements, $s_1$ and $s_2$. Obviously, these elements may lie on several routes. Furthermore, assume that $C$ is visited by trains of three different series: $r_1$, $r_2$ and $r_3$. The set $C$ can only be considered as a single server point if a train moving along one of the three routes, excludes all other train movements across $C$ at the same time. Therefore, at least one of the elements $s_1$ and $s_2$ must lie on all three routes to ensure that a train movement of one series inevitably excludes the other two.

The expression below approximates the probability of an infra element or a set of them to be blocked for all other train movements as soon as it is occupied by a single train. This "Verkettungszahl" or level of interference is defined as [18, 19]:

\[ \varphi = \frac{1}{N^2} \cdot \sum_{i=1}^{R} \sum_{j=1}^{R} n_i \cdot n_j \]  

(3.1)

where $n_i$ denotes the number of train movements corresponding with the different train series/routes $1, 2, \ldots, R$ and $N$ is the total number of trains traversing during a fixed period of time (e.g. 24 hours).

In other words, a part of the station area may be considered as a TFK if the corresponding value for $\varphi$ equals one [23]. Values less than one indicate that more than one train movement can take place at the same time and consequently, the location should be represented by a multiserver queue.

Note that $\varphi$ can easily be expressed without using the absolute train numbers per series/route combination, thus yielding a more theoretical expression for the Verkettungszahl of train movements at the infra elements under consideration. We propose the following definition:

\[ \varphi = \frac{R + 2 \sum_{i=1}^{R} \sum_{j>i}^{R} \delta_{ij}}{R^2} \]  

(3.2)

where $R$ is the total number of series/route combinations traversing the infra elements under consideration and $\delta_{ij}$ is defined as:

\[ \delta_{ij} = \begin{cases} 1, & \text{if at least one infra element inside the TFK is used by both } i \text{ and } j, \\ 0, & \text{else}. \end{cases} \]  

(3.3)

Figure 3.1 shows an example of a TFK demarcation. Trains visiting the displayed part of the station run along four different routes: $r_1$ from point A, via junction 1 (switch) and junction 2 (level crossing) to point D, $r_2$ from A via junctions 1, 3 (switch), 4 (switch) and 2 to C, $r_3$ from A via junctions 1, 3 and 4 to E and $r_4$ from B via junctions 3 and 4 to E. At junction 1 trains along $r_1$ and $r_2$ are
mutually excluded so TFK$_1$ should at least contain junction 1. Junction 2 cannot be included in this TFK as it is not used by trains running along route $r_3$. Finally, three TFK’s are obtained. Obviously, it would also be allowed to identify two separate TFK’s instead of TFK$_3$ containing junctions 3 and 4 respectively. This shows that the identification of TFK’s is not necessarily unique. However, the number of TFK’s should be restricted to a minimum as a queueing model is to be specified for each TFK. This can be achieved by letting the individual TFK’s consist of as much conflicting points as possible. Still, at each TFK it must hold that the train movements along each pair of routes are mutually excluded at the same time at least at one point inside the TFK.

![TFK's identified for a simple station layout](image)

### 3.3 Arrival process specification

Wakob assumes the trains to arrive in random order because timetables are generally unavailable for capacity planning purposes. Furthermore, he assumes that the observed interarrival times at the identified TFK’s can be approximated by an Erlang distribution [30]. This type of distribution is fully characterised by the mean interarrival time $1/\lambda$ and the number of phases $k$.

The intensity $\lambda$ of the arrival process can obviously be estimated by

$$\hat{\lambda} = \frac{N}{T},$$

where $T$ is the observation period within which the arrival process is assumed to be stationary. Again, $N$ denotes the total number of trains that arrived during this period.
The parameter $k$, also called the coefficient of variation for the interarrival times, must be an integer in case of an Erlang-type process. Since the expected value for the interarrival time $A$ equals $E[A] = 1/\lambda$ for an Erlang-$k$ distribution, while its variance is given by $\sigma_A^2 = 1/(k\lambda^2)$ (see [9]), it follows directly that:

$$k = \frac{(E[A])^2}{\sigma_A^2}. \quad (3.5)$$

The sample variance of the interarrival times, denoted by $S_A^2$, is used as an estimator for $\sigma_A^2$. It can be obtained via the method of moments. Substitution of $\lambda$ and $S_A^2$ gives the following estimator for $k$:

$$\hat{k} = \left(\frac{1}{\hat{\lambda}}\right)^2 \cdot \frac{1}{S_A^2}. \quad (3.6)$$

Generally, the estimate of $k$ will not be an integer regardless of the applied estimation procedure. Hence, one should focus on the Gamma-distribution with density

$$f(x) = \frac{1}{\Gamma(t)} x^{t-1} e^{-ax}, \quad (3.7)$$

or take $k = \lfloor \hat{k} \rfloor$.

For the Erlang-distribution, $t = k$ and $\alpha = k\lambda$ should be substituted in (3.7). The maximum likelihood estimators for $\lambda$ and $k$, based on a sample of interarrival times $x_1, \ldots, x_n$, are given by

$$\frac{1}{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{T}{N},$$

as before, and

$$\log \hat{k} - \log \hat{\lambda} + \frac{1}{n} \sum_{i=1}^{n} \log x_i = \int_{0}^{\infty} y^{k-1}(\log y)e^{-y} dy \frac{\Gamma'(k)}{\Gamma(k)}. \quad (3.8)$$

The estimator $\hat{k}$ can be found numerically by solving (3.8) with Gauss-Newton iteration.

Note that the coefficient of variation $\hat{k}$ is directly related to the number of trains visiting the TFK under consideration (see (3.5)). Moreover, equal numbers of trains visit all TFK’s on the same route. As a result, Wakob’s arrival process specification indeed reproduces similar properties (at least to some extent) for TFK’s on the same route.
3.4 Service process specification

The characterisation of the TFK service times is based on the minimal headways between consecutive train movements. For each possible train succession, the minimal headway is determined along the common route part through the entire station rather than along the TFK under consideration only (see §2.2). By doing so, the same minimal headway applies to specific train successions at adjacent TFK's. In other words, the traffic processes at adjacent TFK's will indeed behave in a similar way, at least to some extent, as was required. Minimal headway procedures for station areas can be found in Schwanhäußer [24] (analytical expressions) and Uebel [28] (using simulations). Throughout the paper, we denote the minimal headway between a train of series $i$ followed by one of series $j$ by $z_{ij}$. In general, $z_{ij} \neq z_{ji}$ unless $i = j$, whereas $z_{ij} = 0$ if $i$ and $j$ do not interfere at all.

Recall that Wakob assumes that the service times are also Erlang distributed. Hence, the service process is fully characterised by the mean service time $1/\mu$ and the number of phases $l$. Therefore, the previously calculated minimal headways $z_{ij}$ are translated into a single 'average' minimal headway distribution for each TFK. To achieve this, Wakob introduces the relative frequencies of all possible train successions. These relative frequencies can be regarded as conditional probabilities. They are obtained as follows. Recall that $N$ denotes the total number of trains that arrive during the given period of time. Therefore, $N = \sum_{i=1}^{R} n_i$ where $n_i$ represents the number of trains of series $i$. $R$ denotes the total number of train series traversing the TFK. Since the trains arrive in random order, a train of series $j$ appears with probability $n_j/N$. Hence, the conditional probability that $j$ supplies the next arrival given a train of series $i$ has arrived before is equal to:

$$\text{Prob}\{j \text{ arrives next } | i \text{ has arrived before}\} = \frac{n_j}{N} \cdot n_i . \quad (3.9)$$

The right-hand side of (3.10) is equivalent to the relative frequency expression proposed by Wakob. From now on we denote this quantity by $h_{ij}$. Consequently, we have

$$h_{ij} = \frac{n_i \cdot n_j}{N} . \quad (3.10)$$

Clearly, $h_{ij} = h_{ji}$.

The weighted minimal headway $b_{ij}$ is subsequently obtained by multiplying the minimal headway $z_{ij}$ with the corresponding relative frequency $h_{ij}$:

$$b_{ij} = h_{ij} \cdot z_{ij} . \quad (3.11)$$

Finally, the 'average' minimal headway is defined as

$$\overline{B} = \frac{1}{N} \sum_{i=1}^{R} \sum_{j=1}^{R} b_{ij} = \frac{1}{N} \sum_{i=1}^{R} \sum_{j=1}^{R} h_{ij} \cdot z_{ij} . \quad (3.12)$$
Example:
Let a TFK be traversed by two different train series 1 and 2 with \( n_1 = 4 \) and \( n_2 = 2 \) respectively. Let the corresponding minimal headways be given by \( z_{11} = 3 \) min., \( z_{12} = 6.5 \) min., \( z_{21} = 8 \) min. and \( z_{22} = 4 \) min. Then the following results are obtained using (3.10) – (3.12):

\[
N = n_1 + n_2 = 6 \quad \Rightarrow \quad h_{11} = \frac{4^2}{6} = \frac{8}{3}, \\
h_{12} = h_{21} = \frac{4 \cdot 2}{6} = \frac{4}{3}, \\
h_{22} = \frac{2^2}{6} = \frac{2}{3}
\]

\[
\Rightarrow \bar{B} = \frac{\frac{8}{3} \cdot 3 + \frac{4}{3} \cdot \frac{18}{2} + \frac{4}{3} \cdot 8 + \frac{2}{3} \cdot 4}{6} = \frac{90}{18} = 5 \text{ [min]}
\]

Next, Wakob introduces minimal headway fluctuations by considering the order in which pairs of trains are processed. In fact, he considers the extreme situation in which both trains arrive at the same time. Note that each train succession like “\( i \)” before “\( j \)” and “\( j \)” before “\( i \)” will occur equally often due to the assumed random order of train arrivals. However, the order in which they traverse the TFK depends on the priority rules. Normally, in case no timetable is available, the train with the highest priority is allowed to run first. Consequently, the other train incurs an additional blocking time on top of the minimal headway due to the simultaneous arrival of “\( i \)” and “\( j \)”. This extra blocking time is referred to as the disposition time. Its amount depends on whether the priority rules are accomplished. The disposition time, denoted by \( d_{ij} \), can be interpreted as the minimal time slot required to allow the trains to run such that their order of succession is the opposite of the priority rule:

\[
d_{ij} = \begin{cases} 
  z_{ji}, & \text{if } i \text{ has the highest priority}, \\
  0, & \text{if } i \text{ and } j \text{ have the same priority}, \\
  -z_{ij}, & \text{if } j \text{ has the highest priority}.
\end{cases} \tag{3.13}
\]

The negative value of \( d_{ij} \) (last option in (3.13)) indicates that the train of series \( i \) wins time equal to \( z_{ji} \) because it is allowed to proceed the train of series \( j \), whereas the latter one would normally have priority.

Note that \( d_{ij} + d_{ji} = 0 \) for all train series \( i \) and \( j \).
The worst case minimal headway, denoted by \( z^*_{ij} \) is then given by

\[
  z^*_{ij} = z_{ij} + d_{ij} .
\] (3.14)

Combining (3.13) and (3.14), we obtain: \( z^*_{ij} \in \{0, z_{ij}, z_{ij} + z_{ji}\} \). Accordingly, (3.14) results in either an upper bound or a lower bound for the minimal headway that will apply in practice. Evidently, the upper and lower bound are equal if both train series have the same priority. That is, it makes no difference which train is allowed to pass first.

Note that the worst case minimal headways are based on the assumption that pairs of trains arrive at the same time. In practice however, only calls for the release of a particular TFK can be simultaneously submitted. That is, the actual occupation of a TFK by successive trains always takes place with certain intervals (realisations of the planned interarrival times according to the timetable). Consequently, less disposition time will generally be required. In other words, the uncertainty in the order of succession will have fewer impact on the train movements in daily operation.

A definition for the average worst case minimal headway is obtained by replacing \( z_{ij} \) in (3.11) and (3.12) by \( z^*_{ij} \). This quantity is denoted by \( \overline{z}^* \). Using the properties \( d_{ij} + d_{ji} = 0 \) and \( h_{ij} = h_{ji} \), it follows that

\[
  \overline{z}^* = \frac{\sum_{i=1}^{R} \sum_{j=1}^{R} h_{ij} z^*_{ij}}{N} = \frac{\sum_{i=1}^{R} \sum_{j=1}^{R} h_{ij} z_{ij} + \sum_{i=1}^{R} \sum_{j=1}^{R} h_{ij} d_{ij}}{N} \equiv \overline{B} .
\] (3.15)

Wakob uses the above expression as an estimation for the mean service time at the TFK under consideration. Note that (3.15) implies that on average the service time is not affected by the additional disposition times.

Example (continued):

Let a TFK be traversed by the same train series 1 and 2 whose characteristics are given in the example on page 20. Assuming trains of series 1 would normally have the highest priority, the following worst case headways are obtained according to (3.14):

\[
\begin{align*}
  z^*_{11} & = z_{11} + d_{11} = 3 + 0 = 3 \text{ [min]} \\
  z^*_{12} & = z_{12} + d_{12} = z_{12} + z_{21} = 6.5 + 8 = 14.5 \text{ [min]} \\
  z^*_{21} & = z_{21} + d_{21} = z_{21} - z_{21} = 8 - 8 = 0 \text{ [min]} \\
  z^*_{22} & = z_{22} + d_{22} = 4 + 0 = 4 \text{ [min]}
\end{align*}
\]

It is remarked that both Anlage 1 in [30] and [24] use this quantity in the weighted blocking time definition instead of \( z_{ij} \) shown in (3.11), thus including the disposition time \( d_{ij} \). Regardless of which definition is chosen, the same result is obtained for (3.12) thanks to the property that \( d_{ij} + d_{ji} = 0 \forall i \forall j \).
Consequently, the average worst case minimal headway is found by substituting the resulted amounts into (3.15):

\[ Z^* = 5 \text{ [min]} \]

The sample variance of the worst case minimal headway \( Z^* \) can be derived as follows, using (3.15). Let \( M \) denote the sampled number of train successions \( (M = N^2) \). Using standard statistic procedures we get:\(^1\)

\[
S_{Z^*}^2 = \frac{1}{M-1} \cdot \left\{ \sum_{i=1}^{R} \sum_{j=1}^{R} n_i n_j \left( z_{ij}^* - Z^* \right)^2 \right\}
\]

\[
= \frac{N}{M-1} \cdot \left\{ \sum_{i=1}^{R} \sum_{j=1}^{R} h_{ij} (z_{ij}^*)^2 - 2Z^* \sum_{i=1}^{R} \sum_{j=1}^{R} h_{ij} z_{ij}^* + Z^*^2 \sum_{i=1}^{R} \sum_{j=1}^{R} h_{ij} \right\}
\]

\[
= \frac{N}{M-1} \cdot \left\{ \sum_{i=1}^{R} \sum_{j=1}^{R} h_{ij} \cdot (z_{ij} + d_{ij})^2 - N(Z^*)^2 \right\} \quad (3.16)
\]

Now define

\[
P = \frac{N}{2T} \cdot \sum_{i=1}^{R} \sum_{j=1}^{R} h_{ij} \cdot (z_{ij} + d_{ij})^2 \quad (3.17)
\]

which is the sum of the blocking times imposed on \( j \) due to occupation by \( i \) (see [30]).

Using this definition for \( P \), equation (3.16) can be rewritten as

\[
S_{Z^*}^2 = \frac{N}{M-1} \cdot \left\{ \frac{2TP}{N} - N(Z^*)^2 \right\} = \frac{2TP - MB^2}{M-1} \quad (3.18)
\]

The service time coefficient of variation, \( t \), can subsequently be estimated according to

\[
\hat{t} = \frac{Z^*^2}{S_{Z^*}^2} \quad (3.19)
\]

Substituting (3.15) and (3.18) in (3.19) gives

\[
\hat{t} = \frac{B^2 (M-1)}{2TP - MB^2} \quad (3.20)
\]

\(^1\)This derivation differs from the ones stated in [30] and [21]. Although the same approximate expression for \( \hat{t} \) is obtained in the end, our report shows that the estimate is already accurate if the squared number of trains \( M \) is large enough.
For large $M$, (3.20) can be approximated according to

$$\hat{l} \approx \left( \frac{2TP}{MB^2} - 1 \right)^{-1}.$$  \hspace{1cm} (3.21)

Generally, equation (3.21) will result in a non-integer value for $\hat{l}$ implying the service times to be \textit{Gamma} distributed rather than \textit{Erlang} distributed as assumed by Wakob. Some further remarks on the estimation of the coefficients of variation $k$ and $l$ are made in §4.2 and in §A.4.

### 3.5 Conclusions regarding the theoretical soundness

The previous sections show that the model specifications are mathematically sound. Moreover, the steps could be presented in a straightforward manner. In contrast with some critics found in literature, we thus conclude that Wakob's 'model' is theoretically consistent if the interference between adjacent TFK's is neglected. The stochastic properties of the interarrival times and service times, however, are represented only superficially. Therefore, the model can only be adopted to predict an upper bound for the average waiting time to be expected at the identified TFK's.
Chapter 4

TFK WAITING TIME CALCULATIONS

4.1 Introduction

In the previous chapter we described the construction of approximate single server queueing processes for the individual TFK's. Both the interarrival times and service times were assumed to be Erlang distributed in the sequel of this report. This type of queue is denoted by $E_k(\lambda)/E_l(\mu)/1$, according to the Kendall notation (see [9]), where the mean interarrival time equals $1/\lambda$ and the mean service time equals $1/\mu$. This chapter prescribes the stationary waiting time calculations to be performed for this type of queue. The chapter is organised as follows.

First we explain how an approximate $E_k(\lambda)/E_l(\mu)/1$ system can be properly obtained for the original $GI/GI/1$ system. Section 4.3 gives the definition for the TFK occupancy rate (or traffic rate) as well as its estimator. Conditions concerning the stationarity of the queueing system are based on this particular quantity. Section 4.4 presents the waiting time approximations proposed by Wakob. In §4.5 special attention is paid to the calculation of waiting times due to folding and crossing. Section 4.6 briefly discusses how these total waiting time calculations are adopted to assess the capacity of a railway station. Finally, section 4.7 contains some conclusions and remarks on the presented waiting times calculations.

4.2 Proper construction of the $E_k(\lambda)/E_l(\mu)/1$ system

As indicated in the previous chapter, Wakob assumes that the (unknown) TFK queueing process can be approximated by means of an $E_k(\lambda)/E_l(\mu)/1$ system.
Suppose that the real system is a $GI/GI/1$ queue with mean interarrival time $\bar{a}$ and mean service time $\bar{b}$. The standard deviations are $\sigma_a$ and $\sigma_b$, respectively. Below we present the standard way of approximating this $GI/GI/1$ queue by a queue with Erlang interarrival times and service times.

Let $m_a$ denote the smallest integer greater or equal than $\bar{a}^2/\sigma_a^2$ (cf. (3.6)) and set $\lambda_a = m_a/\bar{a}$, then $E_{m_a}(\lambda_a)$ is a suitable approximation of the interarrival time distribution. Similarly, we obtain $m_b$ and $\mu_b = m_b/\bar{b}$ and approximate the service time distribution by $E_{m_b}(\mu_b)$. As a result, the original $GI/GI/1$ queue is approximated by an $E_{m_a}(\lambda_a)/E_{m_b}(\mu_b)/1$ queue.

Some remarks on this approximation need to be stated here. First, the method is exact in the sense that if we start with an, say, $GI/E_1(\mu)/1$ queue, then the approximation yields $m_b = 1$ and $\mu_b = \mu$, that is, it recovers the $E_1(\mu)$ distribution. Analogously, exact representations are also obtained for original queues of types $E_k(\lambda)/GI/1$ and $E_k(\lambda)/E_1(\mu)/1$. A second point is that the approximation is only feasible if $\bar{a}/\sigma_a$ and $\bar{b}/\sigma_b$ are both greater than or equal to one. This stems from the fact that the coefficient of variation, that is, the squared mean value divided by the variance, of an Erlang-$k$ distribution is always greater or equal to one. Fitting an Erlang-$k$ distribution to a distribution with coefficient of variation less than one will underestimate the variance of the distribution.

It is easily verified that a coefficient of variation equal to one corresponds with an exponential distribution. The symbol $E_1$ will therefore be replaced by $M$ (Markov process) in the sequel of this chapter which is in accordance with the Kendall notation.

### 4.3 The TFK occupancy rate

The occupancy rate or traffic load has to be determined before we can calculate the mean stationary waiting time. The occupancy rate denoted $\varrho$ of a single server queue is defined as the ratio between the expected arrival and service rates:

$$\varrho = \frac{1/\mu}{1/\lambda} = \frac{\lambda}{\mu}. \quad (4.1)$$

For each TFK, the arrival rate $\lambda$ is estimated according to (3.4). The service rate is defined as the reciprocal value of the average worst case minimal headway $Z^*$ which in turn is given by (3.15). As a result, the TFK occupancy rate can be estimated by

$$\hat{\varrho} = \frac{\hat{\lambda}}{\hat{\mu}} = \frac{N}{T} \cdot \frac{1}{\overline{Z^*}} = \frac{N \cdot \overline{B}}{T}. \quad (4.2)$$
It is well known (cf. [2]) that for \( \rho < 1 \), a single server queue has a stationary waiting time distribution. From now on we will assume that this condition is satisfied.

Equation (4.1) implies that \( \rho, \lambda \) and \( \mu \) depend on each other through \( \lambda = \rho \mu \). Therefore, we may assume without loss of generality \( \mu = 1 \). To see this, assume that \( \mu' \neq 1 \) is the original mean service time, then \( \lambda' = \rho \mu' \) yields the adjusted arrival rate.

**Convention:**

For \( \mu = 1 \), the \( E_k(\lambda)/E_l(\mu)/1 \) queue is completely determined by \( k, l \) and \( \rho \). From now on, this queue is therefore denoted as \( E_k/E_l/1/\rho \). Furthermore, the mean stationary waiting time (from now on: average waiting time) of this particular queue is notated as \( E[W(k,l,\rho)] \).

Note that, in the Kendall notation, \( E_k(\lambda)/E_l(\mu)/1/Z \) denotes an Erlang queue with finite waiting room: the system can contain at most \( Z \) customers. Since \( 0 < \rho < 1 \), the misuse of our notation cannot give rise to confusion.

### 4.4 Waiting time calculations

Wakob's approximation for the average waiting time is based on the following formula:

\[
E[W(k,l,\rho)] = f(k,l) \cdot E[W(1,l,\rho)] .
\]  
(4.3)

The second factor at the right-hand side of (4.3) denotes the stationary waiting time of a queue with exponentially distributed interarrival times. The average waiting time of this queue is given by the *Pollackzek-Khintchine* formula (see [9]):

\[
E[W(1,l,\rho)] = \frac{\rho}{1-\rho} \cdot \frac{l+1}{2l} .
\]  
(4.4)

Combining (4.3) and (4.4) yields

\[
E[W(k,l,\rho)] = f(k,l) \cdot \frac{\rho}{1-\rho} \cdot \frac{l+1}{2l} .
\]  
(4.5)

The function \( f(k,l) = f_\rho(k,l) \) is called the *factor of variation*. For given traffic load \( \rho \), the function \( f(k,l) \) is obtained via multiple linear regression involving variations of the \( E_k/E_l/1 \) queueing system. This factor of variation improves the one introduced by Gudehus [11] and is one of Wakob's main results. However, the following calculations can also be performed if the factor of variation is obtained in a different way (e.g. via the methods according to Grübel [10] or Tijms [27]). We therefore postpone the description of Wakob's approximation to the appendix.
Note that $E[W(k, l, e)]$ denotes the average waiting time for the individual trains. However, Wakob is interested in the total amount of waiting time arising at the TFK under consideration. In order to achieve this, Wakob considers the mean queue length rather than the mean waiting time. Denote the mean queue length by $L_w$.

Little’s formula states that [9]

$$L_w = \lambda \cdot E[W] \quad (4.6)$$

In case of an $E_k/E_l/1/\rho$ process we obtain from (4.5) that

$$L_w(k, l, e) = \frac{\varrho^2}{1 - \varrho} \cdot \frac{l + 1}{2l} \cdot f(k, l) \quad (4.7)$$

when we use that $\mu = 1$ which implies $\varrho = \lambda$.

Recalling $\hat{\varrho} = NB / T$ and substituting (3.21) for $l$ in (4.7), we get the following estimation for $L_w(k, l, e)$:

$$\hat{L}_w(k, l, e) := \frac{\hat{\varrho}^2}{1 - \hat{\varrho}} \cdot \frac{\hat{l} + 1}{2\hat{l}} \cdot f(\hat{k}, \hat{l}) = \frac{1}{1 - \hat{\varrho}} \cdot \frac{N^2B^2}{T^2} \cdot \frac{1}{2} \left(1 + \frac{1}{\hat{l}}\right) \cdot f(\hat{k}, \hat{l})$$

$$\approx \frac{1}{1 - \hat{\varrho}} \cdot \frac{N^2B^2}{T^2} \cdot \frac{TP}{N^2B^2} \cdot f(\hat{k}, \hat{l}) = \frac{1}{1 - \hat{\varrho}} \cdot \frac{P}{T} \cdot f(\hat{k}, \hat{l}) \quad (4.8)$$

Eventually, the total waiting time is found by multiplying the expected queue length, $L_w$ by the predefined observation period $T$:

$$\hat{W}_{tot}(k, l, e) = \hat{L}_w(k, l, e) \cdot T = \frac{P}{1 - \hat{\varrho}} \cdot f(\hat{k}, \hat{l}) \quad (4.9)$$

which is easily explained by assuming unity waiting costs per unit of time.

Recall that $P = \sum_{i=1}^{R} \sum_{j=1}^{R} p_{ij}$ where $p_{ij}$ is the blocking time imposed on a train of series $j$ due to the occupation by a train of series $i$. Consequently, we can also represent the total waiting time as the sum of the waiting time of each train succession $i$ before $j$:

$$\hat{W}_{tot}(k, l, e) = \sum_{i=1}^{R} \sum_{j=1}^{R} w_{ij} \quad (4.10)$$

where

$$w_{ij} = \frac{p_{ij}}{1 - \hat{\varrho}} \cdot f(\hat{k}, \hat{l}) \quad (4.11)$$

The waiting time denoted by $w_{ij}$ in (4.11) is imposed on the train of series $j$ (the train of series $i$ does not have to wait). Recalling the modeled arrival and service process characteristics (see §3.3 and §3.4) this quantity reflects the waiting time
caused by two trains arriving simultaneously and being processed in random order. However, these waiting times should not be mistaken for observable waiting times (i.e. the actual waiting of trains inside the station)! Section §4.7 indicates how these waiting times should be interpreted then. Indeed, trains do hardly ever arrive at the same time nor are they processed in random order. However, with respect to capacity planning, this scenario provides insight in the worst case situation. By collecting information about all possible combinations of pairs of trains (series), Wakob’s method allows for assessing the station capacity (cf. (4.10) and see §4.6).

4.5 Waiting times due to folding and crossing

If trains cross or fold at a given TFK then the resulting waiting times may involve the train movements at several locations in front of it to be temporarily hindered. As a result, additional waiting times may arise at other waiting positions. Evidently, this spillback effect plays a decisive role in capacity planning of stations. Wakob therefore also records the waiting times due to folding and crossing. They can be easily retrieved from the total waiting time per TFK if the routes through the station area known: they correspond with the total waiting times $w_{ij}$ of train successions $i$ before $j$ which enter the TFK via different tracks (direction independent; see [30, pp. 88]).

Wakob indicates graphically which waiting positions are affected by waiting times due to folding and crossing (see e.g. [30, Figure 5.6, page 100]). However, we could not find any explicit statements on how to distribute these waiting times over the affected waiting positions. This seems insufficient for numerical analysis. However, we still expect the approach to be quite robust as it is based on worst case scenarios: the backward propagation of waiting times is expected to be much smaller than the impact of simultaneous arrivals.

4.6 Procedures for station capacity assessment

Wakob’s final objective is to assess the capacity of railway stations using the prescribed waiting times calculations. He does so, by increasing the number of trains until the total waiting time in (4.10) exceeds a predefined limit (called “Hochrechnung”). In [26] this principle is illustrated for the main railway station of Ulm, Germany.

The station capacity depends on the most intensively occupied TFK or, equivalently, the TFK with the lowest admitted increase of train movements. The
following upper limits for the admitted total waiting time are applied in Germany:

- \(0.6 \cdot T\) for the entire station area and
- \(0.3 \cdot T\) for the individual TFK's.

The above figures are based on experience and applications of the approximation by Gudehus [11]. For instance, an old rule of thumb is to never utilise more than 70\% of the capacity. This may explain why the admitted total waiting time for the entire station is set to \(0.6 \cdot T\). However, as indicated in [32] these values have not yet been validated in practice. Instead, German Railway adopts a critical limit of 200 minutes of total waiting time during the day (24 hours) at railway stations to achieve a good level of service [26]. Of course, this limit is also based on experience.

Schwanhäußer has proposed several procedures for determining the admitted train increase for the individual TFK's as well as the entire station (see [23]). Roughly speaking, an increase factor is determined for each TFK. This factor indicates to what extent the current number of trains visiting the individual TFK's may be increased before the admitted total waiting time is exceeded locally. Subsequently, the minimal value of these increase factors is chosen as the one that applies to the entire station. Note however, that adjusting the train amounts per TFK also affects the characteristics of both the arrival and service process. Therefore, the model specification and subsequent calculations (see chapter 3 and above) have to be repeatedly performed. Nowadays, computers allow to perform these iterative procedures rather fast and easy.

4.7 Conclusions and remarks regarding the waiting time calculations

The results in [7] show that Wakob’s waiting time approximation is very accurate. However, the approximation method returns only the mean value of the waiting time. Nowadays, numerical procedures have been developed to estimate very accurately the entire waiting time distribution which is often desired. Therefore, using such methods (e.g. Grübel [10]) is preferred.

It should be noted carefully that the calculated waiting times do not correspond with the physical waiting of trains. Instead, they indicate to what extent the desired timetable paths of the corresponding train series need to be shifted to a later time slot. These shiftings either imply (see [21, 25, 26])
• longer *scheduled* dwell times at the platform tracks, or

• longer *scheduled* dwell times at previously visited stations, or

• longer *scheduled* running times along the lane sections in between two consecutive stations.

Obviously, it is very difficult to observe or verify all these particular deviations in practice. According to Wendler [32] this is the main reason why the outcomes of Wakob’s model have not yet been validated. However, we expect that it will be very difficult to validate the method at all (see §2.4).

Nevertheless, the approach may be useful as a first approximation of the capacity of individual TFK’s and the railway station as a whole. In fact, this is confirmed by the studies performed in Germany (see DB [4], and Sitzmann and Eilers [26]). Moreover, it has not yet been assessed whether the method yields also reasonable results when applied to railway stations in other countries. Therefore, it still makes sense to perform a case study such as the one presented in the next chapter.
Chapter 5

CASE STUDY: STATION THE HAGUE HS

5.1 Introduction

For assessing the empirical validity of Wakob's method, the previously described procedures have been applied in a case study. The case study has been performed at station The Hague Holland Spoor (The Hague HS), one of the two main stations of the city of The Hague. This city lies in the western part of the Netherlands (see Figure 5.1).

Figure 5.1: Map of the Netherlands
Each day (24 hours) this station is visited by nearly 500 trains, belonging to 8 passenger lines in both directions, including the Thalys (the high speed train from Amsterdam to Paris and vice versa) and two freight lines (which pass during the night only).

Most interesting about this station is that trains heading to The Hague Central Station and vice versa make level crossings with trains coming from Amsterdam/Leyden and heading towards Rotterdam and vice versa (see Figure 5.2). Because the level crossings highly affect the station capacity, the approach by Wakob should at least be able to locate the bottlenecks caused by these crossing movements.

Figure 5.2: Detail of passenger lines serving the area near station The Hague HS

The following steps have been executed before the waiting times for each of the identified TFK's could be determined:

- definition of the priority rules for railway operation in the Netherlands,
- identification of the train series to be considered,
- identification of the TFK's at The Hague HS,
Chapter 5. CASE STUDY: STATION THE HAGUE HS

- specification of the arrival processes and
- specification of the service processes.

The above steps are discussed in subsequent sections.

5.2 Priority rule definition

Four types of trains (read: line services) are distinguished in the current Dutch railway network, represented by increasing priorities:

- freight trains (FR),
- stop trains (AR = AggloRegional),
- express trains (IR = InterRegional) and
- intercity trains, including the Thalys (IC = InterCity).

The above implies that intercity trains always get priority, whereas freight trains never get priority.

5.3 Definition of unique train series

As explained on page 7, each train series is characterised by a set of features (direction, train type, route, frequency). Trains of the same series but running across different routes through the station are therefore denoted by different numbers (e.g. 1501 and 1503). Initially, 21 different serial numbers have been identified for The Hague HS (based on the timetable during the season 1996–1997, see Nederlandse Spoorwegen [16]). However, several series use the same route as well as the same types of trains. They can accordingly be represented by only one of them which is referred to as the unique train series. Obviously, the corresponding (hourly) frequencies should be adapted by adding those of the left out (equivalent) series. By doing so, a collection of 16 unique train series was obtained for The Hague HS. The freight trains are assumed to visit the station only four times during the night. Moreover, they do not dwell at The Hague HS. Table 5.1 shows some information about the identified unique passenger train series. There is no difference between frequencies during peak hours and off-peak hours.

The Thalys trains serve The Hague HS only four times a day (in both directions). In order to specify hourly frequencies, its given daily number of visits has been equally distributed over the day (= 24 hours). This explains the noninteger values shown in Table 5.1 for the unique series 2101 and 9302.
<table>
<thead>
<tr>
<th>serial number</th>
<th>train type</th>
<th>direction from</th>
<th>to</th>
<th>freq.</th>
<th>platform track</th>
</tr>
</thead>
<tbody>
<tr>
<td>1501</td>
<td>IC</td>
<td>Rotterdam CS</td>
<td>The Hague CS</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>1502</td>
<td>IC</td>
<td>The Hague CS</td>
<td>Rotterdam CS</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1901</td>
<td>IR</td>
<td>Rotterdam CS</td>
<td>The Hague CS</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>1902</td>
<td>IR</td>
<td>The Hague CS</td>
<td>Rotterdam CS</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2101</td>
<td>IC</td>
<td>Vlissingen</td>
<td>Amsterdam CS</td>
<td>2.17</td>
<td>6</td>
</tr>
<tr>
<td>2102</td>
<td>IC</td>
<td>Amsterdam CS</td>
<td>Vlissingen</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3401</td>
<td>IR</td>
<td>Rotterdam CS</td>
<td>Hoorn</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3402</td>
<td>IR</td>
<td>Hoorn</td>
<td>Rotterdam CS</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3404</td>
<td>IR</td>
<td>Hoorn</td>
<td>Amsterdam CS</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5101</td>
<td>AR</td>
<td>Rotterdam CS</td>
<td>The Hague CS</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>5102</td>
<td>AR</td>
<td>The Hague CS</td>
<td>Rotterdam CS</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>5104</td>
<td>AR</td>
<td>The Hague CS</td>
<td>Rotterdam CS</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5401</td>
<td>IR</td>
<td>Dordrecht</td>
<td>Amsterdam CS</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>9302</td>
<td>IC</td>
<td>Amsterdam CS</td>
<td>Paris</td>
<td>0.17</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 5.1: The unique passenger train series visiting The Hague HS (1996-1997)

5.4 Identification of TFK’s at The Hague HS

A set of ten TFK’s has been identified at The Hague HS, according to the procedure given in §3.2 and based on the station layout and the routes used by the identified unique train series. Figure 5.3 presents the result schematically. The picture shows only the relevant infra elements of the station. Examining the configuration, we expected that the level crossings mentioned earlier should give rise to high amounts of waiting times at TFK$_2$, TFK$_7$, and TFK$_8$. Table 5.2 shows which unique train series pass the individual TFK’s.

5.5 Arrival process specification

The available timetable (for the season 1996–1997) is in fact an hourly pattern of train arrivals and departures which is repeated throughout the day. Therefore, it does not really make any difference whether the analysis is performed for only one hour or for a multiple of 60 minutes. However, several unique train series have fractional hourly frequencies, which is inconvenient to work with, whereas the daily frequencies of all unique series are integer valued. Hence, the observation period is set to $T = 1440$ min $\sim 24$ hours.
The average interarrival times (expressed in minutes) at each TFK can be straightforwardly estimated by the summed frequencies of the passing unique train series divided by $T$ (see equation (3.4)). However, estimation of the coefficient of variation, $k$, appears to be more complicated since it requires also the estimation of the interarrival time variance (see equation (3.5)). Wakob applies statistical analyses on real life train operation data to obtain these sample variances. However, no such data was available for The Hague HS. Instead, we have used the platform track occupation schedules corresponding with the timetable (in Dutch: Basis SpoorOpstellingen). These schedules indicate the moments of arrival and departure of the individual train series as well as the corresponding platform tracks that are used for dwelling. The estimation of sample interarrival time variances is performed in two steps. First, a matrix is constructed which contains the interarrival times between all relevant train successions (for the TFK under consideration). These interarrival times result by subtracting the corresponding arrival moments according to the platform track occupation scheme (results between 1 and 60 minutes). Afterwards, both the sample interarrival time average and variance can be straightforwardly obtained from this new data set. Figure 5.4 illustrates this procedure schematically.

One might argue that the resulting interarrival time variances may be quite different from the ones corresponding with situations in which the timetable is unknown. However, we always need a sample of interarrival times to estimate the unknown variance. In fact, Wakob also uses timetable information (indirectly) since the train operations he observes are actually realisations of scheduled processes. In fact, the train arrivals at The Hague seem to be planned quite randomly according to the coefficients of variance, $k$, which are only slightly greater than one for all TFK's (see Table 5.4, fifth column).
### Table 5.2: An overview of the unique train series visiting the individual TFK's

<table>
<thead>
<tr>
<th>TFK number</th>
<th>is used by the unique train series</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 3 5401 3401 2101</td>
</tr>
<tr>
<td>2</td>
<td>5101 3402 1901 2102 1501</td>
</tr>
<tr>
<td>3</td>
<td>1 5101 1901 1501</td>
</tr>
<tr>
<td>4</td>
<td>1 3402 2102</td>
</tr>
<tr>
<td>5</td>
<td>3 5101 5401 1901 3401 1501 2101</td>
</tr>
<tr>
<td>6</td>
<td>5102 5104 3404 1902 9302 1502</td>
</tr>
<tr>
<td>7</td>
<td>1 5102 3404 2102</td>
</tr>
<tr>
<td>8</td>
<td>5101 3404 1901 9302 1501</td>
</tr>
<tr>
<td>9</td>
<td>3 5101 5401 1901 1501</td>
</tr>
<tr>
<td>10</td>
<td>3 1901 3401 2101</td>
</tr>
</tbody>
</table>

#### 5.6 Service process specification

The specification of service times starts with the minimal headway calculations. The minimal headway for each possible train succession has been estimated as follows. First, we have determined which signalling blocks are visited along both routes of the corresponding unique train series through the entire station. The minimal headway for these common route parts has been set equal to the critical blocking time for the heading train plus the braking time required for the following train to come to a complete stop after running at maximum speed. The calculations are based on the distances between main block signals and the average train lengths and dynamic characteristics (see e.g. Hansen, De Kort and Wiggenraad [12] and Schwanhäußer [24]). We have to use average train characteristics since we distinguish types of lines services rather than types of rolling stock.

Additional assumptions have been made on the dwell times per train type. By doing so, we have achieved a compromise between the expected accuracy of the minimal headways and the required amount of detailed information about the infrastructure and the train characteristics. In fact, the minimal headway calculations required most of the computation time during the case study. Table 5.3 contains the values of train lengths, maximum speeds, acceleration and braking rates as well as dwell times for the different train types, that are assumed to apply inside the station area.
Chapter 5. CASE STUDY: STATION THE HAGUE HS

Chapter 5. CASE STUDY: STATION THE HAGUE HS

5.7 Numerical results

Table 5.4 contains the results of Wakob’s method applied to the TFK’s identified at The Hague HS. The rows present the following quantities for the individual TFK’s (from left to right): the index according to Figure 5.3, the total number of trains visiting the corresponding TFK’s during time slot $T$, between brackets the number of trains that would be observed during $T$ if the TFK’s were be visited by the same number of trains as during the busiest hours, the mean interarrival time, the mean service time, the interarrival time coefficient of variation ($\hat{k}$), the service time coefficient of variation ($\hat{l}$), the occupancy rate ($\tilde{q}$) and the total waiting time ($W_{tot}$).

Note that the coefficients of variation $\hat{k}$ and $\hat{l}$ are real values for all TFK’s. The procedures do not fully justify the use of real valued parameters $k$ and $l$ (see §A.4). However, we might interpret this by saying that Gamma distributions are considered rather than Erlang distributions.
Furthermore, the service time coefficients of variation, $\hat{\theta}$, appear to be reasonably high. This indicates that the specified minimal headway fluctuations are quite small. In other words, the dynamic characteristics inside the station area are comparable for all train types.

The results show that TFK$_5$ and TFK$_2$ produce the highest total waiting times as was expected due to the presence of level crossings. However, the total waiting time at TFK$_8$ is substantially smaller whereas level crossing train movements also occur at this location. Presumably, this is explained by the extra two minutes interarrival time (on average) compared to the mean value at TFK$_2$ (9.7 minutes versus 7.5 minutes). Indeed, that the outer track leading from Amsterdam towards the station is less frequently used than the inner track (2.17 trains per hour versus 4.17 trains per hour on average according to Table 5.1). Consequently, we expect the total waiting time at TFK$_8$ to be of the same order as the ones found at TFK$_2$ and TFK$_5$ if the arrival rate of the outer track is increased.

Observe that large amounts of waiting time also arise at TFK$_6$ and TFK$_7$, caused by the level crossings and merging movements of trains from Amsterdam with those coming from The Hague CS. In fact, the total waiting time at TFK$_7$ is comparable to that at TFK$_2$. However, this is mainly due to the presence of freight train movements via TFK$_7$ which require rather large minimal headways (see fourth column). Presumably, the same argument holds for the remarkably large total waiting time at TFK$_9$. Therefore, the results can be easily verified by discarding all freight train movements (implying that TFK$_1$, TFK$_3$ and TFK$_4$ do not have to be examined). Total waiting times between 70 and 80 minutes would then be expected to arise at TFK$_7$ and TFK$_9$, that is, comparable to that of TFK$_{10}$.

Note that the total waiting times at TFK$_5$ (139.8 minutes) is still less than the German Railway standard of 400 minutes (see [26]). This suggests that an increase of the number of train movements is still allowed. However, according to Railned, the organisation which is responsible for the capacity management of the railway network in the Netherlands, the capacity of a junction or an infra element is reached if the corresponding occupancy rate, $\phi$, is approximately 0.3. Indeed, this occurs both at TFK$_5$ and TFK$_2$, indicating that the available capacity at these elements is already fully utilised. Further research is required in order to assess which measure should be adapted to make it applicable to railway stations in the Netherlands.
current station: The Hague HS

observation period: $T = 1440 \text{ [min]}$

<table>
<thead>
<tr>
<th>TFK</th>
<th>#trains (max)</th>
<th>mean interarrival service</th>
<th>$\hat{k}$</th>
<th>$\hat{l}$</th>
<th>$\hat{\phi}$</th>
<th>$W_{\text{tot}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>time (min)</td>
<td>time (min)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>156 (168)</td>
<td>9.2</td>
<td>2.3</td>
<td>1.3</td>
<td>23.1</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>192 (192)</td>
<td>7.5</td>
<td>2.3</td>
<td>1.3</td>
<td>22.2</td>
<td>0.30</td>
</tr>
<tr>
<td>3</td>
<td>100 (100)</td>
<td>14.4</td>
<td>3.1</td>
<td>1.2</td>
<td>32.7</td>
<td>0.21</td>
</tr>
<tr>
<td>4</td>
<td>100 (100)</td>
<td>14.4</td>
<td>3.7</td>
<td>1.2</td>
<td>44.6</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>248 (264)</td>
<td>5.8</td>
<td>1.9</td>
<td>1.6</td>
<td>14.4</td>
<td>0.33</td>
</tr>
<tr>
<td>6</td>
<td>148 (168)</td>
<td>9.7</td>
<td>2.6</td>
<td>1.2</td>
<td>25.7</td>
<td>0.27</td>
</tr>
<tr>
<td>7</td>
<td>124 (124)</td>
<td>11.6</td>
<td>3.5</td>
<td>1.2</td>
<td>36.5</td>
<td>0.30</td>
</tr>
<tr>
<td>8</td>
<td>148 (168)</td>
<td>9.7</td>
<td>2.2</td>
<td>1.2</td>
<td>24.5</td>
<td>0.23</td>
</tr>
<tr>
<td>9</td>
<td>148 (148)</td>
<td>9.7</td>
<td>2.9</td>
<td>1.3</td>
<td>27.5</td>
<td>0.30</td>
</tr>
<tr>
<td>10</td>
<td>128 (148)</td>
<td>11.3</td>
<td>2.9</td>
<td>1.2</td>
<td>35.3</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Table 5.4: Results of Wakob’s approach applied to The Hague HS

5.8 Conclusions

The results of the case study corroborate that Wakob’s method is able to detect the bottlenecks of the station. This confirms that the approach is also useful to obtain a first approximation of the capacity of railway stations in the Netherlands. We expect the method to discriminate the performance of the TFK’s (in terms of total waiting times) even better if larger fluctuations appear in the minimal headways (e.g. due to more train series, types of line services or routes).

Originally, the method was performed by filling in special forms containing a stepwise calculation of the total waiting times. Obviously, most of these procedures allow for execution by means of a computer which drastically decreases the required calculation time. However, the implementation is rather complex, mainly where the specification of the infrastructure is concerned (TFK definition in terms of main signals, locating switches, level crossings and merging points). In fact, the specifications should be updated each time new series or routes are introduced. This makes the algorithms rather inflexible. Therefore, the method seems to be rather uncomfortable for the practical use by railway staff.
Chapter 6

OVERALL CONCLUSIONS
AND FURTHER RESEARCH

6.1 Overall conclusions

In this report we have assessed the practical value and the empirical validity of Wakob's approach for modeling and analysis of train movements at railway stations. The procedures proposed by Wakob appear to be theoretically sound, if the interference between adjacent TFK's is neglected. However, we expect the approach to be quite robust since the backward propagation of waiting times is much smaller than the impact of the assumed simultaneous train arrivals.

Wakob does not give a queueing model of an entire railway station. Instead, he adopts queueing theory as an analytical framework for capacity assessment. In fact, Wakob's approach is a worst case analysis since it predicts the extreme waiting time incurred by the simultaneous arrival and random processing of two trains at isolated parts of the infrastructure. These characteristics are suitable for capacity planning purposes. However, the relevant sources of randomness involving delay propagation are substantially different from the ones assumed by Wakob. Therefore, the method is inappropriate for delay propagation analysis.

Wakob's method is a "timetable"-free approach. In this report we have shown that systems running under a given timetable cannot be used to verify or falsify Wakob's approach, nor that they can provide reasonable input data. That is, the results of Wakob's approach cannot be compared to daily observations. Additionally, the waiting times are expected to be generally larger than those obtained via simulations. Therefore, Wakob's method should only be adopted as a first approximation for the capacity assessment of railway stations.

The case study has indicated that Wakob's method is also able to locate the bottlenecks of railway stations in the Netherlands. However, the method
seems to be rather uncomfortable for the practical use by railway staff since the implementation and maintenance of the prescribed steps require considerable efforts.

6.2 Further research

The method appears also to be useful for capacity planning purposes in the Netherlands. However, we showed that the prescribed limits for the admitted total waiting times do not fully agree with the values adopted by Railned. Applications of Wakob’s method to other railway stations in the Netherlands may give insight in the critical waiting time limits that apply to the Dutch situation. Furthermore, some drastic changes have been established since 1997, both with respect to the timetable (line services and frequencies) and the infrastructure (from two to four tracks between Rotterdam and Leyden). We therefore recommend the entire case study to be repeated in order to assess how these changes have affected the capacity of station The Hague HS.

Our main concern is the analysis of delay propagation. Since Wakob’s method cannot be used for this purpose, our future research will concentrate on the development of a new model. More specifically, we will first investigate how the initial delay of a single train develops when running along a series of conflict points (block signals, switches, level crossings) inside the station area. Several basic scenarios are to be examined, represented by a (common) railway station topology and a set of train series specifications.

The model should take into account the real sources of randomness (delays). That is, assuming small variations for the initial delay distributions and large ones for the dwell time distributions. The signalling system causes interactions between the train movements at these conflict points and will eventually give rise to delay propagation. The model should reflect train type dependent running times, minimal headways and routes (see §1.2). The outcomes will be validated by statistical analysis of train detection data and by simulations of train movements in real world railway stations. Validation is only possible if the results of the model have a practical meaning. That is, they should be somehow related to quantities or realisations that are easy to observe in practice.
BIBLIOGRAPHY


[6] DIJK, N.M. VAN “Where Would We Be Without Trains?” A Queueing Modelling to Determine Train Delays Internal paper, University of Amsterdam, 1994


[18] POTTHOFF, G. Verkehrsströmungslehre Transpress, VEB Verlag für Verkehrswesen, Bd. 1, Berlin, 1968 (in German)

[19] PRAHL, J. Ermittlung der Leistungsfähigkeit komplexer Fahrstraßenknoten Eisenbahntechnische Rundschau (ETR), 22, no. 6, pp. 245-250, 1976 (in German)


[22] SCHWANHAUSER, W. Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn Veröffentlichungen des verkehrswissenschaftlichen Institutes der RWTH Aachen, 20, 1974 (in German)


[24] SCHWANHAUSER, W. Eisenbahnbetriebswissenschaft, §3.4-1 - §3.4-3, Lehrstuhl für Verkehrswirtschaft, Eisenbahnbau und -betrieb der RWTH Aachen, 1993 (in German)


[28] Uebel Durchsatz von Strecken und Stationen Signal + Draht, 90, no. 4: pp. 5–10, 1998 (in German)


Appendix A

WAKOB’S STATIONARY WAITING TIME APPROXIMATION

A.1 Why using approximations?

In case of the $E_k(\lambda)/E_l(\mu)/1$ queue, the waiting time distribution and its mean value can be expressed with the help of the Laplace transforms of the involved Erlang distributions. A valuable source is Cohen, [3, page 329]. Denote by $W$ a random variable representing the stationary waiting time of the $E_k(\lambda)/E_l(\mu)/1$ queue. Provided $\rho = \lambda/\mu < 1$, we obtain

$$E[W] = \frac{\mu \rho}{2(1 - \rho)} \cdot \left\{ \frac{1}{(k + 1)\lambda^2} + \frac{1}{(l+1)\mu^2} + \frac{2}{\mu} - \frac{2}{\lambda^2} \right\} + \sum_{i=2}^{k} \frac{1}{\delta_i(1)}. \quad (A.1)$$

where $\delta_1(r), \ldots, \delta_k(r)$ are the unique zeros (defined for $r < 1$) of:

$$(\xi + l\mu)^i(-\xi + k\lambda)^k = r(l\mu)^i(k\lambda)^k, \quad (A.2)$$

with real part of $\xi$ larger than 0, and the values $\delta_i(1)$ are the limits for $r \uparrow 1$.

For practical purposes it is inconvenient to obtain these zeros. Therefore waiting time approximations are often used. Different approximations have been proposed in literature (e.g. Fischer and Hertel [8], Grübel [10], Gudehus [11] and Tijms [27]). Below we discuss the approach provided by Wakob [30].

To understand what follows we recall that we consider a “standardized” queueing system, i.e. we set $\mu = 1$ and therefore the traffic rate $\rho$ determines the system completely (see §4.3). Furthermore, the mean stationary waiting time of the standardized queueing system $E_k/E_l/1/\rho$ is notated by $E[W(k,l,\rho)]$. 

49
A.2 The "ersatz" systems

Wakob's idea is to analyse the influence of the phase number of the interarrival time and the service time separately. More precisely, he considers two versions of the $E_k/E_t/1/\varrho$ queue. One version with the phase number of the interarrival times equal to one, i.e. similar to an $E_1/E_t/1/\varrho$ queue, and the other with the phase number of the service times equal to one, i.e. similar to an $E_k/E_t/1/\varrho$ queue. Next, he substitutes the $E_1/E_t/1/\varrho$ queue by an $M/M/1$ queue with appropriate traffic rate $\varrho^*$, called the $\varrho^*$-system, and the $E_k/E_t/1/\varrho$ queue by an $M/M/1$ queue with appropriate traffic rate $\bar{\varrho}$, called the $\bar{\varrho}$-system. Then, he determines functions $a(k, l)$ and $b(k, l)$, such that

$$E[W(k, l, \varrho)] \approx E[W(k, 1, \varrho)]$$

with

$$E[W(k, 1, \varrho)] = a(k, l) \left( E[W(1, 1, \varrho^*)] \right)^{b(k, l)} ,$$

or, equivalently,

$$\log(E[W(1, 1, \varrho)]) = \log(a(k, l)) + b(k, l) \log(E[W(1, 1, \varrho^*)]) . \quad (A.3)$$

Generally, the traffic rates $\bar{\varrho}$ and $\varrho^*$ depend on $k$ and $l$. However, we suppress this dependency in our notation for the sake of simplicity. The log-linear connection (A.3) between the $\varrho^*$-system and the $\bar{\varrho}$-system enables $E[W(1, 1, \varrho)]$ to be easily calculated out of $E[W(1, 1, \varrho^*)]$. In the following we discuss the role of these "ersatz" systems in more detail.

A.2.1 The $\varrho^*$-system

Consider the $E_k/E_t/1/\varrho$ queue. Now disregard the phase number $k$ and consider the "remaining" $M/E_t/1/\varrho$ queue. From the Pollackzek–Khintchine formula we obtain for this system (see e.g. [9]):

$$E[W(1, l, \varrho)] = \frac{\varrho}{1 - \varrho} \cdot \left( \frac{l + 1}{2l} \right) . \quad (A.4)$$

By setting

$$\varrho^* = \left( 1 + \frac{1}{1 - \varrho} \left( \frac{l + 1}{2l} \right) \right)^{-1} \quad (A.5)$$

an $M/M/1$ system with traffic rate $\varrho^*$ is obtained so that

$$E[W(1, 1, \varrho^*)] = E[W(1, l, \varrho)] . \quad (A.6)$$

Equation (A.6) can be easily checked, since $\varrho^*$ is such that $\varrho^*/(1-\varrho^*) = E[W(1, l, \varrho)]$. The $\varrho^*$-system can be interpreted as capturing the influence of the phase number $l$ on the waiting time while disregarding $k$. 
A.2.2 The $\bar{\rho} - $ system

Suppose that, for given $\bar{\rho}$, the mean waiting time of the $E_k/E_\ell/1/\bar{\rho}$ queue is known. As a first step, this system is replaced by an $E_k/M/1$ queue with suitable traffic rate $\bar{\rho}$. That is, $\bar{\rho}$ is chosen such that

$$E[W(k, l, \bar{\rho})] = E[W(k, 1, \bar{\rho})].$$

It is well-known that the waiting time of an $E_k/M/1/\bar{\rho}$ queue is given by

$$E[W(k, 1, \bar{\rho})] = \frac{v}{1-v},$$

where $v$ is the solution of

$$v = \left(1 + \frac{1-v}{k\bar{\rho}}\right)^{-k}$$

on the interval $[0, 1)$.

The following two steps are performed in order to obtain $\bar{\rho}$.

1. The $E_k/E_\ell/1/\bar{\rho}$ waiting time is interpreted as the waiting time of an $E_k/M/1/\bar{\rho}$ queue, by setting

$$E[W(k, l, \bar{\rho})] = \frac{v}{1-v},$$

which yields for $v$

$$v = \left(1 + \frac{1}{E[W(k, l, \bar{\rho})]}\right)^{-1}.$$

2. For given $v$, $\bar{\rho}$ is found by means of (A.8) as

$$\bar{\rho} = \frac{1-v}{k(v^{-1/k} - 1)}.$$

The above algorithm returns a $E_k/M/1/\bar{\rho}$ queue with its waiting time equal to that of an $E_k/E_\ell/1/\bar{\rho}$ queue. Next, the phase number $k$ is disregarded: we simply suppress the fact that the queue has $k$ phases by considering it as an $M/M/1$ queue but with the same traffic rate $\bar{\rho}$. The resulting $M/M/1/\bar{\rho}$ queue then has waiting time $E[W(1, 1, \bar{\rho})].$

The $\bar{\rho}$-system can be interpreted as capturing the influence of the phase number $k$ on the waiting time while ignoring $\ell$. 
A.3 An algorithm for Wakob’s waiting time approximation

In this subsection we describe how (A.3) can be used to actually derive an approximation for $E[W(k, l, \varphi)]$. The algorithm involves three steps:

1. **(Deleting “l”)** Evaluate the $g^*$ system, that is, calculate $E[W(1, 1, g^*)]$ by means of equations (A.5)–(A.6).

2. **(Wakob’s formula)** Apply (A.3) to obtain $E[W(1, 1, \hat{g})]$, where we write $\hat{g}$ to distinguish the traffic rate obtained via the $g^*$ system, using (A.3), from the traffic rate $\bar{g}$ obtained via the algorithm given in §A.2.

Since the waiting time of an $M/M/1$ queue with traffic rate $\bar{g}$ is $\frac{1}{1 - \bar{g}}$, this determines $\hat{g}$ as

$$\hat{g} = \left(1 + \frac{1}{a(k, l) (E[W(1, 1, g^*)])^{b(k, l)}}\right)^{-1}.$$

3. **(Inserting “k”)** In order to obtain the waiting time of an $E_k/M/1/\hat{g}$ queue, find $\hat{\varphi}$ so that

$$\hat{\varphi} = \left(1 + \frac{1 - \hat{\varphi}}{k \hat{g}}\right)^{-1}.$$

Note that this is the “inverse” of “suppressing k” as mentioned in the paragraph concerning the $\bar{g}$-system (see §A.2.2). The waiting time is

$$E[W(k, 1, \hat{\varphi})] = \frac{\hat{\varphi}}{1 - \hat{\varphi}}$$

and we take

$$E[W(k, l, \varphi)] \approx E[W(k, 1, \hat{\varphi})] = \frac{\hat{\varphi}}{1 - \hat{\varphi}}. \quad (A.9)$$

A.4 The functions $a(k, l)$ and $b(k, l)$

The complexity of the algorithm in §A.3 depends on that of the functions $a(k, l)$ and $b(k, l)$. Wakob proposes simple-structured functions in order to make the above algorithm efficient. However, the functions have to be fitted to a given situation. First, a set of sample points $(k_i, l_i, \varphi_i)$, $1 \leq i \leq n$, is chosen and the $w_i = E[W(k_i, l_i, \varphi_i)]$ are evaluated either numerically or by looking them up in tables. Next, a parameterized model is built by assuming that $a(k, l)$ is a function of $k, l$ and two design parameters, say $A$ and $\alpha$. Using (multiple) linear regression, $A$ and $\alpha$ are fitted so that $a(k, l)$ yields the best approximation in the least squares sense. Similarly, $b(k, l)$ is written as a function of $k, l$ and two
design parameters, say $B$ and $\beta$. Again, linear regression is used to fit $B$ and $\beta$ to the data.

In particular, Wakob proposes

$$a(k, l) = 1 - A \cdot \left( \frac{k - 1}{k + 1} \right) \left( \frac{l - 1}{l + 1} \right) - (1 - A) \cdot \left\{ 1 - \left( \frac{2}{k + 1} \right)^\alpha \right\} \left\{ 1 - \left( \frac{2}{l + 1} \right)^\alpha \right\}$$

(A.10)

and

$$b(k, l) = 1 - B \left\{ \left( \frac{k - 1}{k + 1} \right) \left( \frac{l - 1}{l + 1} \right) \right\}^{\beta}$$

(A.11)

Wakob has determined the corresponding parameters as $\alpha = 0.2924$, $A = 0.8060$, $\beta = 0.1278$ and $B = 1.1375$. Observe that $\varphi$ is no parameter of $a(k, l)$ and $b(k, l)$, that is, the parameters are uniformly fitted in $\varphi$. On the interval $(0.2, 0.8)$ the relative error of this approximation is less than or equal to 4%. In the approximation, the phase number $l$ has the largest influence. For small $k$, the relative error is not greater than 1%. For extreme situations, like $\varphi$ close to 1, or $l << k$, the expressions for $a(k, l)$ and $b(k, l)$ have to be adjusted, see [30, §3.3.5, pp. 52–60].

The functions $a(k, l)$ and $b(k, l)$, as defined in (A.10) and (A.11), are well-defined for real-valued parameters $k$ and $l$. On the other side, the phase number $l$ is obtained by taking the smallest integer greater or equal than $\hat{k} = \tilde{a}^2/\sigma_a^2$ according to §4.2. One might regard $\hat{k}$ as the "true" value and want use $\hat{k}$ rather than the integer-valued $k$ for numerical evaluations. Wakob does so indeed. Although the above analysis does not fully justify the use of real-valued parameters $k$ and $l$, one might interpret this by saying that Gamma distributions are considered rather than Erlang distributions.

### A.5 Concluding remarks on Wakob’s approximation method

Wakob has proposed an approximation method for estimating $EW(k, l, \varphi)$ which is

- exact whenever the original system is either an $M/GI/1$, a $GI/M/1$ or an $M/M/1$ queue,

- feasible if the coefficients of variance $k$ and $l$ of the (original) interarrival times and service times respectively are both greater or equal to one and

- computationally efficient (fitting the parameters requires multiple linear regression but this is needed only once and one can resort to the values provided by Wakob under the conditions mentioned in §A.4).
A.6 A comparison of several approximation methods

Table A.1 shows some numerical results for arbitrary sets of parameters \((\lambda, k, \mu\) and \(l)\) and different types of distributions. The displayed waiting times are calculated by four different approximation methods, namely those by Wakob [30], Grübel [10], Tijms [27] and Fischer and Hertel [8]. Grübel uses fast fourier transforms and is used as a reference because of its proved high accuracy (typically, errors of order \(10^{-4}\); see also [7]).

<table>
<thead>
<tr>
<th>arrival distribution</th>
<th>service distribution</th>
<th>(\lambda^{-1})</th>
<th>(k)</th>
<th>(\mu^{-1})</th>
<th>(l)</th>
<th>(\phi)</th>
<th>Wakob</th>
<th>Grübel</th>
<th>Tijms</th>
<th>Fischer</th>
<th>Hertel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>Exponential</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0.7</td>
<td>4</td>
<td>3.9999</td>
<td>4</td>
<td>2.7062</td>
<td>2.7062</td>
</tr>
<tr>
<td>Erlang</td>
<td>Erlang</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0.7</td>
<td>2.7062</td>
<td>2.7062</td>
<td>2.7062</td>
<td>2.4495</td>
<td></td>
</tr>
<tr>
<td>Erlang</td>
<td>Erlang</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0.7</td>
<td>1.7557</td>
<td>1.7500</td>
<td>1.7659</td>
<td>1.7247</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>Gamma</td>
<td>3</td>
<td>2.1</td>
<td>2</td>
<td>1.9</td>
<td>0.7</td>
<td>1.7464</td>
<td>1.7411</td>
<td></td>
<td>1.7077</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>Gamma</td>
<td>3</td>
<td>2.5</td>
<td>2</td>
<td>1.5</td>
<td>0.7</td>
<td>1.8203</td>
<td>1.8168</td>
<td></td>
<td>1.7302</td>
<td></td>
</tr>
<tr>
<td>Erlang</td>
<td>Weibull</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0.7</td>
<td>1.7557</td>
<td>1.7454</td>
<td>1.7659</td>
<td>1.7247</td>
<td></td>
</tr>
<tr>
<td>Erlang</td>
<td>Lognormal</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0.7</td>
<td>1.7557</td>
<td>1.7622</td>
<td>1.7659</td>
<td>1.7247</td>
<td></td>
</tr>
<tr>
<td>Erlang</td>
<td>Pareto</td>
<td>3</td>
<td>2</td>
<td>1.5</td>
<td>3</td>
<td>0.5</td>
<td>0.4775</td>
<td>0.4874</td>
<td>0.4858</td>
<td>0.5152</td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>Erlang</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0.7</td>
<td>1.7557</td>
<td>1.8055</td>
<td></td>
<td>1.7247</td>
<td></td>
</tr>
<tr>
<td>Lognormal</td>
<td>Erlang</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0.7</td>
<td>1.7557</td>
<td>1.6096</td>
<td></td>
<td>1.7247</td>
<td></td>
</tr>
<tr>
<td>Pareto</td>
<td>Erlang</td>
<td>1.5</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0.7</td>
<td>0.6741</td>
<td>0.4624</td>
<td></td>
<td>0.6754</td>
<td></td>
</tr>
</tbody>
</table>

Table A.1: Waiting time approximations for several different set-ups

Observe that Wakob's waiting time approximations are very accurate, even if the coefficients of variation \((k\) and \(l)\) are not integer values (see row 1–5), i.e. if Gamma distributions apply rather than Erlang ones which Wakob assumes as input distributions. Wakob outperforms the method by Fischer and Hertel as well as Tijms in almost all tested situations.

Note that Wakob only takes account of the first and second moment of both the interarrival time and service time distribution. The same is true for the approach by Fischer and Hertel. The accuracy of both approximation methods thus depend on how well the shapes of the entire distributions are covered by these two moments. It is known that the waiting times are not very sensitive for the service process as long as its first two moments are well determined. This is confirmed by the small differences of Grübel's approximations for different types
of service time distributions (see row 3, 6 and 7 in Table A.1). In these cases, Wakobs produces errors of approximately 1% which is still acceptable.

The results in rows 9–11 of the ninth column (Grübel) show significant differences. This indicates that a similar rule of thumb does not apply to the interarrival times. That is, the first two moments of the interarrival time distribution are generally insufficient to obtain accurate waiting time approximations. In these cases, Wakob produces much larger errors. Still, this does not indicate that Wakob's approximation may be inappropriate for typical non-Erlang (or non-Gamma) interarrival processes. Instead, one must then conclude that different regression functions have to be defined and subsequently calibrated in order to improve the results (see §A.4).
TRAIL Studies in Transportation Science

A series of The Netherlands TRAIL Research School for fundamental and theoretical studies on transport, infrastructure and logistics.


Forthcoming:

Sales and distribution:
Delft University Press
P.O. Box 98
2600 MG Delft
Telephone: +31 (0)15 278 32 54
Telefax: +31 (0)15 278 16 61