Propagation of train delays in stations

J. Yuan, R.M.P. Goverde & I.A. Hansen

Faculty of Civil Engineering and Geosciences,
Delft University of Technology, The Netherlands

Abstract

Train delays are daily practice. Moreover, propagation of delays is a main source for delays experienced by the Dutch Railways. A statistical analysis shows that the mean train delay increases 15 to 45 seconds for most train lines at The Hague HS station, the Netherlands. Late arrival delays and corresponding dwell delays can generally be fitted to an exponential distribution and a normal distribution, respectively. A stochastic model is presented that forecasts the distribution of departure delays of late trains based on the distributions of late arrival delays and dwell delays. The model incorporates some main constraints of train operations. The model is solved by the Monte Carlo sampling approach. Model validation shows that the model predicts the propagation of train delays reasonably well in railway stations.

1 Introduction

Punctuality of train services has become a major issue during the last decade due to increased mobility of people, capacity of rail systems and competition with other transport modes [1]. Therefore, train punctuality analysis and improvement is an important issue in the Netherlands.

In spite of the application of buffer times, train delays are inevitable due to the complex stochastic interactions among people, vehicle, infrastructure and weather condition [2]. Moreover, late arriving trains often suffer additional delays in railway stations due to excess of scheduled dwell times. Sources of the additional delays are variation of alighting and boarding times, transfers between connected trains, and late setting of outbound train routes due to conflicting train movements. Furthermore, limited discipline of train drivers, conductors and control centre personnel may also contribute to additional delays.
The extent of delay propagation in a station indicates whether the station capacity is sufficient to operate a given timetable at a predefined level of punctuality. Much achievement has already been obtained in delay propagation analysis on railway open tracks [3,4]. On the other hand, only limited literature is available on the analysis of capacity utilisation and delay propagation in complex station areas [5,6] and here the variation of train running times and dwell times are not considered according to real operation data. Therefore, it is necessary to analyse and model train delay propagation at stations on the basis of empirical train detection data by a stochastic approach.

To this end, train traffic has been statistically analysed in Eindhoven, the Netherlands [2]. The research results indicate that the mean dwell time of late trains is generally 40 seconds longer than the scheduled dwell time. To confirm whether these findings are valid in general, statistical analysis of train delays has also been performed at the station The Hague Holland Spoor (The Hague HS), the Netherlands, based on train detection data recorded during September 1999.

This paper concentrates on the case The Hague HS. It starts with a brief introduction to this station and the collection of train detection data. Then some main research results obtained from the statistical analysis of train delays in the station are presented. Furthermore, a stochastic approach to model delay propagation in railway stations is suggested and validated. Finally several conclusions are drawn.

2 The Hague HS station and train detection data

The Hague HS is one of the two main railway stations in the city of The Hague, the Netherlands. There are two main platforms and four main tracks in three directions (Amsterdam/Leiden, Rotterdam and The Hague CS, see Figure 1), at which passengers board and alight the trains.

![Figure 1: Track layout of The Hague HS railway station](image-url)
According to the scheduled timetable (Table 1), each hour 24 passenger trains arrive and depart in The Hague HS. These trains correspond to 9 different lines in both southbound and northbound directions, including 1 high speed train (HST) line, 1 international (INT) line, 4 intercity (IC) lines, 1 interregional (IR) line and 2 local (AR) lines.

Table 1: The timetable for trains passing through The Hague HS (1999/2000)

<table>
<thead>
<tr>
<th>Type</th>
<th>Line</th>
<th>Origin-Destination</th>
<th>Southbound</th>
<th>Northbound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Train Arr.</td>
<td>Arr. Dwell</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>/Hr time</td>
<td>Track nr.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[Min]</td>
<td>[Min]</td>
</tr>
<tr>
<td>AR</td>
<td>5000</td>
<td>Leiden-Dordrecht</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>AR</td>
<td>5100</td>
<td>The Hague CS-Roosendaal</td>
<td>2</td>
<td>00</td>
</tr>
<tr>
<td>IR</td>
<td>2200</td>
<td>Amsterdam-Breda</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>IC</td>
<td>1900</td>
<td>The Hague CS-Venlo</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>IC</td>
<td>2100</td>
<td>Amsterdam-Vlissingen</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>IC</td>
<td>2400</td>
<td>Amsterdam-Dordrecht</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>IC</td>
<td>2500</td>
<td>The Hague CS-Heerlen</td>
<td>1</td>
<td>54</td>
</tr>
<tr>
<td>INT</td>
<td>600</td>
<td>Amsterdam-Brussels</td>
<td>1</td>
<td>04</td>
</tr>
<tr>
<td>HST</td>
<td>9300</td>
<td>Amsterdam-Paris</td>
<td>1</td>
<td>34</td>
</tr>
</tbody>
</table>

During September 1999, the Dutch Railways have logged train operations data in the control area The Hague by a TNV-system (TNV stands for the Dutch abbreviation of “train number following”). The software package TNV-Prepare [7] has been used to convert the recorded logfiles into data suitable for analysis. In addition, since the precise arrival and departure times at the platform are not measured directly by the TNV-system, the program TNV-Filter [8] has been adopted to obtain the estimated arrival delays, departure delays and dwell times at the station.

In total, nearly 450 trains per day passed through The Hague HS station. The amount of data used for the statistical analysis corresponds to around 10,000 trains.

3 Statistical analysis of train delays

Train lines have distinct routes through The Hague HS station and the corresponding trains generally have different dynamic characteristics. These factors eventually lead to a different mean, standard deviation, and probability distribution of the train delays. Hence, a statistical analysis has been performed of train delays at The Hague HS station, per individual train line in both southbound and northbound directions.

The punctuality of a train line is expressed as the percentage of trains arriving or departing at a platform no later than a certain time in minutes. Delays shorter than 3 minutes are usually not considered as delays by the Dutch Railways. Figure 2 therefore shows the percentages of both arrival delays and departure delays longer than 3 minutes.
In most cases, the train delays longer than 3 minutes increase by 1% to 5% between arrival and departure at The Hague HS station, except for an international and two local southbound train lines. These train lines have longer scheduled dwell times and no conflicting outbound routes, by which the arrival delays are compensated during the stop.

The above analysis shows a delay increase for most train lines at The Hague HS station. To quantify the growth of the delays, the observed dwell times in case of late arrivals have been analysed for each train line per direction. In this paper, the term late arrival is used to distinguish from early arrivals. Figure 3 displays the mean dwell time of late trains and the associated scheduled dwell time.

In general, the mean dwell time in case of late arrivals is 15 to 45 seconds longer than scheduled, except for an international train line, two local southbound train lines, and the northbound high speed train line. Hence, most train lines obtain 15 to 45 seconds additional delay on average at The Hague HS.

An investigation at several Dutch railway stations [9] shows that about 20% of observed dwell times is unused during the dwelling process at stations. So, the additional delays at The Hague HS station might be caused by the limited discipline of train drivers and conductors or late setting of outbound train routes due to conflicting train movements.
In this section, the probability distributions of late arrival delays and dwell delays at The Hague HS station are studied. The identified distributions will be used in the train delay propagation modelling. It should be mentioned that during the whole measurement period, 6 southbound train lines had an alternative route from Sept. 13 to Sept. 18, 1999. So, for these trains two separate periods have been considered. In total, 24 alternatives of train lines and routes are taken into account during the distribution fitting for train delays.

The computation of statistics and the estimation of distribution parameters can be influenced by outliers, i.e., data points that deviate from the bulk of the data [2]. Boxplots and scatterplots [10] have been used to identify the outliers of a dataset. Discarding the outliers, the distributions of train delays are fitted by the Kolmogorov-Smirnov (KS) test [10]. The software package S-Plus 2000 [101 has been used during these detailed analyses.

There is statistical evidence that late arrival delays and corresponding dwell times can be modelled by an exponential and normal distribution respectively in Eindhoven station [2]. So it has been checked whether these distributions also apply in The Hague HS.

The KS tests indicate that late arrival delays are fit well to an exponential distribution in 20 alternatives (83% of all cases) at The Hague HS. The exponential distribution has only been rejected for the late arrival delays of southbound trains originating from The Hague CS and a single northbound train line. In those special cases, the share of late arrivals is nearly 100% or only 24%. It seems that late arrival delays generally follow an exponential distribution.
unless a structural arrival delays exists. The dwell times of late arrivals can be fit to a normal distribution in 21 alternatives (88% of all cases) at The Hague HS. For 18 of all 24 considered alternatives, the late arrival delays and corresponding dwell delays are fitted to an exponential distribution and a normal distribution, respectively.

To model delay propagation at a station, we define dwell delay as the difference between the observed dwell time and the scheduled dwell time in the sequel. Although the mean of observed dwell times is generally longer than the scheduled dwell time, some delayed trains have shorter dwell times than scheduled to compensate for the arrival delays. Thus the dwell delays here include not only positive, but also negative values. Obviously, dwell delays can also be fitted to a normal distribution in most cases, where the mean is the difference between the mean of observed dwell times and scheduled dwell time.

5 Modeling delay propagation at stations

In this section, a stochastic model of delay propagation at a station is presented and validated. Based on the distributions of late arrival delays and dwell delays, the model predicts not only the mean, but also the distribution of departure delays in case of late arrivals, which is more important in train planning and scheduling.

In the proposed model, two main train operation constraints are considered. First, a train is not allowed to depart earlier than the scheduled departure time at a station, so the departure delay is nonnegative. This means that the sum of late arrival delay and dwell delay should be nonnegative. Second, the dwell time of a train is longer than a certain minimum value, for instance 30 seconds. Note that some other train operation characteristics are not considered in this model. For example, a late train might decrease its dwell time to compensate for the arrival delay. On the other hand, some late trains may increase dwell times in case of late setting of outbound train routes due to conflicting train movements. The statistical analysis of train traffic at The Hague HS shows that the correlation coefficient between late arrival delays and dwell delays is between -0.15 and +0.15 for most train lines per direction, except for an international train line, the two local southbound train lines, and the high speed train line in both directions. This behaviour is neglected in the model, which is partly justified by the small correlation coefficients.

Predicted delay propagation at a station will indicate whether the station capacity is sufficient to operate a given timetable at a predefined level of punctuality. In addition, accurate prediction of departure delay can support online train dispatching.

5.1 Model specification

Late arrival delays, dwell delays and the corresponding departure delays are indicated by $X$, $Y$ and $Z = X + Y$. According to basic probability theory


\[ E(Z) = E(X) + E(Y), \quad (1) \]

and

\[ Var(Z) = Var(X) + Var(Y) + 2Cov(X, Y). \quad (2) \]

Here, \( E(\cdot) \) and \( Var(\cdot) \) denote the mean and variance of a random variable, and \( Cov(\cdot, \cdot) \) is the covariance of two random variables.

Given the distributions of late arrival delays and dwell delays, the corresponding mean and variance are known, by which the mean departure delay is obtained from eqn (1). Taking into account the constraint that the sum of late arrival delay and dwell delay should be nonnegative, implies that \( Cov(X, Y) \) is not zero. Moreover, \( Cov(X, Y) \) is unknown, by which the variance of the departure delays will be calculated from the predicted distribution.

The distribution function of departure delays in case of late arrivals can be described as,

\[ F(z) = \int_{0}^{z} f(u)du. \quad (3) \]

Here, \( F(\cdot) \) and \( f(\cdot) \) are the cumulative distribution function and the probability density function of departure delays in case of late arrivals, respectively.

The integral in eqn (3) is actually a double definite integral as \( Z = X + Y \). However, eqn (3) can not be simplified as a multiplication of two single integrals, because of the constraints between \( X \) and \( Y \). In this case, the Monte Carlo sampling approach \[11\] is very efficient to solve the problem.

The principle of the Monte Carlo sampling approach is to approximate a probability with the corresponding frequency. Given the distributions of late arrival delays \( X \) and dwell delays \( Y \), the corresponding random numbers can be generated. Assume that we have \( N \) groups of generated random numbers, of which \( M \) groups satisfy the condition \( Z = X + Y \leq z \), then

\[ F(z) \approx \frac{M}{N}. \quad (4) \]

The random numbers can not be produced independently in the model as \( Y \geq -X \) has to be satisfied. Furthermore, if the scheduled dwell time is \( c \) and the minimum observed dwell time is \( d \), then also the constraint \( Y \geq d - c \) should be satisfied. Therefore, the generation of random numbers is performed as follows.

1) Generate a random number \( X_i \) from the distribution of late arrival delays;
2) Generate a random number \( Y_i \) from the distribution of dwell delays;
3) If \( Y_i < \max(-X_i, d-c) \) then repeat step 2), until \( Y_i \geq \max(-X_i, d-c) \);
4) Repeat step 1) to step 3) until \( N \) groups of random number are obtained for the late arrival delays \( X \) and dwell delays \( Y \).
The probability of departure delays less than \( z \) can be predicted by eqn (4), and hence the probability of departure delays larger than \( z \) is approximated by,

\[
1 - F(z) \approx 1 - \frac{M}{N}.
\]

In addition, the variance of departure delays in case of late arrivals are easily calculated based on the generated \( N \) random samples \( Z_i = X_i + Y_i \),

\[
Var(Z) \approx \frac{N \sum_{i=1}^{N} Z_i^2 - (\sum_{i=1}^{N} Z_i)^2}{N(N-1)}.
\]

### 5.2 Model validation

In this sub-section, the presented model is validated with the train detection data recorded at The Hague HS station in September 1999. The model is applied and verified respectively on the 18 alternatives mentioned in Section 4.

![Figure 4: Probability density lines of departure delays in case of late trains (IC2100S on its normal route, The Hague HS, Sept. 1999)](image)

Figure 4 shows the modelled probability density and the observed density estimate from the train detection data for one of the 18 alternatives. The model
predicts the observations quite well. As for the probability that departure delays are longer than 3 minutes, the absolute difference between modelled results and the results obtained directly from the train detection data is less than 3% for most train lines, except for a high speed train line in both directions. The modelled standard deviation is within 20% of the one obtained from the train detection data in most cases, except for an international train line and a southbound high speed train line. In general, the presented model predicts reasonably well the distribution of departure delays in case of late arrivals.

6 Conclusions

The detailed statistical analysis indicates that the train delays of most train lines increase 15 to 45 seconds on average at The Hague HS station, although some trains can compensate for the arrival delays. Late arrival delays and corresponding dwell delays can generally be fitted to an exponential distribution and a normal distribution, respectively. These results confirm the conclusions of a previous analysis in Eindhoven. In case of late arrivals, the presented stochastic delay propagation model predicts reasonably well the distribution of departure delays.

Future research will focus on analyzing the dependence of dwell delays on late arrival delays and scheduled dwell times. The interrelation among different train series will also be studied. Finally, optimal capacity utilisation at a station is expected to be obtained with a predefined level of punctuality.

Acknowledgement

This publication is a result of the research programme Seamless Multimodal Mobility, carried out within the TRAIL Research School for Transport, Infrastructure and Logistics, and financed by Delft University of Technology.

References


