Synchronization Control of Scheduled Train Services to Minimize Passenger Waiting Times

TRAIL Research School, Delft, December 1998

Author:
Drs. Rob M.P. Goverde
(Transportation Planning and Traffic Engineering Section,
Faculty of Civil Engineering and Geo Sciences, Delft University of Technology)

Thesis supervisor:
Prof.Dr-Ing. I.A. Hansen
(Transportation Planning and Traffic Engineering Section,
Faculty of Civil Engineering and Geo Sciences, Delft University of Technology)

Contents

Abstract

During operation a transportation service may wait on delayed feeder services to secure scheduled transfers. For low-frequent connecting services this has a major positive impact on the transfer waiting times. However, the resulting synchronization control time of the connecting service also affects the waiting times of originating and through passengers on the current transfer station, as well as the waiting times on subsequent stations resulting from delay propagation in the service network.

A systematic mathematical model has been developed to compute all affected waiting times of initial train departure delays. The delay propagation is modelled as a discrete event dynamic system, and in particular as a max-plus linear system. It explores the effect of buffer times to compensate for arrival delays.

The model can be utilized to evaluate the optimal synchronization control policy with respect to given arrival delays: secure or dissolve a transfer. The objective is the minimization of the total relative (generalized) passenger waiting time. Another application is the analysis of existing buffer times on passenger waiting times in service network timetables.

1 Introduction

Public transportation networks often operate according to a predetermined timetable. However, during operation running times usually have random variations that may result in arrival delays at transfer stations. To prevent missed transfers, operators may hold connecting vehicles to secure connections. The resulting departure delay is called synchronization control time. This paper investigates the question of how large a synchronization control time may get or whether a connection should be cancelled, with respect to the waiting times of all passengers involved. The focus is on scheduled train services as they represent a particularly interesting general case.

In the Netherlands, the yearly established WRT (Wachttijden voor ReizigersTreinen) (NS Reizigers, 1995) corresponding to the annual timetable gives guidelines to the train process operators for the maximum admissible synchronization control times, the so-called synchronization control margins. It contains some general guidelines, e.g., an intercity (IC) train waits no longer than 2 minutes on other intercity trains and/or (international) interregional (IR) trains. Additionally, it has a large list of exceptions for particular connections. The numbers are all based on rules of thumb and experience. In Germany a similar list of guidelines exists, the WZV (Wartezeitvorschrift für den Personenverkehr) (Kannengießer & Wiche, 1987).

A static list of guidelines like the WRT does not satisfy well in all situations. For instance, if two feeder trains have a moderate arrival delay then a larger synchronization control time is beneficial to relative many transferring passengers. For this reason, the static list is merely a guideline. In practice, the process operators base their decisions also on the actual situation and their experience.

A more suitable approach to assist the process operators is a decision support system (DSS) that utilizes online information on the actual state of the train operations to determine whether a connection should be secured or cancelled. Such a tool can quickly evaluate various scenarios on for instance passenger waiting times and propagation of delays. The consequences of a decision can be very complex and a DSS can help the process operator to derive an optimal decision in any particular situation.

The mathematical model described in this paper can also be utilized to offline compute static guidelines for various situations, like for a range of combinations of arrival delays of all feeder train services. Another example is that passenger flows may differ during various periods of a day, which might effect optimal synchronization control times. The optimal synchronization control times can then be computed numerically for different scenarios of passenger flows. In this way, a pseudo-dynamic and improved "WRT" is obtained that is not just the result of experience.

Most studies about transfer optimization are concerned with the scheduling process, see e.g., Bookbinder & Désilets (1992) and Knoppers & Muller

(1995). The latter derive optimal transfer buffer times for a single connection, with respect to stochastic arrival times of a feeder service and also incorporate a (given) synchronization control margin.

Dispatching solutions to schedule deviations in transportation networks have been discussed rarely. Most papers in this field consider schedule perturbations of a single line only and do not consider the effect on other lines. These studies are mainly concerned with the control of punctuality (on-time arrivals/ departures) or regularity (maintaining the scheduled headway) (Adamski & Turnau, 1998; Ruyter *et al.*, 1990). Some studies consider swapping the order of two scheduled services with respect to minimum departure headways or knock-on delays on the subsequent single-track sections (Carey & Kwiecinski, 1994; Jovanovic & Harker, 1991; Ruyter *et al.*, 1990). These studies are again not concerned with the effects on a network level. Yan & Yang (1996) consider the flight schedule perturbation problem due to breakdown of an aircraft. The problem is constructed as a network flow problem and considers all affected connections. However, the problem is completely different from the one dealt with here. We are not concerned with the consequences of the breakdown of a train (or vehicle) but with random arrival times of several services. Betgé-Brezetz *et al.* (1998) and Stolk (1998) are recent studies devoted to conflict solving.

The current paper presents an approach to evaluate the effect of synchronization control times on all passenger (generalized) waiting times. The optimal dispatching strategy then follows by comparison of the various possible scenarios. The mathematical model is based on a discrete event description of the waiting times. It includes a delay propagation forecasting model formulated as a max-plus linear system.

Section 2 describes the concepts of connection buffer times and synchronization control times, and formulates the general synchronization control problem. Section 3 then considers the various passenger waiting times resulting from the arrival delays at the current transfer station, and at the subsequent stations resulting from delay propagation. The delay propagation model is the subject of Section 4. Section 5 gives an illustration of deriving optimal synchronization control times for an example train service system. Finally, Section 6 gives conclusions and some remarks about possible extensions of the model.

2 Synchronization Control

2.1 Introduction

Small arrival delays of feeder trains can usually be compensated by connection buffer times that are incorporated in the timetable. For larger arrival delays a connecting train may wait to secure the connection. The amount of necessary synchronization control time is limited by the negative impact on the waiting times of other passengers and the perturbation of scheduled train movements. If a feeder train is too much delayed then the connection is cancelled and the connecting train may depart as scheduled. The transferring passengers from this delayed feeder train then miss their transfer and have to wait on the next train. If the frequency of the connecting line is low the resulting transfer waiting times are relatively large. For a more detailed description, see Goverde (1997).

We thus have the following operational control policies at our disposal: secure or dissolve a connection. Or to be more specific: how can we decide that a synchronization control time is no longer favourable and the connection has to be cancelled? This question is answered in Section 2.3 after the introduction of some concepts and assumptions in Section 2.2.

2.2 Connection Buffer Times and Synchronization Control Time

The train service network is modelled such that each train service between adjacent transfer stations is denoted separately. Thus, an individual train run of a train series is modelled as a different service for each trip between consecutive transfer stations on its route. A connection ij at a transfer station between an arriving train service i and a departing service j then either corresponds to a stop or a transfer.

The scheduled transfer time (or changeover time) between a feeder train service i and a connecting train service j consists of two components: a minimum transfer time necessary for changing trains and a transfer buffer time (or recovery time) that compensates for small arrival delays of the feeder train. This transfer buffer time thus reduces the probability of missing a connection but increases the scheduled transfer time. The transfer time is given as

$$t_{ij}^{\mathrm{t}} = t_{ij}^{\mathrm{t,min}} + r_{ij}^{\mathrm{t}} . \qquad (2.1)$$

The minimum transfer time consists of alighting time, walking time (including possible orientation), and boarding time. It depends on individual walking speed and acquaintance with the station, the relative position of the arrival and departure platform (cross-platform, two platforms apart, etc.), the geography of the station (platform lengths, distances between platforms, widths of corridors and door-ways, presence of escalators, etc.), and the pedestrian flows and densities in the station. It is assumed in this paper that the minimum transfer time is a constant, determined in such a way that all transferring passengers are able to successfully transfer with high probability. It then follows that also the transfer buffer time is a constant since the scheduled transfer time is fixed.

The stopping time of a train service i at a transfer station consists of two components: a minimum stopping time necessary for passengers to alight and board the train and a stopping buffer time induced by a scheduled connection. The departure time of the next service is postponed to give transferring passengers of a feeder train the opportunity to change trains. The stopping time is thus given as

$$t_{ij}^{\mathrm{s}} = t_{ij}^{\mathrm{s,min}} + r_{ij}^{\mathrm{s}}, \tag{2.2}$$

where j is the subsequent service of the stopping train from service i. The minimum stopping time is again assumed to be constant and thus also is the stopping buffer time.

In general, a connection time is thus composed of a minimum connection time and a connection buffer time, which correspond to either a stop or a transfer. If it is clear from context whether we deal with a stop or a transfer the superscript identification is also dropped and we speak, e.g., of the connection buffer time r_{ij}. The connection buffer times can compensate small arrival delays of arriving trains. However, for larger arrival delays the scheduled departure times may be endangered.

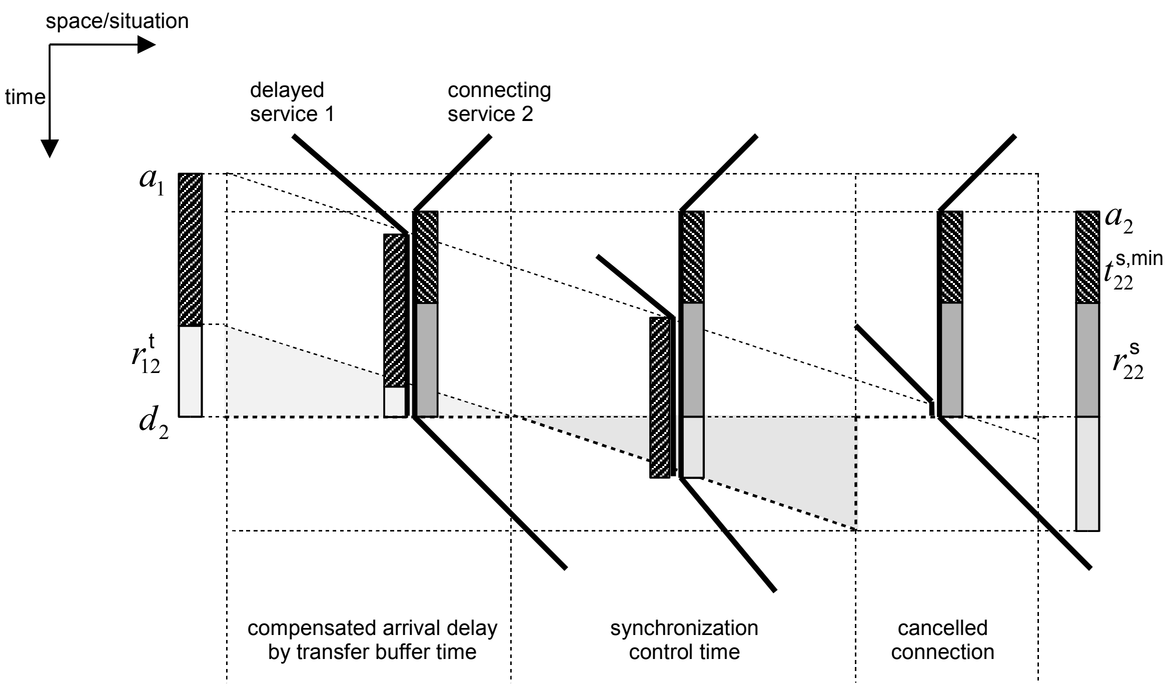


Figure 2.1 Synchronization control

Therefore, during operation a connecting train may wait for delayed transferring passengers if a feeder train arrives behind schedule. The resulting synchronization control time is the additional time above the scheduled departure time that the connecting train waits at the platform to secure the connection for the delayed transferring passengers. Large synchronization control times result in high waiting times for other passengers and may endanger connections at subsequent stations. To avoid this, a so-called synchronization control margin defines a maximum admissible synchronization control time to each (connecting) train. If a feeder train is that much delayed that synchronization control would result in exceeding the synchronization control margin then the connection is cancelled and the train may leave as scheduled, see Figure 2.1. The computation of optimal synchronization control times (online) or optimal synchronization control margins (offline) is the subject of this paper.

2.3 Optimal Synchronization Control

A stop relates to a physical connection and can not be controlled in the sense that train j can not depart before train i has arrived. A transfer, however, relates to a connection between two different trains and can be cancelled, i.e., the connecting train may depart before the feeder train has arrived. The cancelled connection then results in missed transfers. This might be useful when a feeder train is highly delayed. Although the transferring passengers miss the connection, other passengers benefit from the on-time departure. On the other hand, it might be beneficial that a train waits for a (or more) slightly delayed feeder train(s). This synchronization control time results in a secured connection by which transferring passengers from the delayed feeder train do not miss the transfer. Note that missed transfers may result in large waiting times for the transferring passengers depending on the frequency of the connecting train service.

The aim of synchronization control is to prevent large waiting times for transferring passengers who tend to miss a connection at the cost of small waiting times for other passengers. So a sensible objective is to minimize the overall inconvenience for all passengers involved, including the transferring passengers, through passengers, and originating passengers, as well as passengers on subsequent stations. An obvious measure for the inconvenience of an individual passenger is waiting time. The objective can then be defined as the minimization of the total passenger waiting times.

From modal split studies it is known that a passenger judges various travel time components differently, see, e.g., Van Goeverden *et al*. (1990) and Van der Waard (1989). Therefore, it makes sense to define the passenger inconvenience as its generalized waiting time. The simplest of those models consist of a linear relationship where the travel time component is multiplied by a weighting factor to take account for its relative importance. The above mentioned studies derive that a travel component outside a vehicle weighs 3 times higher than inside a vehicle. Adopting this result gives that the transfer waiting times (outside the train) weigh three times higher as the through waiting times of the through passengers inside the connecting train. For originating passengers and transferring passengers from other feeder trains the weights are defined according to their status of being inside or outside the train. The objective then becomes to minimize the total generalized waiting times. The determination of the various waiting times is the subject of

Section 3 where a discrete event formulation describes all waiting times. Here we continue with the formulation of the optimization problem.

Consider a transfer station where several train series stop. Each stopping train may be both a feeder train and a connecting train to other trains that have arrived earlier or have to arrive soon. Assume train service j is a connecting service to one or more feeder services. Then each feeder train i may cause a synchronization control time if its arrival delay p_i exceeds the connection buffer time. The candidate synchronization control time for service j with respect to connection ij then is

$$s_{ij} = p_i - r_{ij}^{\mathrm{t}} . \tag{2.3}$$

If the maximum of those candidate synchronization control times is applied then all transfer connections are secured. Otherwise one or more connections are cancelled. In this way, the candidate synchronization control times give rise to various control scenarios. Computing the objective function for each one of them shows which one is the optimal control policy. If more trains at a transfer station may be subject to synchronization control then all possible combinations of proposed departure times can be evaluated. Note that the simultaneous application of synchronization control to various trains effects the waiting times.

Resuming we obtain the following optimization problem: determine the synchronization control times for the connecting trains with respect to all transfer connections that minimizes the total (generalized) waiting time.

Additionally, some side constraints may be taken into account. These may include

- a maximum tolerable departure delay;
- a maximum number of periods before a resulting departure delay has disappeared, i.e., the number of subsequent stations that are also affected by the current synchronization control time;
- stability of train circulations, i.e, the guaranteed on-time departure of trains from their starting terminal station.

The last two criteria are determined implicitly by the delay propagation model of Section 4 which is applied to determine waiting times at subsequent stations. These constraints can thus only be checked on feasibility after computing the objective function.

3 Passenger Waiting Times

3.1 Introduction

The cost of a particular synchronization control action is measured by the total (generalized) waiting time of all passengers involved. Two main classes of (passenger) waiting times can be distinguished:

- *primary waiting time*: the sum of all involved waiting times at the transfer station, and
- *secondary waiting time*: the sum of all waiting times at subsequent (transfer) stations.

Computation of the secondary waiting time requires a delay propagation forecasting model which is considered in Section 4.

The primary and secondary waiting time can again be composed according to the four passenger classes:

- *originating passengers*: passengers whose travel starts at the transfer station;
- *through passengers*: passengers inside a train stopping at the transfer station;
- *transferring passengers*: passengers who transfer from a feeder train to a connecting train;
- *terminating passengers*: passengers who end their travel at the transfer station.

From these classes the terminating passengers will not be considered separately. At the transfer station where the synchronization control takes place they are of no concern, and at subsequent transfer stations they have already been incorporated in the classes of through or transferring passengers at the preceding station(s).

We assume that the passenger flows, i.e., the number of concerned passengers, are known for each passenger class. These can either be a real-time forecast of the actual passenger flows or an indication of the relative passenger flows. In the latter case, the total waiting time is actually a measure of the relative waiting time.

As mentioned in Section 2 we are also concerned with the *generalized waiting time*. Basically, the generalized primary and secondary waiting times are computed by weighing waiting times outside a train three times higher than in the train, see Table 1 (Van der Waard, 1989; Van Goeverden *et al.*, 1990).

Table 1 Modal split weights

waiting time	weight
in train	1
on platform	3

The next two sections consider the various components of the primary and secondary waiting time, respectively, as well as their generalized variants.

3.2 Primary Waiting Time

As mentioned above, the primary (generalized) waiting time is composed of the waiting times corresponding to three passenger classes that are considered successively below.

Originating Passengers

The waiting time for a passenger who starts his/her travel in a train of service j is exactly the (possible) synchronization control time of train service j. Hence, the originating waiting time for sevice j is

$$w_j^{\mathrm{o}} = s_j . \tag{3.1}$$

The generalized waiting time for originating passengers also equals (3.1), which can be understood as follows. Possible waiting time at the departure platform due to an arrival delay of train service j is not a consequence of the (current) synchronization control. Therefore, it is assumed that originating passengers board the train on time and wait inside the train for departure. The modal split weight of the waiting time then equals 1. Note that for secondary waiting times the modal split weight would be 3 since then the waiting time at the platform is a result of the current synchronization control.

Through Passengers

The waiting time for a through passenger equals the excess stopping time of the train at the transfer station above the minimum stopping time. In our model we differentiate between arriving and departing trains. Therefore, consider an arriving train service i that stops at the transfer station and resumes as departing train j, i.e., ij is a physical connection. Then the through waiting time is composed of the sum of the stopping buffer time and synchronization control margin corrected by the arrival delay,

$$w_{ij}^{\mathrm{s}} = r_{ij}^{\mathrm{s}} + s_j - p_i . \tag{3.2}$$

The through waiting time is identified with a supercript ‘s’ as it corresponds to a stop.

The modal split weight for a through passenger clearly equals 1 since the waiting time in the train is part of the trip of the through passenger. It follows that the generalized through waiting time also equals (3.2).

Transferring Passengers

The waiting time of a transferring passenger depends on whether or not the passenger misses the connection. If the connection is cancelled then it is assumed that the passengers take the next train of the same connecting service j. Moreover, it is assumed that this next train will depart on time. Note that there is no information about future dispatching decisions. The total transfer waiting time then is

$$w_{ij}^{\mathrm{t}} = \begin{cases} r_{ij}^{\mathrm{t}} + s_j - p_i & \text{if } p_i \le r_{ij}^{\mathrm{t}} + s_j \\ r_{ij}^{\mathrm{t}} + h_j - p_i & \text{if } p_i > r_{ij}^{\mathrm{t}} + s_j. \end{cases} \tag{3.3}$$

Here, h_j is the interdeparture time of the connecting train service j. Note that if the synchronization control time is due to the current feeder train service i then the transfer waiting time (3.3) is zero.

The modal split weight for waiting time inside a train is 1 and outside the train is 3. If a connection is secured then the modal split weight equals 1 since a possible waiting time is assumed inside the connecting train for the same reasoning as in the case of originating passengers. For cancelled connections, however, the transferring passengers have to wait on the next train implying a modal split weight of 3. The generalized transfer waiting time then is

$$g_{ij}^{\mathrm{t}} = \begin{cases} r_{ij}^{\mathrm{t}} + s_j - p_i & \text{if } p_i \le r_{ij}^{\mathrm{t}} + s_j \\ 3(r_{ij}^{\mathrm{t}} + h_j - p_i) & \text{if } p_i > r_{ij}^{\mathrm{t}} + s_j. \end{cases} \tag{3.4}$$

Note that with respect to the primary waiting times, only missed transfers result in a high penalty due to the modal split weights.

All Passengers

The total primary waiting time is the sum over all waiting times at the transfer station. Let n_j be the amount of originating passengers or *originating flow* with respect to service j, and n_{ij} be the *through flow* or *transfer flow* depending on whether the connection ij is a stop or a transfer. Moreover, let O_{p} be the set of departing train services at the transfer station, S_{p} be the set of all pairs of stop connections and T_{p} be the set of all pairs of transfer connections, i.e.,

O_{p} := { j | j is a departing train service from the transfer station},

S_{p} := { (i,j) | ij is a stop connection at the transfer station},

T_{p} := { (i,j) | ij is a transfer connection at the transfer station}.

Then the total (generalized) primary waiting time is given as

$$w^{\mathrm{prim}} = \sum_{j \in O_{\mathrm{p}}} n_j w_j^{\mathrm{o}} + \sum_{(i,j) \in S_{\mathrm{p}}} n_{ij} w_{ij}^{\mathrm{s}} + \sum_{(i,j) \in T_{\mathrm{p}}} n_{ij} w_{ij}^{\mathrm{t}} . \tag{3.5}$$

Here the individual waiting times for the various passenger classes are either the normal waiting times defined by (3.1), (3.2) and (3.3) or the generalized waiting times defined by (3.1), (3.2) and (3.4).

3.3 Secondary Waiting Time

The computation of the secondary waiting times requires a delay propagation forecasting model. Section 4 describes such a model based on discrete event system modelling. This model computes the departure delays $y_i(k)$ of all kth departures (k = 1,2,…) of any train service i. The counter k is called the period. Note that a synchronization control time can effect a train service at a subsequent transfer station several times if the service lies on a circuit where the delay propagation reduces slowly. Therefore, the counter k denotes the successive periods to which the delay refers.

In this section we assume that future departure delays have been computed. The waiting times for the various passenger classes in future periods k can then be computed from the forecasted departure delays and the scheduled connection buffer times. These waiting times are again dealt with successively below.

Originating Passengers

The waiting time for the originating passengers at a particular period k is proportional to the departure delay, i.e.,

$$w_j^{\mathrm{o}}(k) = y_j(k)\,. \tag{3.6}$$

The generalized originating waiting time depends on whether the departure delay of train service j is the result of an arrival delay or a secured transfer connection. In the former case the passenger has to wait on the departure platform whereas in the latter case (part of the) waiting time may be inside the train.

Let i be the arriving train that proceeds as the departing train j, i.e., ij is a stop connection. If the arrival delay of train i does not exceed the stopping buffer time then the originating passengers can board the train and do not have to wait on the departure platform. Otherwise the platform waiting time equals the arrival delay reduced by the stopping buffer time. In our delay propagation model we assume deterministic train running times and hence the arrival delay equals the departure delay of the train at its preceding transfer station. The platform waiting time for originating passengers of train j in period k then is

$$\left(y_i(k-1) - r_{ij}\right)^+, \tag{3.7}$$

where $(x)^+ = \max(0,x)$. Here, we use the convention that the current period is k = 0. The departure delay then corresponds to a possible synchronization control time. The waiting time inside the train is the remaining time until departure, that is

$$y_j(k) - \left(y_i(k-1) - r_{ij}\right)^+. \tag{3.8}$$

Now, the generalized originating waiting time in period $k \geq 1$ is given as the sum of (3.7) multiplied by weighting factor 3 and (3.8) with unit weight, i.e.,

$$\begin{aligned} g_j^{\mathrm{o}}(k) &= 3\left(y_i(k-1) - r_{ij}\right)^+ + y_j(k) - \left(y_i(k-1) - r_{ij}\right)^+ \\ &= y_j(k) + 2\left(y_i(k-1) - r_{ij}\right)^+ . \end{aligned} \tag{3.9}$$

Note that if the departure delay is a result of the train's own arrival delay then the departure delay equals the reduced arrival delay, i.e., $y_i(k-1) - r_{ij} = y_j(k)$ and the generalized originating waiting time (3.9) is three times the departure delay. This corresponds to the fact that the originating passengers have to wait at the platform on the arrival of the delayed train that then departs immediately after boarding and alighting. On the other extreme, if the arrival delay is compensated by the stopping buffer time then the second term in (3.9) is zero and the generalized originating waiting time equals (one times) the departure delay. This corresponds to on-time boarding and subsequently waiting inside the train on its delayed departure.

Through Passengers

We assume that early arrivals do not occur or at least do not have to be taken into account. The through waiting time then equals the sum of the departure delay and the stopping buffer time corrected by the arrival delay. Let ij be a stop connection. Assuming deterministic train running times in the forecasting model, the arrival delay of train i equals its departure delay from its preceding transfer station. The through waiting time in a period $k \geq 1$ then is

$$w_{ij}^{\mathrm{s}}(k) = r_{ij} + y_j(k) - y_i(k-1) . \tag{3.10}$$

Note that this equation is nonnegative for a realistic delay propagation forecasting model and thus well-defined.

The through waiting time is by definition inside the train by which the modal split weight is one. Hence, the generalized through waiting time in period k also equals (3.10).

Transferring Passengers

Let now ij be a transfer connection. Recall the assumption that all connections are secured with respect to the initial departure delay caused by the current (candidate) synchronization control. Similar to the through passenger case, the transfer waiting time in a period $k \geq 1$ then is

$$w_{ij}^{\mathrm{t}}(k) = r_{ij} + y_j(k) - y_i(k-1) . \tag{3.11}$$

The generalized transfer waiting time depends on the arrival time of both the feeder train and the connecting train. If the connecting train has not yet arrived when the transferring passengers arrive at the departure platform, then some platform waiting time results.

Let h be the arriving train service that proceeds as train service j, and let i be the feeder train service that has a transfer connection to j. Then transferring passengers have to wait at the platform if

$$r_{ij} - y_i(k-1) > r_{hj} - y_h(k-1). \tag{3.12}$$

If this inequality is valid then the corresponding waiting time equals the difference between the left-hand side and the right-hand side of (3.12). Hence, in general the platform waiting time is

$$\left(r_{ij} - r_{hj} + y_h(k-1) - y_i(k-1)\right)^+ .$$

The remaining waiting time inside the train before departure is then given as

$$r_{ij} + y_j(k) - y_i(k-1) - \left(r_{ij} - r_{hj} + y_h(k-1) - y_i(k-1)\right)^+ . \tag{3.13}$$

The generalized transfer waiting time in a period $k \geq 1$ is now given as the weighted sum of (3.12) and (3.13), resulting in

$$\begin{aligned} g^{\mathrm{t}}_{hij}(k) &= 2\left(r_{ij} - r_{hj} + y_h(k-1) - y_i(k-1)\right)^+ + r_{ij} + y_j(k) - y_i(k-1) \\ &= \begin{cases} r_{ij} + y_j(k) - y_i(k-1) & \text{if } r_{ij} - y_i(k-1) \leq r_{hj} - y_h(k-1) \\ 3r_{ij} - 2r_{hj} + y_j(k) - 3y_i(k-1) + 2y_h(k-1) & \text{if } r_{ij} - y_i(k-1) > r_{hj} - y_h(k-1). \end{cases} \end{aligned} \tag{3.14}$$

If the departure delay of the connecting service is the result of its own arrival delay, then the departure delay equals the reduced arrival delay $y_j(k) = y_h(k-1) - r_{hj}$. Substituting this equation in the lower part of (3.14) results in thrice the (normal) transfer waiting time corresponding to the fact that the total transfer waiting time is on the departure platform.

Note that for the computation of the secondary generalized transfer waiting times a list of both stop connections and transfer connections is needed additionally to the forecasted departure delays.

All Passengers

The total secondary waiting time is now the sum over all waiting times in the service network in all periods. Let n be the number of train services in the service network, S_{s} be a list of all stop connections and T_{s} be a list of all transfer connections in the service network, i.e.,

S_{s} := { (i,j) | ij is a stop connection in the service network},

T_s := { (i,j) | ij is a transfer connection in the service network}.

Moreover, let K be the number of periods in which all departure delays have vanished, n_j be the originating passenger flow for train j, and n_{ij} be the through or transfer passenger flow, respectively. Here, the originating passenger flow may include the passengers that board the train at intermediate stops of service j. It is then assumed that the departure delay is not reduced before the next transfer station. Then the total secondary waiting time is

$$w^{\mathrm{sec}} = \sum_{k=1}^{K} \left[\sum_{j=1}^{n} n_j w_j^{\mathrm{o}}(k) + \sum_{(i,j)\in S_\mathrm{s}} n_{ij} w_{ij}^{\mathrm{s}}(k) + \sum_{(i,j)\in T_\mathrm{s}} n_{ij} w_{ij}^{\mathrm{t}}(k) \right]. \tag{3.15}$$

Here the individual waiting times for the various passenger classes are the waiting times defined by (3.6), (3.10) and (3.11). A similar equation holds for the secondary generalized waiting time, where the individual generalized waiting times are defined by (3.9), (3.10) and (3.14). If the (relative) passenger flows are known online or are estimated with respect to particular periods, then they are taken dynamically, i.e., $n_j \equiv n_j(k)$ and $n_{ij} \equiv n_{ij}(k)$ for all positive integers k.

3.4 Relative Waiting Time

The delay propagation in the service network depends on the particular values of the synchronization control times. For different candidate synchronization control times the effected connections through the service network differ. Therefore, a reference system is necessary for the comparison of various synchronization control policies.

A natural reference system is the secondary waiting time that arises with respect to a punctual operation, i.e., when all departure times are as scheduled. The reference secondary waiting time w_0^{sec} is computed with respect to all connections (over all periods) for which the feeder train has an arrival delay, since these connections give rise to a change in waiting times. For all other connections the waiting times are unaffected and the difference with the waiting times arising in punctual operation is zero. The waiting times for punctual operation are just the connection buffer times. Note that the waiting time for originating passengers is zero for on-time departures. The reference secondary waiting time thus is

$$w_0^{\mathrm{sec}} = \sum_{k=1}^{K} \left[\sum_{(i,j)\in S_\mathrm{s}} n_{ij} r_{ij} + \sum_{(i,j)\in T_\mathrm{s}} n_{ij} r_{ij} \right].$$

The secondary generalized waiting time follows from the derived equations in Section 3.3. The generalized originating waiting time is zero and the generalized through waiting time equals the stopping buffer time. The generalized transfer waiting time depends on whether or not the transferring passengers can board the connecting train immediately after arrival at the platform,

$$g_{hij}^{\mathrm{t}} = \begin{cases} r_{ij} & \text{if } r_{ij} \le r_{hj} \\ 3r_{ij} - 2r_{hj} & \text{if } r_{ij} > r_{hj}. \end{cases}$$

The reference secondary generalized waiting time therefore is

$$g_0^{\mathrm{sec}} = \sum_{k=1}^{K} \left\lfloor \sum_{(i,j)\in S_{\mathrm{s}}} n_{ij} r_{ij} + \sum_{(i,j)\in T_{\mathrm{s}}} n_{ij} g_{hij}^{\mathrm{t}} \right\rfloor.$$

Similarly, also the primary waiting time is taken relative to the primary waiting time that arises in a punctual operation of the service system. Note that for punctual operation there is no originating waiting time. The reference primary waiting time is the sum of the weighted (stopping and transfer) connection buffer times at the transfer station under consideration,

$$w_0^{\mathrm{prim}} = \sum_{(i,j)\in S_{\mathrm{p}}} n_{ij} r_{ij}^{\mathrm{s}} + \sum_{(i,j)\in T_{\mathrm{p}}} n_{ij} r_{ij}^{\mathrm{t}}$$

and this is also the reference generalized primary waiting time. The total relative waiting time with respect to a certain synchronization control policy then is

$$w^{\mathrm{tot}} = w^{\mathrm{prim}} - w_0^{\mathrm{prim}} + w^{\mathrm{sec}} - w_0^{\mathrm{sec}} \tag{3.16}$$

and the total relative generalized waiting time is

$$g^{\mathrm{tot}} = g^{\mathrm{prim}} - g_0^{\mathrm{prim}} + g^{\mathrm{sec}} - g_0^{\mathrm{sec}}. \tag{3.17}$$

The total relative waiting time can either be nonnegative or negative. In the first case the usage of the connection buffer times has a positive effect on the passenger waiting times. In the latter case the arrival delays are that large that on-time departures are impossible or not attractive.

Note that substituting the derived equations of the various (generalized) waiting times of the former subsections into (3.16) and (3.17) facilitates the expressions to a certain extend.

4 A Delay Propagation Forecasting Model

4.1 Introduction

This section describes a mathematical model for the computation of the delay propagation in a service network with respect to given initial departure delays. The model is formulated as a discrete event dynamic system (DEDS) and in particular as a max-plus algebra linear system.

A DEDS is a dynamic system where the state transitions are initiated by events that occur at discrete instants of time. An event corresponds to the start or the end of an activity. A scheduled service network is a typical example of a DEDS. An event is then an arrival at or a departure from a station. Characteristic is that the start of an activity (e.g., a departure time) depends on the termination of several other activities (train arrivals). Such systems cannot conveniently be described by differential or difference equations, and naturally exhibit a periodic behaviour. This section considers a special class of 'linear' discrete event dynamic systems. Here, the linearity has to be understood with respect to a non-standard algebraic structure, the so-called max-plus algebra.

The model of the train service network described in this section is based on the following assumptions:

(a) services operate according to a cyclic timetable: the scheduled arrival and departure times of the train services repeat regularly with the same interval time, the cycle time.

(b) early departures do not occur: this coincides with the aim of a timetable that a train may not depart until its scheduled departure time.

(c) train running times between transfer stations are deterministic: the running times, including stopping times at intermediate stations, are taken as realistic as possible and are not necessarily the same as the published trip times. This implies that, unlike the modelled scheduled departure times, the modelled scheduled arrival times do not have to coincide with the original timetable.

Assumption (b) implies that the arrival delay of a train at a particular transfer station equals its departure delay at the preceding transfer station. In an actual timetable slack times are usually distributed over running time margins and connection buffer times. Without loss of generality, in the model all slack times are concentrated in the connection buffer times.

Additionally, the delay propagation model has the following assumptions:

(d) scheduled transfers are secured during operation: since the purpose of the model is to evaluate the effect of synchronization control actions, it is reasonable that securing a transfer may not result in missed transfers at following transfer stations.

(e) initial perturbations are considered only: all successive departure delays correspond to propagation of initial delays.

Assumptions (d) en (e) can both be relaxed at cost of increased computational complexity. Relaxing assumption (d) relates to inclusion of controlled connections (De Vries *et al.*, 1998) and relaxing assumption (e) results in a stochastic model, see assumption (c).

The developed train service model is an analytical model rather than a simulation model although the model can also be used as a basis for simulations. Our aim is a comparative evaluation of the relative effect of different dispatching scenarios on the delay propagation in a service network. The analytical model efficiently computes this relevant macroscopic effect. Note that a concise microscopic simulation model including the safety and signalling system that is so characteristic for railway networks is not tractable at a large network level and also goes far beyond our purposes.

4.2 The Precedence Graph

Securing scheduled connections between individual trains at transfer stations is the cause for delay propagation in a train service network. Therefore, the model concentrates on the connections in the service network, and it is assumed that the train running times between transfer stations, including the stopping times at intermediate stops, are fixed. The physical railway network is only of implicit interest. Our main concern is the service network and the *precedence graph*. The latter is a graphical (network) representation of precedence constraints, which we will illustrate here by means of an example.

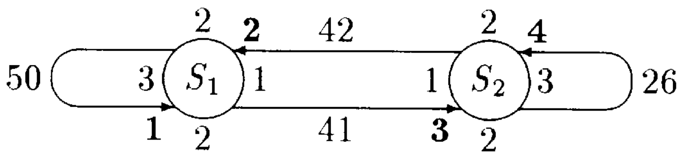


Figure 4.1 The railway network

Figure 4.1 shows an example railway network consisting of two transfer stations S_1 and S_2 and 4 routes (the routes are indicated in bold numbers). Intermediate stops along the routes have not been drawn. The service network consists of 3 lines (train series) with a total rolling stock of 6 trains, see Table 2. The weights at the arcs indicate the running times of the routes, and the weights around the nodes (transfer stations) indicate the minimum stopping or transfer times between the arriving and departing arcs (services). So, trains circulating on the routes 2 and 3 have a minimum stopping time of 1 minute at the transfer stations, and trains of route 1 and 4 have a minimum stopping time of 3 minutes at the transfer stations. The minimum transfer times are all 2 minutes.

Table 2 The line system

line	served routes	number of trains
1	1	2
2	2 & 3	3
3	4	1

Figure 4.2 shows the corresponding *precedence graph*. Each train service between two adjacent transfer stations is denoted as a separate arc. A node corresponds to the start of a train service at a transfer station, and the arcs represent precedence constraints corresponding to connections between train services. The arc weights are defined as the sum of the train running time between the adjacent transfer stations and the connection time between the connected services. The connection time can either be a stopping time or a transfer time depending on whether or not the connecting service is run by the same train.

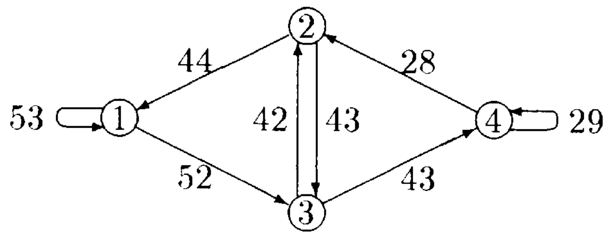


Figure 4.2 The precedence graph

An arc in the constructed precedence graph may represent a train run that covers several periods of the cycle time. This implies that such an arc corresponds to several running trains. As an example, consider the service on route 1. In the precedence graph of Figure 4.2 this service corresponds to arc (1,1). The arc weight is 53 and the cycle time is 30 minutes. This implies that at least two trains run the service to be able to depart every 30 minutes. Table 2 indicates that indeed two trains are allocated to this service.

It is convenient that each arc corresponds to a train run covering only one period. This can be accomplished by decomposing an arc into several segments and adding auxiliary nodes in between. The auxiliary nodes can be interpreted as (stop) connections, with zero connection time, at fictitious stations. The actual service is then augmented to several services in the model graph corresponding to the number of running trains. This procedure is illustrated for the example service network.

Consider the railway network of Figure 4.1. According to Table 2, line 1 serves route 1 with two trains. Therefore, on route 1 an additional fictitious station is assumed at 30 minutes (the cycle time) from station S_1. An auxiliary service 5 then runs from this station back to S_1 in the remaining running time (including the connection time). Note that in this way the departure time from the fictitious station corresponds to the departure time of the next train from station S_1 (with a shift of one period). However, the division of the original arc weight over the several services in the auxiliary precedence graph can be done arbitrary. The same procedure is now applied for line 2 that serves the two routes 2 and 3 with three trains. A fictitious station is now assumed on route 2, at 30 minutes distance from station S_2. An auxiliary service 6 then runs from this fictitious station to the transfer station S_2. The circuit containing the two routes 2 and 3 in the augmented model network now contains three services corresponding to the three running trains. Line 3 serves route 4 with only 1 train. So here no auxiliary services are necessary. Figure 4.3 shows the resulting augmented precedence graph. The 6 nodes coincide with the 6 trains running the service network.

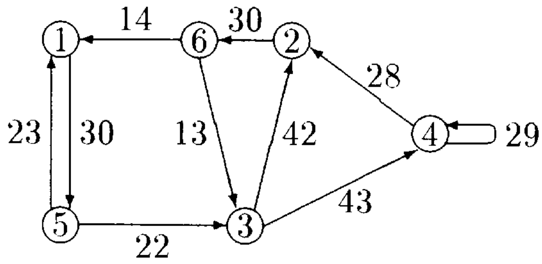


Figure 4.3 The augmented precedence graph

The augmented precedence graph corresponds to a 1st order representation of the service network, relating to the fact that all arcs coincide with train runs covering only one period. This will be made more specific in Sections 4.3 and 4.4. Furthermore, note that the precedence graph is strongly connected, i.e., there is a (directed) path between any node i to any node j, where a path is a sequence of adjacent nodes (without any repetition of nodes). This is consistent with the train circulations of the line system.

4.3 Discrete Event Dynamic Systems

A discrete event dynamic system (DEDS) description of the train service network is a system that expresses the departure times at all transfer stations in terms of the departure times at preceding transfer stations. This description is easily obtained from the precedence graph.

Consider the precedence graph. Assign to each node i a departure time x_i. This departure time depends on the arrival time from the train's preceding trip as well as on the arrival times of its feeder trains. The earliest possible departure time of a train i is therefore formally given as

$$x_i = \max_{j=1,\ldots,n} (a_{ij} + x_j), \qquad i = 1,\ldots,n, \tag{4.1}$$

where a_{ij} is defined as the sum of the train running time of service j and the connection time from service j to i,

$$a_{ij} = \begin{cases} t_j^{\mathrm{r}} + t_{ji}^{\mathrm{s}} & \text{if connection } ji \text{ is a stop} \\ t_j^{\mathrm{r}} + t_{ji}^{\mathrm{t}} & \text{if connection } ji \text{ is a transfer} \\ -\infty & \text{otherwise.} \end{cases} \tag{4.2}$$

Note that assigning $-\infty$ to non-existing connections implies that these trains can also be incorporated in (4.1) as these entries have no influence on the maximization (as long as other connections have finite entries). The train services j for which $a_{ij} \neq -\infty$ correspond to the predecessors of node i in the precedence graph. The fact that a weight a_{ij} corresponds to an arc (j,i) may be confusing at first sight. However, in the next section we show that by doing so, the railway system can be formulated as a familiar linear system $x(k+1)=Ax(k)$.

The trains operate according to a cyclic timetable, i.e., the pattern of arrival and departure times repeats every cycle time. Therefore the departure times are periodic recurrent events. Let k be a counter denoting a specific period. Then the kth departure time of a train service i is $x_i(k)$. Incorporating the periodicity in (4.1) gives

$$x_i(k+1) = \max_{j=1,\ldots,n}(a_{ij} + x_j(k)), \qquad i = 1,\ldots,n. \tag{4.3}$$

The departure time of a service i thus depends on former departure times of the preceding trains. In general a train may also be connected to a train that departed, say, l periods before. However, this situation is reduced to (4.3) by the constructed precedence graph, see Section 4.2. The original service has been augmented to l services in the precedence graph. The train then successively runs the additional auxiliary services before reaching the original connection after l periods.

If the train service network operates according to a timetable then a train may not depart before its scheduled departure time. However, if the train is behind schedule, or has to wait for a delayed feeder train, then the actual departure time may exceed the scheduled departure time. Denote the scheduled departure time of a train service i from a transfer station as d_i. Then the scheduled train service network can be described as

$$x_i(k+1) = \max\big(a_{i1} + x_1(k),\ldots,a_{in} + x_n(k), d_i(k+1)\big), \qquad i = 1,\ldots,n. \tag{4.4}$$

The subsequent scheduled departure times for train service i are given as

$$d_i(k) = d_i(0) + kT\,, \tag{4.5}$$

where T is the cycle time and $d_i(0)$ is an initial departure time of train i.

A timetable naturally contains connection buffer times defined as the intervals between the earliest possible departure times and the scheduled departure times at transfer stations, see Section 2.2. From (4.4) follows that the connection buffer time r_{ji} between an arriving service j and a departing service i is

$$r_{ji} = d_i(k+1) - a_{ij} - d_j(k) \tag{4.6}$$

for any integer $k \geq 0$.

Connection buffer times have a reducing or eliminating effect on arrival delays. In this way, the propagation of initial departure delays at a particular period can be computed. Each subsequent period the departure delays are reduced by the connection buffer time at the various connections. If initial departure times $x_1(0), \ldots, x_n(0)$ are given then the evolution of the train service system (4.3), or the scheduled service system (4.4), is completely determined, i.e., the subsequent departure times of all trains are uniquely fixed. These systems are examples of Discrete Event Dynamic Systems (DEDS). Here, a discrete event is a departure at a transfer station that occurs at a discrete instance in time, the departure time, and the dynamic equation, (4.3) or (4.4), describes the dynamic behaviour over the successive periods k. The above described systems are deterministic. If the parameters a_{ij} also depend on the period k then the system is stochastic. The subsequent running and connection times then are variable.

4.4 Max-Plus Linear Systems

The DEDS models (4.3) and (4.4) are examples of a special class of 'linear' discrete event systems. Here, the linearity has to be understood with respect to a non-standard algebraic structure, the max-plus algebra.

The max-plus algebra is very similar to conventional algebra but the addition is replaced by maximization, denoted as $\oplus$, and multiplication is replaced by the conventional addition, denoted as $\otimes$, i.e.,

$$\begin{aligned} a \oplus b &= \max(a, b), \\ a \otimes b &= a + b. \end{aligned}$$

The set of elements considered in the max-plus algebra is the real numbers and the additional element $\varepsilon = -\infty$. Note that the elements a_{ij} defined in (4.2) belong to this set. The extension to vectors and matrices is equivalent to the linear algebra: addition of two matrices (vectors) is defined componentwise and matrix multiplication is defined as

$$(A \otimes B)_{ij} = \bigoplus_{k=1}^{n} a_{ik} \otimes b_{kj} = \max_{k=1,\ldots,n} (a_{ik} + b_{kj}).$$

Concepts from linear algebra and linear system theory have their counterparts in the max-plus algebra, see for instance Baccelli *et al*. (1992). Table 3 illustrates some similarities between linear algebra and max-plus algebra. Here, we will not give an extensive treatment of the algebraic and system theoretic properties of the max-plus algebra but restrict to the modelling issues. The max-plus algebra modelling and (stability) analysis of (scheduled) railway systems is due to Braker (1993), see also Goverde (1997). A max-plus linear system approach to design service network timetables is presented in Goverde (1998).

Table 3 Linear algebra and max-plus algebra

linear algebra	max-plus algebra
operations	
$(A+B)_{ij} = a_{ij} + b_{ij}$	$(A \oplus B)_{ij} = a_{ij} \oplus b_{ij}$
$(cA)_{ij} = c \cdot a_{ij}$	$(c \otimes A)_{ij} = c \otimes a_{ij}$
$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$	$(A \otimes B)_{ij} = \bigoplus_{k=1}^{n} a_{ik} \otimes b_{kj}$
properties	
$(A+B)+C = A+(B+C)$	$(A \oplus B) \oplus C = A \oplus (B \oplus C)$
$A+B = B+A$	$A \oplus B = B \oplus A$
$(AB)C = A(BC)$	$(A \otimes B) \otimes C = A \otimes (B \otimes C)$
$A(B+C) = AB + AC$	$A \otimes (B \oplus C) = (A \otimes B) \oplus (A \otimes C)$
spectral analysis (eigenvalue and eigenvector)	
$Av = \lambda v$	$A \otimes v = \lambda \otimes v$

The DEDS model (4.3) contains the maximization operation which makes it nonlinear in a linear algebra sense. However, in the max-plus algebra equation (4.3) becomes

$$x_i(k+1) = \bigoplus_{j=1}^{n} (a_{ij} \otimes x_j(k)), \qquad i = 1, \ldots, n, \tag{4.7}$$

where $\oplus_{j=1}^{n} a_j = \max(a_1, \ldots, a_n)$ denotes repeated maximization. In vector notation (4.7) is written as

$$x(k+1) = A \otimes x(k), \tag{4.8}$$

where $x=(x_1,\ldots,x_n)'$ and A is the square $n \times n$ matrix whose ijth entry is a_{ij}. Equation (4.8) is a linear system in the max-plus algebra and is also simply written as $x(k+1) = Ax(k)$. The vector x is called the *state vector* and matrix A is called the *state matrix*. The state matrix corresponds to the precedence graph as defined in Section 4.2 where the entry a_{ij} is the weight of arc (j,i) with the convention that if $a_{ij} = \varepsilon$ then there is no arc from j to i.

As an example, the matrix A corresponding to the precedence graph of Figure 4.3 is

$$A = \begin{pmatrix} \varepsilon & \varepsilon & \varepsilon & \varepsilon & 23 & 14 \\ \varepsilon & \varepsilon & 42 & 28 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 22 & 13 \\ \varepsilon & \varepsilon & 43 & 29 & \varepsilon & \varepsilon \\ 30 & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 30 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \end{pmatrix}. \tag{4.9}$$

Vice versa, the precedence graph of this matrix A is the graph in Figure 4.3. The scheduled service network (4.4) is also a linear system in the max-plus algebra,

$$x(k+1) = Ax(k) \oplus d(k+1), \tag{4.10}$$

where d is the timetable vector (4.5) which in the max-plus algebra is

$$d(k) = d(0) \otimes T^k = \begin{pmatrix} d_1(0) + kT \\ \vdots \\ d_n(0) + kT \end{pmatrix}. \tag{4.11}$$

Consider again the railway network of Figure 4.1, where now Table 4 shows the timetable of its service network. The state matrix corresponding to the precedence graph of Figure 4.3 is given in (4.9). The timetable vector in the max-plus linear system model (4.10) is given by (4.11) with

$$d(0) = (2,15,0,17,2,15)' \tag{4.12}$$

and T = 30. The first four components of the timetable vector are the scheduled departure times of the four actual train services from the transfer stations and can be read from Table 4 directly. The last two components correspond to the departure times of the auxiliary services from the fictitious stations. Since we assumed 30 minutes running time and zero connection time these departure times correspond to the departure times from the preceding transfer station modulo 30.

Table 4 Timetable

line **station**	**cycle time 30 minutes** **1**	 **2**	 **3**
S_1 arrival	22	27	
S_1 departure	02	00	
S_2 arrival		11	13
S_2 departure		15	17

If an initial vector $x(0) = x_0$ is given then the evolution of the max-plus linear systems (4.8) or (4.10) is completely determined. Usually also an output vector is incorporated. For instance, assume we are interested in the departure times of the train services from the transfer stations. The state vector of the 1st order representation (4.8) also contains auxiliary variables corresponding to departure times from fictitious stations. The output vector is then defined as the vector of the non-auxiliary variables as

$$y(k) = Cx(k) \tag{4.13}$$

for a suitable chosen matrix C.

4.5 The Max-Plus Delay Propagation Model

The propagation of an initial delay in a scheduled service network can now be computed from the max-plus linear system (4.10) with the timetable dynamic equation (4.11), an initial state vector incorporating the initial departure delays and an *output vector* representing the departure delays in a period. Here, the initial vector is

$$x_0 = d(0) + y_0. \tag{4.14}$$

where y_0 is the initial departure delay vector, and the output vector is defined as

$$y(k) = x(k) - d(k). \tag{4.15}$$

The output vector $y(k)$ contains the departure delays in the kth period for all trains. The initial delay has disappeared as soon as the output vector becomes zero.

The delay propagation forecasting model is thus given as

$$\begin{cases} x(k+1) = Ax(k) \oplus d(k+1) \\ d(k) = d(0) \otimes T^k \\ y(k) = x(k) - d(k) \\ x(0) = x_0. \end{cases} \tag{4.16}$$

Note that the subtraction in the definition of the output vector has to be understood in the conventional linear algebra. Let K be the *settling period* in which the output vector has become zero, i.e., when the initial delays have vanished. Then, the output vectors

$$y(1), \ldots, y(K)$$

give the departure delays for all train services in the service network for all successive periods as caused by the synchronization control times.

Consider again the example train service network. The state matrix A is given by (4.9), the initial timetable vector $d(0)$ is given by (4.12), and $T = 30$. Assume that the train services 2 and 4 from station S_2 have an initial departure delay of 3 and 5 minutes, respectively. Then the initial state is given as

$$
\begin{aligned}
x_0 &= d(0) + y_0 \\
&= (2,15,0,17,2,15)' + (0,3,0,5,0,0)' \\
&= (2,18,0,22,2,15)'.
\end{aligned}
$$

The delay propagation in the service network can now be computed from (4.16). Table 5 shows the successive departure delays $y(k)$. In 7 periods the delays have vanished. However, in period 6 only the auxiliary service 6 still has a departure delay of 1 minute (which is the departure delay of actual service 2 in period 5). This results in an arrival delay of service 6 at transfer station S_1 which is compensated by the connection buffer times. Thus, the last actual departure delays are observed after 2½ hour (5 periods) and the last arrival delay is observed after 3½ hour.

Table 5 The departure delays $y(k)$

k **service**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	0	0	0	2	1	0	0	0
2	3	5	4	3	2	1	0	0
3	0	0	1	3	2	1	0	0
4	5	4	3	2	1	0	0	0
5	0	0	0	0	2	1	0	0
6	0	3	5	4	3	2	1	0

The implementation of the delay propagation model in the synchronization control evaluation model is now as follows. The modelled railway network includes the concerned transfer station and is formulated as the max-plus linear system (4.16). The initial state vector x_0 is defined as

$$
x_0 = \begin{cases} d_i(0) + s_i & \text{if train service } i \text{ gets a synchronization control time} \\ d_i(0) & \text{otherwise.} \end{cases} \tag{4.17}
$$

Naturally, the train services that obtain synchronization control time are departing services from the concerned transfer station. All departure times from other services in the network are assumed to be as scheduled. If online departure delays at other transfers stations in the network are known then these can also be incorporated in the initial delay vector.

For large models, i.e., service networks with many connections, the computation time for the delay propagation may get very high in a straightforward implementation. For instance, the entire Netherlands train service network is modelled by 435 variables/connections (Subiono, 1996).

However, the delay propagation of an initial delay vector can be computed by considering, for each period k, the connections with delayed arriving services only. Departing services of other connections depart on time, as the delays have not reached these connections. This speeds up the computations considerably since the state matrix is naturally a sparse matrix, i.e., each column of the state matrix only has a small number of non-zero entries (in a max-plus algebra sense the zero element is ε) corresponding to the connecting services of a single arriving train. Therefore, usually a limited number of computations are necessary, depending on the initial delay vector.

Another computational reduction is obtained by reformulation of the max-plus linear system (4.16). The delay propagation model (4.16) can be reformulated by noting that at each period the maximal delay reduction with respect to a particular connection equals the connection buffer time. Consider the following max-plus linear system

$$\begin{cases} x(k+1) = Rx(k) \oplus e_n \\ y(k) = x(k) \\ x(0) = y_0, \end{cases} \tag{4.18}$$

where e_n is the n-dimensional vector with each entry equal to the max-plus algebra neutral element $e = 0$ (and n is dimension of the state vector), and R is the matrix defined as

$$(R)_{ij} := \begin{cases} -r_{ji} & \text{if } ji \text{ is a connection} \\ -\infty & \text{otherwise.} \end{cases} \tag{4.19}$$

with the connection buffer time r_{ji} as defined in (4.6). The state vector can here be interpreted as the departure delay vector, which is also the output vector. The matrix R can be viewed as the delay absorption matrix. The following theorem shows that the max-plus linear systems (4.16) and (4.18) have the same output.

Theorem 4.1 *Consider the max-plus linear systems* (4.16). *Let the max-plus linear system* (4.18) *be defined with matrix R as in* (4.19) *with*

$$r_{ji} = d_i(k+1) - a_{ij} - d_j(k)$$

for any integer $k \geq 0$. Then system (4.16) *and* (4.18) *have the same output vectors.*

Proof Consider system (4.16). Then for each entry i

$$x_i(k+1) = \max\left[\max_{j=1,\dots,n}(a_{ij} + x_j(k)), d_i(k+1)\right] \quad \Leftrightarrow$$

$$x_i(k+1) - d_i(k+1) = \max\left[\max_{j=1,\dots,n}(a_{ij} + x_j(k) - d_i(k+1)), 0\right] \quad \Leftrightarrow$$

$$y_i(k+1) = \max\left[\max_{j=1,\dots,n}(-d_i(k+1) + a_{ij} + d_j(k) + x_j(k) - d_j(k)), 0\right] \Leftrightarrow$$

$$y_i(k+1) = \max\left[\max_{j=1,\dots,n}(-r_{ji} + y_j(k)), 0\right].$$

And this last equation is exactly the max-plus linear system (4.18). q.e.d.

As a result of Theorem 4.1, the delay propagation model (4.16) may also be replaced by the equivalent model (4.18). Note that this implies that the computations for the timetable vector and the output vector in each period in (4.18) are no longer necessary. The max-plus linear system (4.18) thus gives the same results with less computational effort.

As an illustration, the matrix R corresponding to the example system is

$$R = \begin{pmatrix} \varepsilon & \varepsilon & \varepsilon & \varepsilon & -7 & -3 \\ \varepsilon & \varepsilon & -3 & 0 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & -6 & -2 \\ \varepsilon & \varepsilon & -4 & -1 & \varepsilon & \varepsilon \\ 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \end{pmatrix}. \tag{4.20}$$

The delay propagation model for the example system is then completely determined by (4.18) with matrix R defined in (4.20) and given initial delay vector y_0. For example, for initial delay vector $y_0 = (0,3,0,5,0,0)'$ again Table 5 results.

5 Example

Consider the railway network of Figure 4.1. We here compute the optimal synchronization control times for various scenarios of arrival delays with respect to the connections at transfer station S_2.

First, consider the train service network. The max-plus delay propagation model is given as (4.16), with the state matrix given in (4.9), the initial timetable vector given in (4.12), and $T = 30$. The initial state vector is given as (4.17) for the various candidate synchronization control times.

With respect to the computations of the secondary waiting time (3.15), the set of stop connections and transfer connections are

$$S_s = \{(1,5),(5,1),(2,6),(6,3),(3,2),(4,4)\},$$

$$T_{\mathrm{s}} = \{(6,1),(5,3),(3,4),(4,2)\}.$$

Note that the conjunction of both sets comprise all connections in the precedence graph of Figure 4.3.

Next, the (relative) passenger flows have to be defined. Suppose that the surroundings of both transfer stations are industrial areas attracting a lot of commuters through the train service network. The relative passenger flows have been determined as follows: The through flows of line 2 (see Table 2) at both transfer stations are zero. The through flows of lines 1 and 3 are 1 (relative unit) reflecting the small amount of passengers that pass the transfer stations as an intermediate station on routes 1 and 4, respectively. The transfer flows from line 1 and 3 to line 2 are both 2 units reflecting the commuters from the areas around the transfer stations who are employed near the other transfer station. The transfer flows on the opposite directions, from line 2 to line 1 and 3, respectively, are both 1 unit. Finally, the originating flows of all lines are 5 units. Note that this includes the passengers who start at an intermediate station on a particular route. From the above flows also the terminating flows can be derived as 4 units from routes 2 and 3 (line 2), and 2 units from routes 1 and 4, respectively. Additional attention is required for the passenger flows for the auxiliary variables in the max-plus model, i.e., at the fictitious stop connections on the routes 1 and 2. Here, the originating, transfer, and terminating flows are zero, and the through flows correspond to the passengers flows of a running train between the (actual) transfer stations. The through flow for the auxiliary variable 5 on route 1 is 5 units, and for the auxiliary variable 6 on route 2 it is 6 units. Note that this represents a consistent system of passenger flows.

We now concentrate on the transfer station S_2. At this station two lines (train series) stop with transfers in both directions. The arriving services are denoted as service 3 and 4, and the departing services are 2 and 4. It follows that for the computation of the primary waiting time (3.15) the set of departing train services is $O_{\mathrm{p}} = \{2,4\}$, the set of stop connections is $S_{\mathrm{p}} = \{(3,2),(4,4)\}$, and the set of transfer connections is $T_{\mathrm{p}} = \{(3,4),(4,2)\}$. The problem now is to find the optimal synchronization control policies for the departing services 2 and 4 with respect to arrival delays of arriving services 3 and 4.

The max-plus delay propagation model and the computations of the primary and secondary (generalized) waiting times have been implemented in Matlab. For the example system, the effect has been computed of the candidate synchronization control times of departing services 2 and 4 on the passenger waiting times with respect to arrival delays of the arriving services 3 and 4 ranging from 0 to 10 minutes.

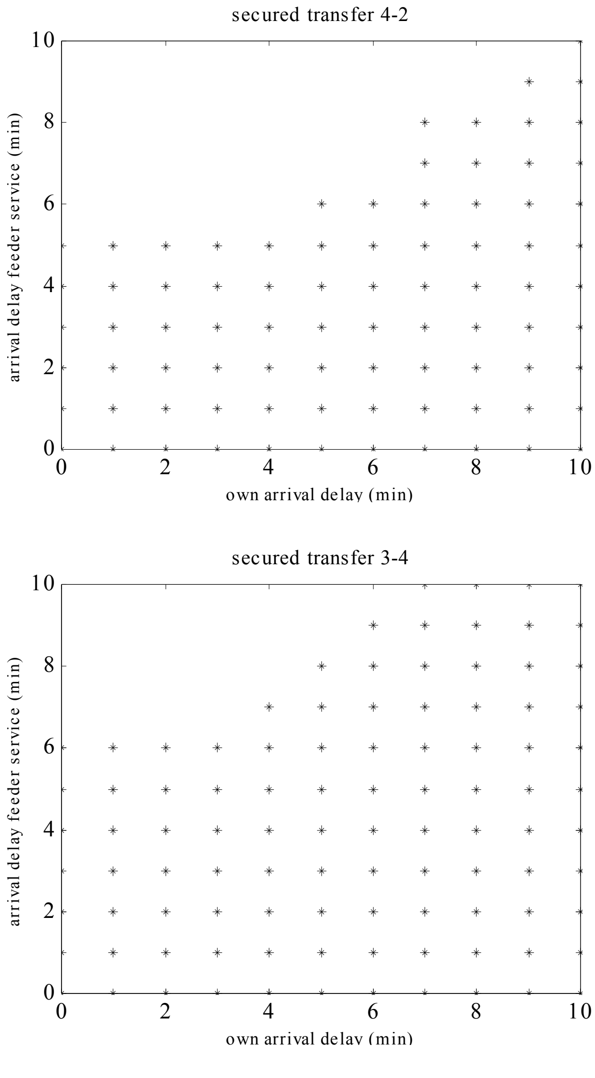
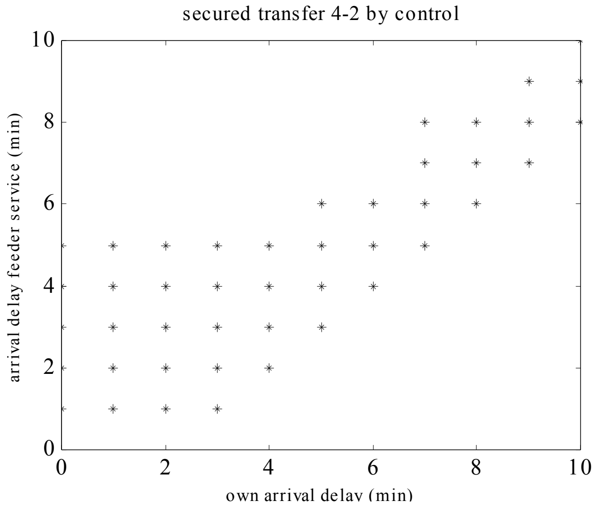


Figure 5.1 All secured transfers for various arrival delay pairs at transfer station S_2 that minimize the total waiting time

Figure 5.1 shows the secured transfers 4-2 and 3-4, respectively, as a function of the arrival delay of the train itself (horizontal axis) and of the feeder train (vertical axis). Here, the objective function of the synchronization control problem is the minimization of the total relative waiting time. Note that the arriving service 3 is run by the same train that proceeds as service 2, and the feeder service of connection 4-2 is train 4, so that for connection 4-2 the horizontal axis corresponds to the arrival delay p_3 and the vertical axis to p_4. For connection 3-4 this is vice-versa.

A wide range of transfers is secured by the transfer buffer time and/or an arrival delay of the train itself. Figure 5.2 shows the additional secured transfers as a direct result of securing the connection by means of the optimal synchronization control.



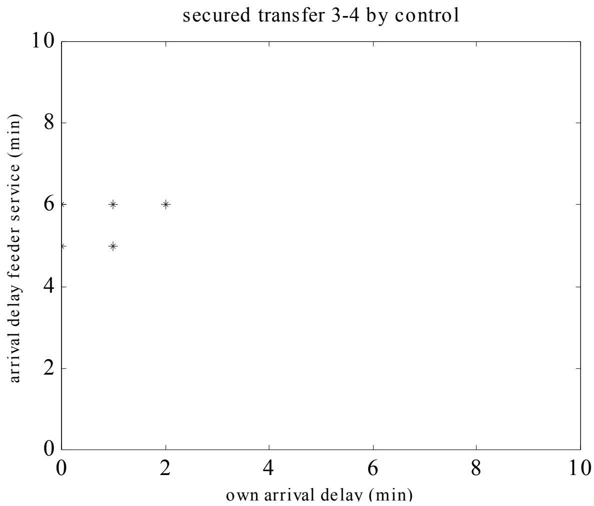


Figure 5.2 Secured transfers for various arrival delay pairs as a result of synchronization control with objective function the total waiting time.

The transfer buffer time of connection 4-2 is 0, see (4.20). So without control, the transferring passengers from service 4 to 2 miss their connection immediately if the feeder train has an arrival delay. Figure 5.2 (upper picture) shows the result of the synchronization control. If the feeder train has an arrival delay of 5 minutes or smaller and so does the feeder train then the connection is secured. For larger arrival delays of both trains the optimal policy secure/dissolve depends on the situation. Note that the stopping buffer time of the connecting service of 3 minutes, see (4.20), is also a determining factor. Figure 5.2 (upper picture) also shows that it is still optimal with respect to overall passenger waiting times that a train waits on a delayed feeder train even if its own arrival delay is already considerable. So from an operator point of view it then might be arguable to still dissolve the connection and prevent further perturbation of train movements on the traffic network.

Figure 5.2 (lower picture) shows the optimal policies for connection 3-4. Here, the transfer buffer time is 4 minutes which already compensates for a large amount arrival delays of the feeder train. Therefore, there is only a small amount of additional secured connections as a result of the optimal control policy.

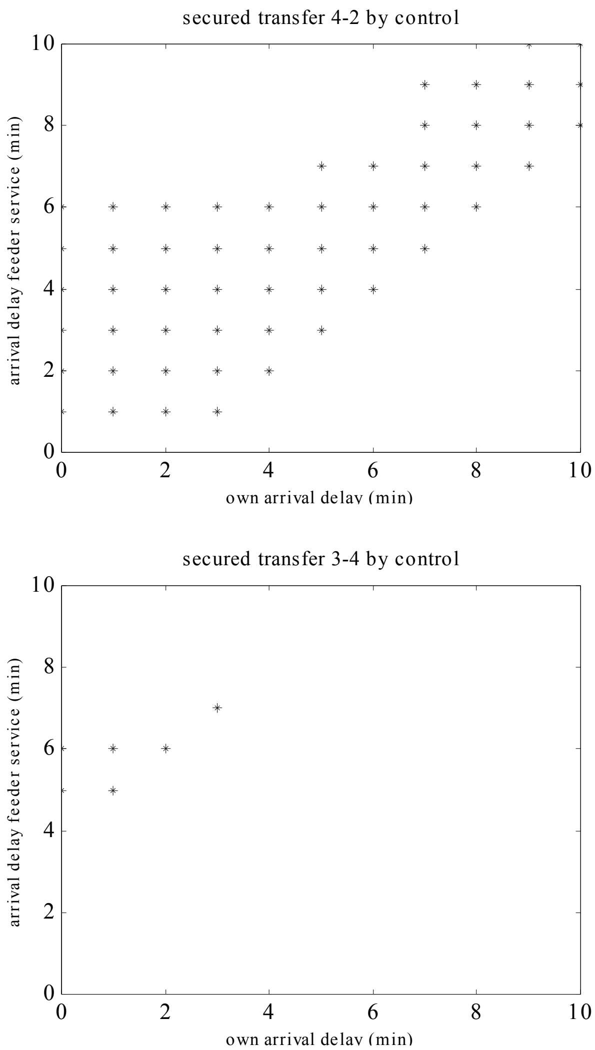


Figure 5.3 Secured transfers for various arrival delay pairs as a result of synchronization control with objective function the total generalized waiting time

Figure 5.3 shows the optimal control policy results when the objective function is the minimization of the relative total generalized waiting time. Now, a few more connections are secured in comparison to Figure 5.2. For connection 3-4 only one additional transfer is secured in the case of $p_3 = 7$ and $p_4 = 3$. It appears that here the generalized total waiting time reduces from 156 to 140 when securing this connection for these particular arrival delays. The total waiting time increases slightly from 76 to 80. This is caused

by an increase of the synchronization control time of service 4 from 2 to 3 minutes, by which the settling period increases from 4 to 5 periods.
For connection 4-2, Figure 5.3 (upper picture) shows 10 additional secured connections when applying the total generalized waiting time as the objective function. As an example, for the arrival delays $p_3 = 0$ and $p_4 = 6$, the total generalized waiting time decreases from 476 to 440, whereas the total waiting time increases slightly from 218 to 226 minutes. The settling time is not changed and stays 7 periods. The corresponding synchronization control time for service 2 has increased from 0 to 5 minutes.

As a final example, securing connection 4-2 for the arrival $p_3 = 5$ and $p_4 = 7$ reduces the total generalized waiting time from 661 to 158, while the total waiting time increases from 305 to 322 minutes and the settling period remains 8 periods. The corresponding synchronization control time for service 2 has increased from 2 to 7 minutes, whereas the departure delay of service 4 is 6 minutes.

6 Conclusions and Recommendations

This paper describes a mathematical model to derive optimal synchronization control times as a dynamic counterpart of existing static guidelines. It can be used either online to compute dynamic synchronization control times and offline to systematically derive pseudo-dynamic synchronization control margins for ranges of arrival delays and periods with different passenger flow patterns.

Several extensions of the model are possible. First, minimum departure headways between services may be taken into account. This can be realized quite straightforward by incorporating the minimum headways in computing the candidate synchronization control times. The effect of knock-on delays between trains at subsequent track sections is more complicated. A possible approach is the inclusion of the optimal order-swapping strategy of Carey & Kwiecinski (1994). Note that the relevant strategy corresponds to their full information case. However, this requires an extension of the results of Carey & Kwiecinski to multiple connections. Also queueing on the preceding track sections of the station may be taken into account.

The presented model of the generalized waiting time is a linear regression model based entirely on the difference of the passenger inconvenience to waiting on a platform or in a train. More accurate models may be formulated. This is a current research topic at the Delft University of Technology. Finally, another extension is the inclusion of route choice of passengers who miss a transfer. In this case the transfer waiting time does not only depend on the headway of the connecting service but also on the expected remaining time to the departure times of other favourable services.

The model has to be validated in a real test environment. With respect to the Netherlands train service network, the max-plus linear system of Subiono (1997) of 435 variables can be used for the timetable of 1996/1997.

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

Adamski, A and Turnau, A. (1998), “Simulation Support Tool for Real-Time Dispatching Control in Public Transport”, *Transportation Research A*, Vol. 32, No. 2, pp. 73-87.

Baccelli, F.L., Cohen, G., Olsder, G.J. and Quadrat, J.-P. (1992), *Synchronization and Linearity: an Algebra for Discrete Event Systems*, Wiley, Chicester.

Betgé-Brezetz, S., Benoliel, S., Bain, G., Ménapace, G. (1998), “Computer Aided Methods for Railway Operation Conflict Solving”, In: Mellit, B., Hill, R.J., Allan, J., Sciutto, G., Brebbia, C.A. (eds.), *Computers in Railways VI*, Computational Mechanics Publications / WIT Press, Southampton, pp. 563-572.

Bookbinder, J.H., Désilets, A. (1992), “Transfer Optimization in a Transit Network”, *Transportation Science*, Vol. 26, No. 2, pp. 106-118.

Braker, J.G. (1993), *Algorithms and Applications in Timed Discrete Event Systems*, PhD Thesis, Delft University of Technology, Delft.

Carey, M. and Kwiecinski, A. (1994), “Swapping the Order of Scheduled Services to Minimize Expected Costs of Delays”, *Transportation Research B*, Vol. 28, No. 6, pp. 409-428.

Goeverden, C.D. van, Behr, F.H., Schoemaker, Th.J.H. and Wiggenraad, P.B.L. (1990), *Onderzoek overstapweerstand binnen het spoorwegsysteem* (in Dutch), Report VK 5302.308, Faculty of Civil Engineering, Delft University of Technology, Delft.

Goverde, R.M.P. (1996), *Max-plus algebra voor de spoorwegen: een inleiding* (in Dutch), Report, TRAIL Research School, Delft.

Goverde, R.M.P. (1997), “Synchronization Management in Railway Systems”, In: *3rd TRAIL Year Congress 1997*, Conference Proceedings, Part 2, The Netherlands TRAIL Research School, Delft.

Goverde, R.M.P. (1998), “The Max-Plus Algebra Approach to Railway Timetable Design”, In: Mellit, B., Hill, R.J., Allan, J., Sciutto, G., Brebbia, C.A. (eds.), *Computers in Railways VI*, Computational Mechanics Publications / WIT Press, Southampton, pp. 339-350.

Jovanic, D. and Harker, P.T. (1991), “Decision Support System for Train Dispatching: An Optimization-Based Methodology”, *Transportation Research Record*, Vol. 1314, pp. 31-40.

Kannengießer, F. and Wiche, B. (1987), “Einfluß der Fahrzeitenzuschläge und Pufferzeiten auf die Pünktlichkeit” (in German), *Die Bundesbahn*, Vol. 63, No. 11, pp. 1001-1007.

Knoppers, P. and Muller, Th. (1995), “Optimized Transfer Opportunities in Public Transport”, *Transportation Science*, Vol. 29, No. 1, pp.101-105.

NS Reizigers (1995), *Wachttijden voor reizigerstreinen (WRT): Dienstregeling 1995/1996* (in Dutch), Nederlandse Spoorwegen, NSR Logistiek, Utrecht.

Ruyter, R. de, Muller, Th.H.J. and Hakkesteegt, P. (1990), *Optimalisatie afwikkeling treinverkeer* (in Dutch), Report VK 4509.301, Faculty of Civil Engineering, Delft University of Technology, Delft.

Stolk, A. (1998), “Automatic Conflict Detection and Advanced Decision Support for Optimal Usage of Railway Infrastructure: Purpose and Concepts”, In: Mellit, B., Hill, R.J., Allan, J., Sciutto, G., Brebbia, C.A. (eds.), *Computers in Railways VI*, Computational Mechanics Publications / WIT Press, Southampton, pp. 629-638.

Subiono (1997), “On Large Scale Max-Plus Algebra Models in Railway Systems”, In: *3rd TRAIL Year Congress 1997*, Conference Proceedings, Part 2, The Netherlands TRAIL Research School, Delft.

Vries, R. de, Schutter, B. de, Moor, B. de (1998), “On Max-Algebraic Models for Transportation Networks”, Proceedings of the Fourth Workshop on Discrete Event Systems (WODES 98), Cagliary, Italy.

Waard, J. van der (1989), *Onderzoek weging tijdelementen: Deelrapport 5: Samenvatting en conclusies* (in Dutch), Report VK 5302.305, Faculty of Civil Engineering, Delft University of Technology, Delft.

Yan, S. and Yang, D.-H. (1996), "A Decision Support Framework for Handling Schedule Perturbation", *Transportation Research B*, Vol. 30, No. 6, pp. 405-419.

Appendix A: List of Symbols

A	state matrix,
$d(k)$	vector of kth scheduled departure times,
g_j^{o}	generalized originating waiting time with respect to service j,
g_{ij}^{s}	generalized through waiting time with respect to stop ij,
g_{hij}^{t}	generalized transfer waiting time w.r.t. transfer ij and stop hj,
g^{prim}	total primary generalized waiting time,
g^{sec}	total secondary generalized waiting time,
h_j	interdeparture time of service j,
k	period,
n_j	number of originating passengers (originating flow) of service j,
n_{ij}	number of through or transfer passengers w.r.t. connection ij,
p_i	arrival delay of service i,
r_{ij}	connection buffer time of connection ij (either transfer or stop),
r_{ij}^{t}	transfer buffer time of transfer connection ij,
r_{ij}^{s}	stopping buffer time of stop connection ij,
s_j	synchronization control time of service j,
s_{ij}	synchronization control time of service j w.r.t. to feeder i,
$\bar{s}_j$	synchronization control margin of service j,
S_i	transfer station i,
t_{ij}^{t}	transfer time of transfer connection ij,
$t_{ij}^{\text{t,min}}$	minimum transfer time of transfer connection ij,
t_i^{r}	running time of service i,
t_{ij}^{s}	stopping time of stop connection ij,
$t_{ij}^{\text{s,min}}$	minimum stopping time of stop connection ij,
T	cycle time,
w_j^{o}	originating waiting time with respect to service j,
w_{ij}^{s}	through waiting time with respect to stop connection ij,
w_{ij}^{t}	transfer waiting time with respect to transfer connection ij,
w^{prim}	total primary waiting time,
w^{sec}	total secondary waiting time,
$x(k)$	vector of kth departure times,
$y(k)$	vector of departure delays in period k,
ε	zero element in max-plus algebra, $\varepsilon = -\infty$ (minus infinity).