

Flexible Connections in PESP Models for Cyclic Passenger Railway Timetabling

Leo G. Kroon, Leon W.P. Peeters, Joris C. Wagenaar,
and Rob A. Zuidwijk

ERIM REPORT SERIES <i>RESEARCH IN MANAGEMENT</i>	
ERIM Report Series reference number	ERS-2012-008-LIS
Publication	May 2012
Number of pages	33
Persistent paper URL	http://hdl.handle.net/1765/32996
Email address corresponding author	lkroon@rsm.nl
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus Universiteit Rotterdam P.O.Box 1738 3000 DR Rotterdam, The Netherlands Phone: + 31 10 408 1182 Fax: + 31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

REPORT SERIES
RESEARCH IN MANAGEMENT

ABSTRACT AND KEYWORDS	
Abstract	In this paper we describe how rolling stock and passenger connections in a cyclic railway timetable can be modeled in a flexible way within the model for the Periodic Event Scheduling Problem (PESP). The PESP model was introduced by Serani and Ukovich (1989). Usually, PESP models assume that the constraints for rolling stock or passenger connections specify in detail which trains should connect with each other. However, the flexibility described in this paper allows the model to choose which trains should connect with each other in a rolling stock or passenger connection. We express the required number of train compositions in terms of the integer cycle variables of the constraint graph. We also describe an abstract framework, demonstrating that, under certain conditions, the extra flexibility can be modeled purely in terms of PESP constraints. The concept of flexible rolling stock and passenger connections is illustrated by an example based on three intercity lines of Netherlands Railways.
Free Keywords	cyclic railway timetabling, rolling stock and passenger connections, PESP
Availability	The ERIM Report Series is distributed through the following platforms: Academic Repository at Erasmus University (DEAR), DEAR ERIM Series Portal Social Science Research Network (SSRN), SSRN ERIM Series Webpage Research Papers in Economics (REPEC), REPEC ERIM Series Webpage
Classifications	The electronic versions of the papers in the ERIM report Series contain bibliographic metadata by the following classification systems: Library of Congress Classification, (LCC) LCC Webpage Journal of Economic Literature, (JEL), JEL Webpage ACM Computing Classification System CCS Webpage Inspec Classification scheme (ICS), ICS Webpage

Flexible connections in PESP models for cyclic passenger railway timetabling

Leo G. Kroon Leon W.P. Peeters
Joris C. Wagenaar Rob A. Zuidwijk

Erasmus University Rotterdam
Rotterdam School of Management
P.O. Box 1738, 3000 DR Rotterdam
The Netherlands

E-mail: lkroon@rsm.nl; jwagenaar@rsm.nl; rzuidwijk@rsm.nl

May 17, 2012

Abstract

In this paper we describe how rolling stock and passenger connections in a cyclic railway timetable can be modeled in a flexible way within the model for the Periodic Event Scheduling Problem (PESP). The PESP model was introduced by Serafini and Ukovich (1989). Usually, PESP models assume that the constraints for rolling stock or passenger connections specify in detail which trains should connect with each other. However, the flexibility described in this paper allows the model to choose which trains should connect with each other in a rolling stock or passenger connection. We express the required number of train compositions in terms of the integer cycle variables of the constraint graph. We also describe an abstract framework, demonstrating that, under certain conditions, the extra flexibility can be modeled purely in terms of PESP constraints. The concept of flexible rolling stock and passenger connections is illustrated by an example based on three intercity lines of Netherlands Railways.

Keywords: Cyclic railway timetabling; Rolling stock and passenger connections; PESP.

1 Introduction

The Periodic Event Scheduling Problem (PESP), introduced by Serafini and Ukovich (1989), can be used for the scheduling of a number of periodically recurring events. Railway operators use the PESP model to develop cyclic timetables. The usual PESP model for cyclic railway timetables uses fixed rolling stock and passenger connections to develop the timetable. However, in this paper we describe how the rolling stock and passenger connections can be modeled in a flexible way within the PESP model, thereby leading to better timetables with respect to passenger connection times, running and dwell times, or the number of train compositions required to operate the timetable.

In an instance of PESP, we are given a set V of events for which the event times have to be determined. The events must be scheduled periodically with time period T . That means that, if event j takes place at time π_j , it should take place at times $\dots, \pi_j - 2T, \pi_j - T, \pi_j, \pi_j + T, \pi_j + 2T, \dots$, as well. As a consequence, for each event $j \in V$, it may be assumed that $\pi_j \in [0, T)$.

The events are pairwise connected by a set of constraints, specifying that the difference of the corresponding event times (modulo T) must be in a certain time interval. That is, there is a set of constraints A , where each constraint $a \in A$ is connected with a pair of events $j, k \in V$, a lower bound ℓ_a and an upper bound u_a , specifying that

$$\ell_a \leq (\pi_k - \pi_j - \ell_a) \bmod T + \ell_a \leq u_a. \quad (1)$$

In the following, the constraint (1) is also denoted as $\pi_k - \pi_j \in [\ell_a; u_a]_T$. More general, if U is a set of real numbers, then we use the notation U_T for the set of real numbers $\{ r \in \mathbb{R} \mid \exists z \in \mathbb{Z} : r - Tz \in U \}$, the periodized set of U .

In the context of cyclic railway timetabling, the events represent the departures and arrivals of trains at the stations. The constraints represent the process times, and are typically used to model the following issues:

- Running times between two consecutive stations of the same train. Here often $\ell_a = u_a$, but in case of mixed traffic (e.g. passenger and freight trains) more flexibility may be needed to get a feasible solution (Kroon and Peeters, 2003);
- Dwell times of trains in the stations, where ℓ_a is either the time the passengers need to board and alight, or some technical time (e.g. change of direction, train coupling or sharing, change of locomotive), and u_a is intended as a service commitment for the through-passengers;
- Passenger connection times in stations, where ℓ_a models the minimum time that is required to change from one train to an other, and u_a is used as an explicit service commitment for the most important transfers;
- Minimum headways on tracks, where –in the easiest case of two trains having the same speed– ℓ_a is just the technical minimum headway, and $u_a := T - \ell_a$ is used to guarantee a minimum headway if the trains run in the reverse order;
- Synchronized departures, where two trains are assumed to provide a balanced service on a common rail segment. Here typically $\ell_a = u_a = \frac{T}{2}$.

In this paper, we show by a number of examples that a priori fixing the rolling stock or passenger connections may lead to suboptimal results with respect to passenger connection times, running and dwell times, and the number of train compositions required to operate the timetable. It may even lead to infeasibility, although a feasible solution would exist if the connections were selected otherwise. We express the number of train compositions needed to operate the timetable in terms of certain cycle variables in the constraint graph. We also present a general framework for describing the flexible connections in terms of PESP constraints. Finally, we show the added value of the flexible rolling stock and passenger connections based on a case involving three intercity lines of Netherlands Railways, the largest operator of passenger trains in the Netherlands.

This paper is structured as follows. In Section 2 we describe some general properties of PESP models. Section 3 deals with rolling stock connections, and with counting the number of train compositions needed to operate a timetable. Section 4 focuses on passenger connections. Next, in Section 5 we describe the general framework for the flexible connections. In Section 6 we present the case involving the three intercity lines of Netherlands Railways. The paper is finished with some conclusions in Section 7.

2 Properties of PESP

In this section we provide some basic properties of PESP that are used in the remainder of this paper. In constraint (1) one may assume without loss of generality that $0 \leq \ell_a < T$ and $0 \leq u_a - \ell_a < T$. Indeed, a constraint not satisfying the latter condition is void in a periodic system with period T . Note that it may happen that $u_a > T$.

An instance of PESP can be represented by a directed constraint graph $D = (V, A)$, where the nodes correspond with the events and the arcs correspond with the constraints. Each arc $a \in A$ is associated with a lower bound ℓ_a and an upper bound u_a .

There may be several arcs between a pair of nodes, but the graph may not contain loops. Furthermore, we may use the terms “events” and “nodes” interchangeably. The same holds for the terms “constraints” and “arcs”.

In the following, we may also write (j, k) for an arc $a \in A$ that is directed from node $j \in V$ to node $k \in V$, in particular if there is only one arc directed from node j to node k .

It is NP-complete to determine whether for a given constraint graph (D, ℓ, u) there exists a vector π satisfying constraint (1) for all arcs $a \in A$ (Odijk, 1996).

The constraint (1) is difficult to handle in MIP models. However, it is well known that constraint (1) is equivalent to the following constraint (2):

$$\ell_a \leq \pi_k - \pi_j + T \times p_a \leq u_a. \quad (2)$$

Here $p_a \in \mathbb{Z}$ is an integer decision variable. Note that (1) is not completely equivalent to the following constraint that is also often mentioned in the context of PESP models:

$$\ell_a \leq (\pi_k - \pi_j) \bmod T \leq u_a. \quad (3)$$

For example, in constraint (3) an upper bound $u_a \geq T$ does not make sense.

Instead of describing the timetabling problem in terms of the events $j \in V$ and the event times $\pi_j \in [0, T)$, one can also use the process times x_a of the arcs $a \in A$. Here x_a represents the periodic time difference between the event times π_j and π_k of nodes j and $k \in V$ if arc a is directed from node j to node k . In that case, the process time x_a is defined as $x_a = (\pi_k - \pi_j - \ell_a) \bmod T + \ell_a$ from constraint (1) or as $x_a = \pi_k - \pi_j + T \times p_a$ from constraint (2). A process time variable x_a satisfies $\ell_a \leq x_a \leq u_a$.

When using the process times instead of the event times in the model, the cycles in the constraint graph play an important role in deciding whether the process times provide a feasible solution. A cycle C in the constraint graph contains forward arcs C^+ and backward arcs C^- , such that changing the direction of the backward arcs leads to a directed cycle. The following result is well known:

Proposition 2.1 (Liebchen and Peeters, 2009). *Let (D, ℓ, u) be a constraint graph representing a PESP instance, and $x \in \mathbb{Q}^A$. Then the following statements are equivalent:*

- *The vector x represents a feasible set of process times for the PESP instance.*
- *For every cycle C in D , we have*

$$\sum_{a \in C^+} x_a - \sum_{a \in C^-} x_a = T \times q_C \text{ for some } q_C \in \mathbb{Z}, \quad \text{and} \quad \ell \leq x \leq u.$$

In particular, if x is a feasible solution, arc a is directed from node j to node k , and arc b is directed from node k to node j , then $x_a + x_b = 0 \pmod T$. Thus, if $0 < x_a, x_b < T$, then $x_a + x_b = T$. The following result of Odijk (1996) is used several times in this paper.

Proposition 2.2 (Odijk, 1996). *Let (D, ℓ, u) be a constraint graph representing a PESP instance. Then for any cycle C in D , the following inequalities are valid*

$$\left\lceil \frac{1}{T} \left(\sum_{a \in C^+} \ell_a - \sum_{a \in C^-} u_a \right) \right\rceil \leq q_C \leq \left\lfloor \frac{1}{T} \left(\sum_{a \in C^+} u_a - \sum_{a \in C^-} \ell_a \right) \right\rfloor \quad (4)$$

The integer variable q_C corresponding to cycle C will be called the *integer cycle variable* of cycle C . Another property of PESP that we will often use in the remainder of this paper is that it allows a process time between two events to lie in either of two periodic time intervals. This property is illustrated in Figure 1.

Proposition 2.3 (Peeters, 2003). *If $0 \leq \ell_1 \leq u_1 < \ell_2 \leq u_2 < T$, then the constraints $\pi_k - \pi_j \in [\ell_1; u_1]_T \vee \pi_k - \pi_j \in [\ell_2; u_2]_T$ are equivalent to the constraints $\pi_k - \pi_j \in [\ell_1; u_2]_T \wedge \pi_k - \pi_j \in [\ell_2; u_1 + T]_T$.*

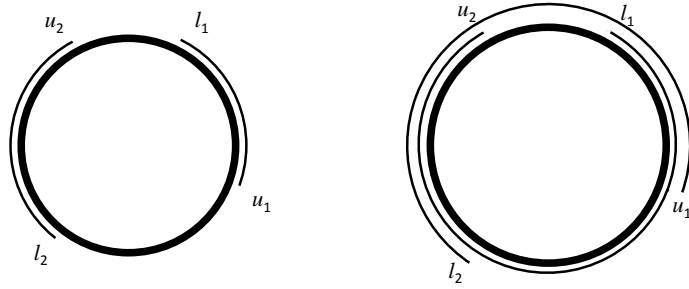


Figure 1: A union of two periodic time intervals is also the intersection of two periodic time intervals.

This property can be generalized easily to more than two periodic time intervals. Thus in a PESP model certain disjunctive constraints can be replaced by conjunctive constraints.

A further relevant property of the integer cycle variables q_C is that they indicate the periodic order in which the events of the corresponding cycle take place, as is expressed

in the following proposition. By definition, the events j , k and l take place in the periodic order $j \rightarrow k \rightarrow l$ if $(\pi_k - \pi_j) \bmod T < (\pi_l - \pi_j) \bmod T$. Note that the events j , k and l take place in the periodic order $j \rightarrow k \rightarrow l$ if and only if they take place in the periodic order $k \rightarrow l \rightarrow j$ if and only if they take place in the periodic order $l \rightarrow j \rightarrow k$.

Proposition 2.4 (Peeters, 2003). *If (D, ℓ, u) is a constraint graph representing a PESP instance, wherein nodes j , k and l are connected by arcs (j, k) , (k, l) , and (l, j) with upper bounds less than T , then the corresponding events are scheduled in the periodic order $j \rightarrow k \rightarrow l$ if and only if the cycle variable of the directed cycle $j \rightarrow k \rightarrow l$ has value 1.*

Note that, if the upper bounds on the arcs are less than T , then the only other value the cycle variable of the directed cycle $j \rightarrow k \rightarrow l$ may have is 2. In that case, the corresponding events are scheduled in the periodic order $j \rightarrow l \rightarrow k$.

3 Rolling stock circulation

In this section we describe how flexible rolling stock connections at the endpoints of a train line are helpful when minimizing the number of train compositions needed for operating the timetable. We also describe how this minimization problem can be expressed in a PESP model. We start with a simple example. This example considers a situation of a line that is operated twice per hour. Note that, if *all* lines of the railway system have the same frequency, then the situation can be simplified by using a smaller time period T . However that is usually not the case. The latter justifies the analysis in this paper.

3.1 Example

In this example we consider a train line that is operated twice per hour between the end stations A and B with exactly 30 minutes between two consecutive departures in the same direction from each (underway) station. The rolling stock on this line can be operated with two separate circulations, or with one combined circulation, depending on the rolling stock connections at the endpoints.

As a consequence, if one fixes the rolling stock connections at the endpoints a priori, then one may not end up with a timetable requiring the minimum number of train compositions. Therefore choosing the rolling stock connections at the endpoints should preferably be part of an optimization model aiming at minimizing the required number of train compositions.

Suppose the running time in each direction of the line must be between 35 and 40 minutes, and the return time (or turn around time) at each endpoint must be between 10 and 25 minutes. Here the return time is the time between the arrival of a train and the departure of the next train in the reverse direction. Then the minimum circulation time of a train is $2 \times (10 + 35) = 90$ minutes, and the maximum circulation time of a train is $2 \times (25 + 40) = 130$ minutes.

Thus, in order to operate the line twice per hour takes at least $2 \times 90 = 180$ minutes and at most $2 \times 130 = 260$ minutes. Since each train composition provides 60 minutes

per hour, it follows that in each feasible timetable at least $\lceil 180/60 \rceil = 3$ and at most $\lfloor 260/60 \rfloor = 4$ train compositions are used for operating this line twice per hour.

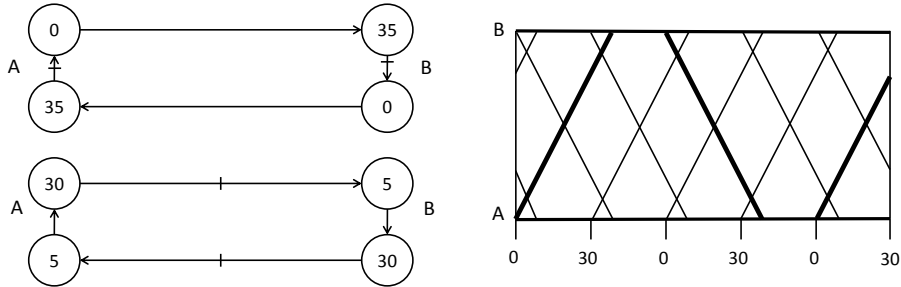


Figure 2: A periodic timetable that needs 4 train compositions

Figure 2 shows a periodic timetable in which the rolling stock connections in stations A and B are such that there are two separate circulations, needing $2 \times 2 = 4$ train compositions. The left part shows the constraint graph with departure and arrival times. The right part shows the corresponding time-space diagram. The short transversal lines in the left part of the figure indicate where the timetable crosses the end of the hour. For example, the arcs in the lower circulation starting at $t = 30$ and ending at $t = 5$ cross the end of the hour. The total number of these transversal lines indicates the required number of train compositions. The right part of the figure indicates the circulation of a single train composition with bold lines. In this timetable all return times are 25 minutes.

Note that also similar timetables with this structure of two separate rolling stock circulations exist with different running and return times. In each separate circulation, the lower bound on the sum of the process times equals $2 \times 35 + 2 \times 10 = 90$ minutes, and the upper bound equals $2 \times 40 + 2 \times 25 = 130$ minutes. Thus operating one circulation requires at least $\lceil 90/60 \rceil = 2$ and at most $\lfloor 130/60 \rfloor = 2$ train compositions: each feasible solution has two train compositions, running times between 35 and 40 minutes in each direction, and return times between 20 and 25 minutes at each endpoint, so that the sum of the running times and the return times equals 120 minutes. Each timetable with this structure that is operated twice per hour requires $2 \times 2 = 4$ train compositions.

Figure 3 shows a timetable with a combined rolling stock circulation that needs only 3 train compositions. Here the return times at the endpoints are just 10 minutes. This timetable with 3 train compositions is more or less unique, as will be demonstrated next.

The dashed arc in the graph in the left part of Figure 3 represents a *synchronization* constraint, specifying that the departures from station B should be 30 minutes apart. Note that the graph contains two cycles consisting of the synchronization arc, two running time arcs and two return time arcs. The lower bound on the sum of the process times on these cycles equals $30 + 2 \times 35 + 2 \times 10 = 120$ minutes, and the upper bound equals $30 + 2 \times 40 + 2 \times 25 = 160$ minutes. Since Proposition 2.1 implies that the sum of the process times along each cycle must be a multiple of 60 minutes, the sum of the process times equals 120 minutes. Thus the running times and the return times are minimal.

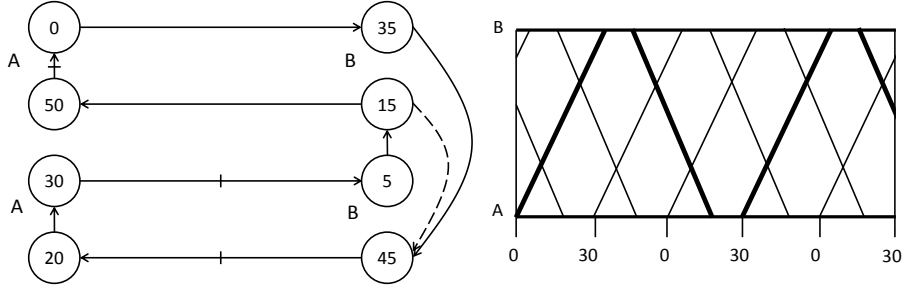


Figure 3: A periodic timetable that needs only 3 train compositions

That is, the running times are 35 minutes and the return times are 10 minutes. Thus the timetable in Figure 3 is basically the only timetable with a single combined rolling stock circulation, apart from trivial time shifts for all events.

This example shows that it is useful not to fix the rolling stock connections at both endpoints of the line a priori: one should have a flexible rolling stock connection at one endpoint of the line, and let the optimization model choose the rolling stock connection. Obviously, adding a flexible rolling stock connection also at the other endpoint of the line will not give any further reduction in the required number of train compositions.

If the arrival and departure times of the trains in station B are denoted by a_j and d_j , respectively, then the flexible connections in station B can be described as follows:

$$(d_1 - a_1 \in [10; 25]_T \wedge d_2 - a_2 \in [10; 25]_T) \vee (d_1 - a_2 \in [10; 25]_T \wedge d_2 - a_1 \in [10; 25]_T)$$

Furthermore, there are synchronization constraints $a_2 - a_1 \in [30; 30]_T$ and $d_2 - d_1 \in [30; 30]_T$ for distributing the arrivals at station B and the departures from station B evenly over the hour. Combining these sets of constraints, and using the result of Proposition 2.3, we get the following constraints:

$$d_1 - a_1 \in [10; 55]_T \wedge d_1 - a_1 \in [40; 85]_T \wedge d_1 - a_2 \in [10; 55]_T \wedge d_1 - a_2 \in [40; 85]_T \\ d_2 - a_1 \in [10; 55]_T \wedge d_2 - a_1 \in [40; 85]_T \wedge d_2 - a_2 \in [10; 55]_T \wedge d_2 - a_2 \in [40; 85]_T$$

These constraints allow the arriving trains in station B to connect to either of the departing trains. The next question is: how to count the number of train compositions needed in each of the two timetables? In Section 3.2 this number will be expressed in the cycle variables of certain cycles in the constraint graph. ■

Nielsen et al. (2006) consider the problem of minimizing the number of train compositions in the timetable of DSB S-train, the operator of local trains in the greater Copenhagen area. Here train lines are operated three times per hour with exactly 20 minutes between two consecutive departures in the same direction from each station. They consider these lines as having a periodic timetable with a period of 20 minutes. However, they recognize that it may be useful to merge the rolling stock circulations of two lines

having the same origin and the same destination. This indeed leads to savings of a few train compositions. Their model focuses on minimizing the return times of the trains at the endpoints of the lines, thereby assuming that the running times of the trains are fixed.

A similar approach is used by Liebchen (2004, 2006), who focuses on minimizing the number of train compositions required to operate the timetable of the Berlin Underground. He also focuses on minimizing the return times of the trains at the endpoints. Moreover, he shows that requiring symmetry for the timetable may lead to a higher minimum number of train compositions. In another paper, Borndörfer and Liebchen (2008) compute the minimum number of train compositions needed for operating a train line, both in case of a periodic timetable and in case of an aperiodic one. They show that, if the timetable is operated long enough, then the minimum number of required train compositions in the periodic case is the same as in the aperiodic case.

3.2 Counting the number of train compositions

In this section we express the number of train compositions needed to operate the timetable in terms of the cycle variables of certain cycles of the constraint graph, also in the case of flexible connections. This expression can be used then in a MIP model to minimize the number of train compositions needed to operate the timetable.

We assume that there are two trains per period from station A to station B and vice versa. The two trains per period from station A to station B are called t_1 and t_2 , and the trains in the reverse direction are called \bar{t}_1 and \bar{t}_2 .

We assume that in station A the trains turn on themselves. That is, in station A arriving train \bar{t}_j returns as departing train t_j ($j = 1, 2$). We consider the following two rolling stock circulations, based on the rolling stock connections in station B:

- (i): In station B, arriving train t_j returns as departing train \bar{t}_j ($j = 1, 2$).
- (ii): In station B, arriving train t_j returns as departing train \bar{t}_{3-j} ($j = 1, 2$).

In the following, a_j denotes the arrival time of train t_j at station B, and d_j denotes the departure time of train \bar{t}_j from station B ($j = 1, 2$). The arrivals of trains t_j in station B are related by a synchronization constraint with time window $[\underline{s}; \bar{s}]$, with $\underline{s} \leq \frac{T}{2} \leq \bar{s}$, and the same holds for the departures of trains \bar{t}_j from station B. Both turn around time windows in station B are equal to $[\underline{c}; \bar{c}]$.

In this section we assume that $\underline{s} + \bar{s} = T$, and that $\bar{c} < \underline{s}$. The latter implies that, for a fixed timetable, only one of the circulations (i) and (ii) is feasible, as can be seen easily. Indeed, $\bar{c} < \underline{s}$ implies that each train has departed from station B before the next one arrives, so that at any point in time there is at most one train in station B. Note that these assumptions also imply that $\bar{c} + \bar{s} < T$.

Figure 4 illustrates the situation in a constraint graph. The dashed arc from node d_1 to node d_2 is a synchronization arc. The four arcs from an arrival node to a departure node are connection arcs, describing the connections of the arrivals to the departures. In the left part of Figure 4, cycle C_j describes the circulation of the trains t_j and \bar{t}_j in circulation (i) ($j = 1, 2$). Cycles C_4 and C_5 play a role in the further analysis: cycle C_4 consists of

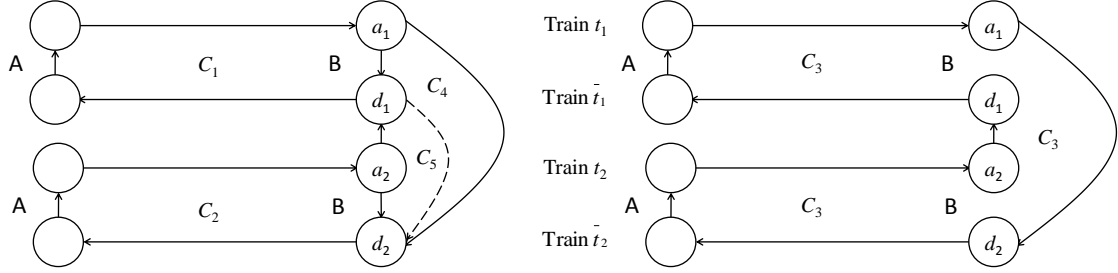


Figure 4: Constraint graph for trains t_j and \bar{t}_j ($j = 1, 2$); Right: Cycle C_3

the nodes a_1 , d_1 , and d_2 and the three arcs in between. Cycle C_5 consists of the nodes a_2 , d_1 , and d_2 and the three arcs in between.

In the right part of Figure 4, cycle C_3 shows the circulation of the trains t_j and \bar{t}_j in circulation (ii). Now the choice between the circulations (i) and (ii) can be modeled by the following constraints:

$$(d_1 - a_1 \in [\underline{c}; \bar{c}]_T \wedge d_2 - a_2 \in [\underline{c}; \bar{c}]_T) \vee (d_1 - a_2 \in [\underline{c}; \bar{c}]_T \wedge d_2 - a_1 \in [\underline{c}; \bar{c}]_T) \quad (5)$$

Using the synchronization constraints $a_2 - a_1 \in [\underline{s}, \bar{s}]$ and $d_2 - d_1 \in [\underline{s}, \bar{s}]$ and the fact that, according to Proposition 2.3, in a PESP model disjunctive constraints can be replaced by conjunctive constraints, constraint (5) is equivalent to constraints (6) - (9).

$$d_1 - a_1 \in [\underline{c}; \bar{c} + \bar{s}]_T \wedge d_1 - a_1 \in [\underline{c} + \underline{s}; \bar{c} + T]_T \quad (6)$$

$$d_1 - a_2 \in [\underline{c}; \bar{c} + \bar{s}]_T \wedge d_1 - a_2 \in [\underline{c} + \underline{s}; \bar{c} + T]_T \quad (7)$$

$$d_2 - a_1 \in [\underline{c}; \bar{c} + \bar{s}]_T \wedge d_2 - a_1 \in [\underline{c} + \underline{s}; \bar{c} + T]_T \quad (8)$$

$$d_2 - a_2 \in [\underline{c}; \bar{c} + \bar{s}]_T \wedge d_2 - a_2 \in [\underline{c} + \underline{s}; \bar{c} + T]_T \quad (9)$$

Because of the constraints (6) - (9), there are actually two arcs and two associated process time variables $x_{j,k}$ and $\bar{x}_{j,k}$ between each pair of nodes a_j and d_k in Figure 4 ($j = 1, 2$ and $k = 1, 2$). These process time variables have time windows $x_{j,k} \in [\underline{c}; \bar{c} + \bar{s}]$ and $\bar{x}_{j,k} \in [\underline{c} + \underline{s}; \bar{c} + T]$. This also means that the graph in Figure 4 contains more cycles than it seems at first sight. For example, cycle C_1 consists in fact of two cycles, since there are two arcs between nodes a_1 and d_1 .

For the choice between the two circulations (i) and (ii), however, we only need to consider the cycles that contain the process time variables $x_{j,k}$: these variables lie in the time windows $[\underline{c}; \bar{c} + \bar{s}]$, so they represent the actual turn around times. Note that the other process time variables $\bar{x}_{j,k}$ are still needed for correctly modeling the choice between the two circulations (i) and (ii).

For $j = 1, \dots, 5$, let q_j be the integer cycle variable of cycle C_j . Then the number of train compositions needed to operate circulation (i) equals $q_1 + q_2$. Similarly, the number of train compositions for circulation (ii) equals q_3 .

Next, consider the cycles C_4 and C_5 in Figure 4, and direct these such that the synchronization arc appears forwardly in both cycles. Then both cycles consist of the forward synchronization arc with time window $[\underline{s}, \bar{s}]$, and of a forward and a backward turn around arc. Recall from the above that we consider these turn around arcs as corresponding to process time variables x_{jk} with time windows $[\underline{c}; \bar{c} + \bar{s}]$. From Proposition 2.2 we obtain the following bounds for the integer variables q_4 and q_5 :

$$\left\lceil \frac{\underline{c} + \underline{s} - (\bar{c} + \bar{s})}{T} \right\rceil \leq q_4, q_5 \leq \left\lfloor \frac{(\bar{c} + \bar{s}) + \bar{s} - \underline{c}}{T} \right\rfloor \quad (10)$$

Since, by assumption, $\underline{s} + \bar{s} = T$ and $\bar{c} < \underline{s}$, we find that $q_4, q_5 \in \{0, 1\}$. The cycle periodicity constraint for cycle C_4 is $x_{a_1, d_1} + x_{d_1, d_2} - x_{a_1, d_2} = T \times q_4$. Now we know that the cyclic sequence $a_1 \rightarrow d_1 \rightarrow d_2$ occurs, if and only if $x_{a_1, d_1} + x_{d_1, d_2} + x_{d_2, a_1} = T$. The latter is equivalent to $x_{a_1, d_1} + x_{d_1, d_2} + (T - x_{a_1, d_2}) = T$ or $x_{a_1, d_1} + x_{d_1, d_2} - x_{a_1, d_2} = 0$. Thus the cyclic sequence $a_1 \rightarrow d_1 \rightarrow d_2$ occurs if and only if $q_4 = 0$. A similar result holds for q_5 . This can be summarized as follows:

$$q_4 = \begin{cases} 1 & \text{if and only if the cyclic sequence } a_1 \rightarrow d_2 \rightarrow d_1 \text{ occurs,} \\ 0 & \text{if and only if the cyclic sequence } a_1 \rightarrow d_1 \rightarrow d_2 \text{ occurs.} \end{cases}$$

$$q_5 = \begin{cases} 1 & \text{if and only if the cyclic sequence } a_2 \rightarrow d_2 \rightarrow d_1 \text{ occurs,} \\ 0 & \text{if and only if the cyclic sequence } a_2 \rightarrow d_1 \rightarrow d_2 \text{ occurs.} \end{cases}$$

Furthermore, we know that either $x_{a_1, d_1} \in [\underline{c}; \bar{c}]$ and $x_{a_1, d_2} \in [\underline{c} + \underline{s}; \bar{c} + \bar{s}]$, or that $x_{a_1, d_1} \in [\underline{c} + \underline{s}; \bar{c} + \bar{s}]$ and $x_{a_1, d_2} \in [\underline{c}; \bar{c}]$. Obviously, the latter case corresponds to $q_4 = 1$, and the former to $q_4 = 0$. Since a similar result holds for cycle C_5 , the above relations can be restated as follows:

$$q_4 = \begin{cases} 1 & \text{if and only if circulation (ii) is operated,} \\ 0 & \text{if and only if circulation (i) is operated.} \end{cases}$$

$$q_5 = \begin{cases} 1 & \text{if and only if circulation (i) is operated,} \\ 0 & \text{if and only if circulation (ii) is operated.} \end{cases}$$

Thus $q_4 + q_5 = 1$. This means that we can use either q_4 or q_5 as a binary variable to indicate which of the circulations (i) and (ii) is operated. Arbitrarily choosing q_5 for this role, the number N of train compositions required for operating the timetable can be expressed as follows:

$$N = q_5 \times (q_1 + q_2) + (1 - q_5) \times q_3 \quad (11)$$

This expression states that, if circulation (i) is operated, then the number of required train compositions equals $q_1 + q_2$, and otherwise this number equals q_3 . Next, we rewrite the non-linear expression (11) into a linear one. To that end, let cycle C_6 be the clockwise directed cycle formed by the four connection arcs from each node a_j to each node d_k ($j = 1, 2$ and $k = 1, 2$) in Figure 4, and let q_6 be its corresponding integer cycle variable. One can check that $q_3 = q_1 + q_2 + q_6$. That means that (11) is equivalent to

$$N = q_5 \times (q_1 + q_2) + (1 - q_5) \times (q_1 + q_2 + q_6) = q_1 + q_2 + (1 - q_5) \times q_6 \quad (12)$$

Furthermore, $q_6 = q_5 - q_4$. Together with the earlier found equality $q_4 + q_5 = 1$, this gives $q_6 = 2q_5 - 1$. Substituting this into the above expression, we obtain

$$N = q_1 + q_2 + (1 - q_5) \times q_6 = q_1 + q_2 + (1 - q_5) \times (2q_5 - 1) \quad (13)$$

In the latter expression, the term $(1 - q_5) \times (2q_5 - 1)$ has the value -1 whenever $q_5 = 0$, and it has the value 0 whenever $q_5 = 1$; that is, $(1 - q_5) \times (2q_5 - 1) = -(1 - q_5)$. Thus for the considered situation in Figure 4, we have obtained the following result:

Proposition 3.1. *If in station B there are two arriving and departing trains, $\bar{c} < \underline{s} \leq \frac{T}{2}$, and $\underline{s} + \bar{s} = T$, then the number N of train compositions required to operate the timetable described by the graph in Figure 4 equals $q_1 + q_2 + q_5 - 1$.*

If the trains belong to the same line, then one may assume that the timetable of the trains t_1 and \bar{t}_1 is very similar to that of the trains t_2 and \bar{t}_2 . In that case we have $q_1 = q_2$, and hence the expression in Proposition 3.1 simplifies to $2q_1 + q_5 - 1$.

Proposition 3.1 implies that the required number of train compositions equals $q_1 + q_2$ if circulation (i) is operated and $q_1 + q_2 - 1$ otherwise. One may be tempted to conclude from this that circulation (ii) is always better than circulation (i). However, that is not the case. One can only conclude from this that, if both circulations are feasible for a fixed timetable, then circulation (ii) is better than circulation (i). However, since $\bar{c} < \underline{s}$, this situation cannot occur. In the general case, it is possible that both circulations are feasible. In fact, it is not difficult to create instances for the following situations:

- (a): Circulation (i) is feasible and circulation (ii) is not.
- (b): Both circulations are feasible, and circulation (i) requires less compositions.
- (c): Both circulations are feasible, and circulation (ii) requires less compositions.
- (d): Circulation (ii) is feasible and circulation (i) is not.

With respect to cases (b) and (c), one can start with a circulation of type (i) with $q_1 + q_2$ compositions, and variable $q_5 = 1$. Then, when changing the circulation into a circulation of type (ii), variable q_5 changes from 1 to 0. But at the same time, the variables q_1 and q_2 may change as well. The net result in the expression $q_1 + q_2 + q_5 - 1$ may be that the circulation of type (ii) requires more or less compositions than the circulation of type (i), or the same number of compositions as the circulation of type (i).

3.3 Three and more trains per period

The analysis of the previous section can also be extended to the case with three or more trains per period between stations A and B. A description is provided in this section.

3.3.1 Three trains per period

In this section, trains t_1 , t_2 and t_3 run from station A to station B, and trains \bar{t}_1 , \bar{t}_2 and \bar{t}_3 run from station B to station A. The arrival and departure times of the trains in station B are denoted by a_j and d_j , respectively ($j = 1, 2, 3$). We assume that trains do

not overtake each other between their departure from station B and their next arrival in station B. That is, we use the following FIFO assumption:

$$\begin{aligned} &\text{In station B, the departures of the trains take place in the cyclic order } d_1 \rightarrow d_2 \rightarrow d_3 \\ &\text{if and only if the arrivals take place in the cyclic order } a_1 \rightarrow a_2 \rightarrow a_3. \end{aligned} \quad (14)$$

Note that this condition is automatically satisfied in the case of two trains per period. As in the previous section, in station A train \bar{t}_j returns as train t_j ($j = 1, 2, 3$). The rolling stock circulation depends on the rolling stock connections in station B. We will prove that, under a reasonable condition, there are only three circulations, namely:

- (iii): In station B, train t_j returns as train \bar{t}_j ($j = 1, 2, 3$),
- (iv): In station B, train t_j returns as train $\bar{t}_{(j+1) \bmod 3+1}$ ($j = 1, 2, 3$),
- (v): In station B, train t_j returns as train $\bar{t}_{j \bmod 3+1}$ ($j = 1, 2, 3$).

In any other circulation than (iii), (iv) and (v), the three arriving trains would be split into a subset of two trains and one single train. In station B, the single arriving train t_j returns as departing train \bar{t}_j , and the other two arriving trains turn on each other. However, if the synchronization constraints are strong enough, then such a timetable is not feasible, as is shown in Proposition 3.2.

Proposition 3.2. *If in station B there are three arriving and three departing trains, $\bar{c} < \underline{s}$, and the FIFO condition (14) holds, then any other circulation than circulations (iii), (iv), or (v) is infeasible.*

Note that Proposition 3.2 does not state that any of the circulations (iii), (iv), or (v) is feasible. It only states that any other circulation is infeasible.

Proof. Suppose the circulation is different from circulations (iii), (iv), or (v). Then there is one arriving train t_j in station B that departs as train \bar{t}_j , and the other two arriving trains turn on each other. Without loss of generality, train t_1 departs as train \bar{t}_1 , $a_1 = 0$, $d_1 \in [\underline{c}, \bar{c}]$, $a_2 \in [\underline{s}, \bar{s}]$, and $a_3 \in [T - \bar{s}, T - \underline{s}]$.

Then, because of the FIFO condition (14), $d_2 \in [\underline{c} + \underline{s}, \bar{c} + \bar{s}]$, and $d_3 \in [T - \bar{s} + \underline{c}, T - \underline{s} + \bar{c}]$. Note that $T - \underline{s} + \bar{c} < T$. As a consequence, we have $d_3 - a_2 = (d_3 - d_2) + (d_2 - a_2) \geq \underline{s} + \underline{c} > \bar{c}$. Thus train t_3 cannot return as train \bar{t}_2 in station B. It follows that train t_j returns as train \bar{t}_j ($j = 1, 2, 3$). Hence, contrary to the assumption, we have circulation (iii). ■

Circulation (iii), together with some additional arcs and cycles, is shown in Figure 5. In this figure, cycle C_j represents the circulation of trains t_j and \bar{t}_j ($j = 1, 2, 3$). In order to keep the figure as simple as possible, it has been split into two parts. Note that in the right part of the figure, trains t_2 and \bar{t}_2 were swapped with trains t_3 and \bar{t}_3 .

Circulations (iv) and (v) are shown in Figure 6. Cycle C_4 belongs to circulations (iv), and cycle C_5 belongs to circulation (v).

As before, the synchronization constraints are indicated by dashed arcs in Figure 5. The synchronization constraints are twofold, and have intervals $[\underline{s}; \bar{s}]$ or $[T - \bar{s}; T - \underline{s}]$. We

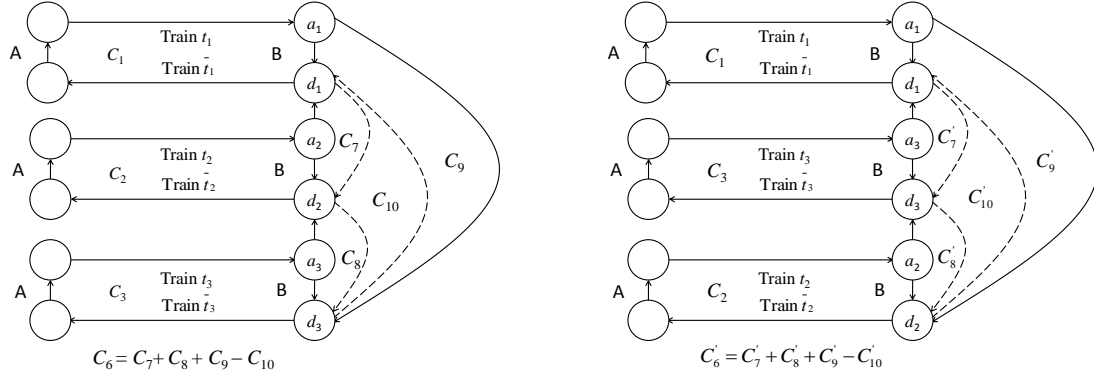


Figure 5: Constraint graph for trains t_j and \bar{t}_j ($j = 1, 2, 3$)

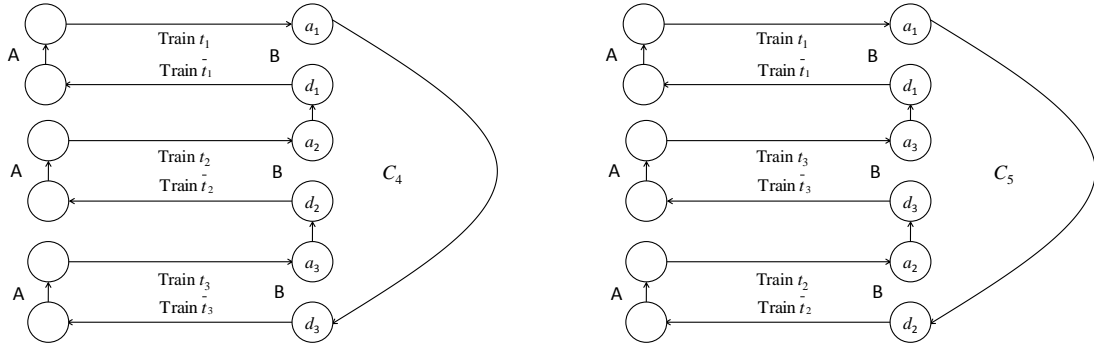


Figure 6: Cycles C_4 and C_5 belonging to circulations (iv) and (v) , respectively

assume $\underline{s} \leq \frac{T}{3} \leq \bar{s}$. The connection constraints are threefold, and have time intervals $[\underline{c}; \bar{c}]$, $[\underline{c} + \underline{s}; \bar{c} + \bar{s}]$, or $[T - \bar{s} + \underline{c}; T - \underline{s} + \bar{c}]$. Again we assume $\bar{c} < \underline{s}$, so that there is at most one train at the same time in station B.

Cycle C_6 is the clockwise oriented cycle consisting of the six connection arcs in the left part of Figure 5 from the arrival nodes to the departure nodes. Cycles C_7 , C_8 , and C_9 are the cycles consisting of one forward connection arc, one backward connection arc, and one synchronization arc in the left part of Figure 5. These cycles are oriented in such a way that the synchronization arc is directed forwardly. Cycle C_{10} is the cycle consisting of the three forward synchronization arcs. Cycles C'_6 , C'_7 , C'_8 , C'_9 and C'_{10} are similar cycles in the right part of Figure 5.

Let q_j be the integer cycle variable corresponding to cycle C_j ($j = 1, 2, \dots, 9$), and let q'_j be the integer cycle variable corresponding to cycle C'_j ($j = 7, \dots, 10$). In the following, we do the calculations with the cycles in the left part of Figure 5, but the same results hold for the cycles in the right part of the figure.

From the structure of the cycles, it follows that $q_4 = q_1 + q_2 + q_3 + q_6$, and $q_6 = q_7 + q_8 + q_9 - q_{10}$, and $q_{10} = 1$ or 2 . Furthermore, the FIFO condition (14) implies that

$q_7 = q_8 = q_9$. Now using the result of Proposition 2.2, we get

$$\left\lceil \frac{\underline{c} + \underline{s} - (T - \underline{s} + \bar{c})}{T} \right\rceil \leq q_7, q_8, q_9 \leq \left\lfloor \frac{(T - \underline{s} + \bar{c}) + (T - \underline{s}) - \underline{c}}{T} \right\rfloor \quad (15)$$

Using the fact that $\bar{c} < \underline{s}$ and that $\underline{s} \leq \frac{T}{3}$, it follows that $0 \leq q_j \leq 1$ ($j = 7, 8, 9$), and in the same way it follows that $0 \leq q'_j \leq 1$ ($j = 7, 8, 9$). By the same arguments as in Section 3.2, we get the following result:

$$q_9 = \begin{cases} 1 & \text{if and only if the cyclic sequence } a_1 \rightarrow d_1 \rightarrow d_3 \text{ occurs,} \\ 0 & \text{if and only if the cyclic sequence } a_1 \rightarrow d_3 \rightarrow d_1 \text{ occurs.} \end{cases}$$

$$q'_9 = \begin{cases} 1 & \text{if and only if the cyclic sequence } a_1 \rightarrow d_1 \rightarrow d_2 \text{ occurs,} \\ 0 & \text{if and only if the cyclic sequence } a_1 \rightarrow d_2 \rightarrow d_1 \text{ occurs.} \end{cases}$$

Thus circulation (iii) is operated if $q_9 = q'_9 = 1$, circulation (iv) is operated if $q_9 = 0$ and $q'_9 = 1$, and circulation (v) is operated if $q_9 = 1$ and $q'_9 = 0$. If $q_9 = q'_9 = 0$, then either of the two circulations (iv) or (v) can be operated, depending on the values of the synchronization arcs. Thus, if the binary variables R_3, R_4 , and R_5 are introduced to represent which of the three circulations is operated, we get:

$$R_3 + R_4 + R_5 = 1 \quad \text{and} \quad R_3 \leq q_9 \quad \text{and} \quad R_3 \leq q'_9 \quad \text{and} \quad R_4 \leq 1 - q_9 \quad \text{and} \quad R_5 \leq 1 - q'_9 \quad (16)$$

Now the required number N of train compositions equals $q_1 + q_2 + q_3$ in circulation (iii), q_4 in circulation (iv), and q_5 in circulation (v). This number N can be expressed as

$$N = R_3 \times (q_1 + q_2 + q_3) + R_4 \times q_4 + R_5 \times q_5 =$$

$$R_3 \times (q_1 + q_2 + q_3) + R_4 \times (q_1 + q_2 + q_3 + q_6) + R_5 \times (q_1 + q_2 + q_3 + q'_6).$$

Using $R_3 + R_4 + R_5 = 1$, and $q_6 = q_7 + q_8 + q_9 - q_{10}$, and $q'_6 = q'_7 + q'_8 + q'_9 - q'_{10}$, and $q_7 = q_8 = q_9$, and $q'_7 = q'_8 = q'_9$, we get:

$$N = q_1 + q_2 + q_3 + R_4 \times (3q_9 - q_{10}) + R_5 \times (3q'_9 - q'_{10}).$$

Since $R_4 = 0$ if $q_9 = 1$ and $R_5 = 0$ if $q'_9 = 1$, we get:

$$N = q_1 + q_2 + q_3 - R_4 \times q_{10} - R_5 \times q'_{10}.$$

Now the problem is to linearize the expressions $R_4 \times q_{10}$ and $R_5 \times q'_{10}$. We denote the former by E_4 and the latter by E_5 . Then we know that $E_4 = 0$ if $R_4 = 0$, and $E_4 = q_{10}$ if $R_4 = 1$. Since $q_{10} = 1$ or 2 , we have $0 \leq E_4 \leq 2$. Similar results hold for E_5 . Thus E_4 and E_5 can be described as follows:

$$0 \leq E_4 \leq 2R_4 \quad \text{and} \quad 0 \leq q_{10} - E_4 \leq 2(1 - R_4) \quad (17)$$

$$0 \leq E_5 \leq 2R_5 \quad \text{and} \quad 0 \leq q'_{10} - E_5 \leq 2(1 - R_5) \quad (18)$$

Indeed, the first pair of inequalities in (17) implies $E_4 = 0$ if $R_4 = 0$. The second pair of inequalities in (17) implies $E_4 = q_{10}$ if $R_4 = 1$. A similar result holds for E_5 . Altogether, we found the following result.

Proposition 3.3. *If in station B there are three arriving and departing trains, the FIFO condition (14) holds, and $\bar{c} < \underline{s} \leq \frac{T}{3}$, then the number of train compositions required to operate the timetable equals $q_1 + q_2 + q_3 - E_4 - E_5$, where E_4 and E_5 satisfy inequalities (17) and (18), respectively, and where R_3 , R_4 and R_5 satisfy inequalities (16).*

Note that the right and the left parts of Figure 5 only differ in the order of trains t_2 and t_3 and of trains \bar{t}_2 and \bar{t}_3 . If these trains belong to the same line, then trains t_2 and t_3 can be considered as interchangeable, and the same holds for trains \bar{t}_2 and \bar{t}_3 . Thus in that case we may assume without loss of generality that the departures of the trains take place in the cyclic order $d_1 \rightarrow d_2 \rightarrow d_3$. That is, $q_{10} = 1$ and $q'_{10} = 2$. The number of train compositions required to operate the timetable equals $q_1 + q_2 + q_3 - R_4 - 2R_5$ then.

3.3.2 More than three trains per period

If the connection and synchronization constraints are strong enough, then the analysis for three arriving and departing trains may also be applied to the case of any *prime* number p of arriving and departing trains. Indeed, if $\bar{c} < \underline{s}$ and $\underline{s} \leq \frac{T}{p} \leq \bar{s}$, then the circulations that may exist in that case are comparable to the circulations for the case with three trains. That is, there are either p separate circulations where arriving train t_j returns in station B as departing train \bar{t}_j ($j = 1, \dots, p$), or there is just one single circulation where arriving train t_j returns in station B as departing train $\bar{t}_{(j+k) \bmod p+1}$ ($j = 1, \dots, p$). The latter type of circulation exists for $k = 0, 1, \dots, p - 2$.

If p is a prime number, then in any other circulation than the two types sketched above, the trains must be split into two or more subsets of *unequal* size that each have their own circulation. However, if the connection and synchronization constraints are strong enough, then this is infeasible, as can be proved in the same way as Proposition 3.2.

For a *composite* number of arriving and departing trains, the situation is different. Then the trains can be split into two or more subsets of the *same* size that each have their own separate circulation. For example, if there are four arriving and departing trains, then there may be one circulation including all four arriving and departing trains, two circulations including two arriving and departing trains each, or four separate circulations including one arriving and one departing train each. In practical situations usually only at most two of these circulations are feasible. Note that frequencies of four or more trains on the same line hardly ever occur in practice.

4 Passenger connections

In this section we show that it is useful to model passenger connections also in a flexible way, and to let the actual passenger connections between the trains be chosen by the optimization model. In this sense, passenger connections are comparable to rolling stock connections, but there are also differences. The main difference is that the rolling stock connections have direct implications for the number of train compositions, and thereby for the complete rolling stock circulation in the timetable. Passenger connections act more locally: they just put additional constraints on the local arrival and departure times of

the trains in the stations, so that passengers can easily transfer from one train to another there. These constraints should be effective, but they should also be as much as possible flexible, thereby not constraining the timetable more than necessary. In this section we first go back to the example in Section 3.1.

4.1 Example

This example is an extension of the example in Section 3.1. In addition to the two trains per hour between stations A and B, there is also a train between stations B and C, which is operated once per hour in both directions. The running time between stations B and C is 45 minutes in both directions. This train has a return time in stations B and C between 10 and 25 minutes. In station B, there must be a passenger connection with a duration between 5 and 15 minutes from an arriving train from station A to the departing train to station C, and also in the reverse direction.

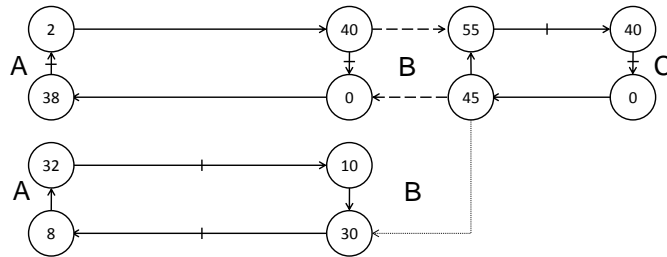


Figure 7: A timetable with separate rolling stock circulations between stations A and B, and passenger connections in station B with one of these circulations

Figure 7 shows the timetable that is obtained if two separate rolling stock circulations are operated between stations A and B, and the trains to and from station C are connected in both directions with the same circulation. The dashed arcs indicate these passenger connections. In this case, the passenger connection times in station B are 15 minutes in both directions. The thin dotted arc shows that the passenger connection from station C to station A could have been chosen also in a different way, namely from the train arriving from station C to the train in the other circulation departing to station A.

Note that in Figure 7 the running times between stations A and B have been stretched from 35 to 38 minutes, because otherwise the return times in station A would become too long, or the return time in station B of the train from and to station C would become too short. Note further that, without loss of optimality, one passenger connection, in this case from station A to station C, may be fixed a priori.

Figure 8 shows the timetable that is obtained if separate rolling stock circulations are operated between stations A and B, and if the trains to and from station C are connected with both circulations. Now the passenger connection times in station B are only 5 minutes in both directions, without stretching the running times between stations A and B. Thus this timetable is better than the timetable shown in Figure 7.

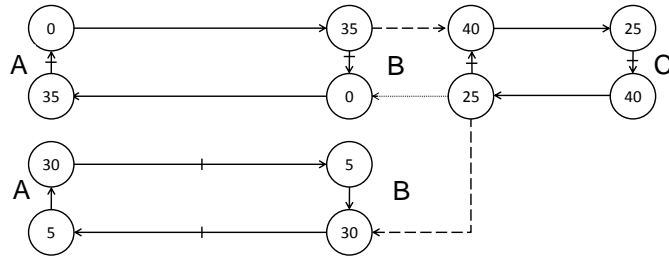


Figure 8: A timetable with separate rolling stock circulations between stations A and B, and passenger connections in station B with both circulations

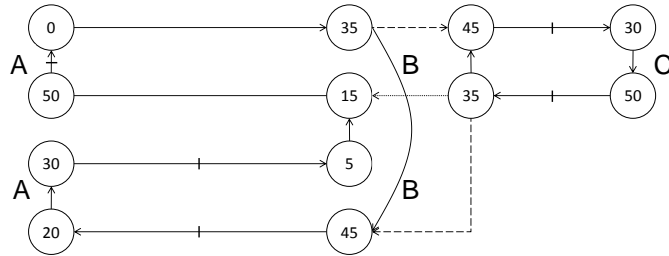


Figure 9: A timetable with a combined rolling stock circulation between station A and B, and flexible passenger connections in station B

Figure 9 shows the timetable that is obtained if the trains between stations A and B are operated with a combined rolling stock circulation, and the train from station C to station B is connected with one of the trains departing from station B to station A in an optimal way. Now the passenger connection times in station B are 10 minutes in both directions. However, as we have seen before, this rolling stock circulation between stations A and B saves one train composition.

Finally, Figure 10 shows that a flexible passenger connection in station B can also be achieved by having flexible rolling stock connections for the trains between stations A and B at both endpoints of this line. The timetable shown in this figure is virtually the same as the one shown in Figure 8. The flexible rolling stock connections at both endpoints of the line make the two trains from station A to station B completely interchangeable, and the same holds for the two trains from station B to station A. A drawback of this method is that it makes counting the required number of train compositions more complex.

The foregoing examples showed that fixing either rolling stock connections or passenger connections a priori may lead to a suboptimal solution. The examples also showed that there is a trade-off between short rolling stock connections and short passenger connections: a timetable that is optimal for the passengers may require more than the minimum number of train compositions, and vice versa.

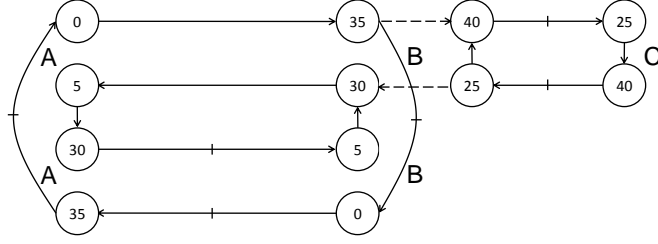


Figure 10: A timetable with flexible rolling stock connections in stations A and B, and fixed passenger connections in station B

4.2 Model

Let the arrival and departure times in station B of the trains from and to station A be denoted by a_1 , a_2 , d_1 , and d_2 , respectively, and let the arrival and departure time in station B of the trains from and to station C be denoted by a and d . Without loss of generality, we may assume that in station B the arriving train t_1 from station A connects to the departing train to station C. This connection can be described by $d - a_1 \in [\underline{c}; \bar{c}]_T$.

The flexible connection in station B in the other direction is represented as follows:

$$d_1 - a \in [\underline{c}; \bar{c}]_T \quad \vee \quad d_2 - a \in [\underline{c}; \bar{c}]_T \quad (19)$$

As before, using the synchronization constraint $d_2 - d_1 \in [\underline{s}; \bar{s}]_T$, constraint (19) can be rewritten as follows:

$$d_1 - a \in [\underline{c}; \bar{c} + \bar{s}]_T \quad \wedge \quad d_1 - a \in [\underline{c} + \underline{s}; \bar{c} + T]_T \quad (20)$$

$$d_2 - a \in [\underline{c}; \bar{c} + \bar{s}]_T \quad \wedge \quad d_2 - a \in [\underline{c} + \underline{s}; \bar{c} + T]_T \quad (21)$$

Constraints (20) and (21) in combination with the synchronization constraint $d_2 - d_1 \in [\underline{s}; \bar{s}]_T$ guarantee that constraint (19) is satisfied, i.e. in station B the arriving train from station C has a passenger connection with one of the two departing trains to station A.

The minimization of the connection times can be expressed as follows:

$$\min \quad w \times ((d_1 - a + T \times p_1) + (d_2 - a + T \times p_2)) + 2\bar{w} \times (d - a_1 + T \times p)$$

Here the integer variables p_j are such that $d_j - a + T \times p_j \in [\underline{c}, \bar{c} + \bar{s}]$ ($j = 1, 2$), and the integer variable p is such that $d - a_1 + T \times p \in [\underline{c}, \bar{c}]$. Furthermore, the parameters w and \bar{w} are weights representing the involved numbers of passengers.

Preferably, one would like to minimize only the “real” connection time, and not the time between the arrival of the “other” train from station A and the departure time of the train to station C. However, the fact that both connection times are included in the minimization is almost equivalent, since the two connection times differ by a value in $[\underline{s}, \bar{s}]$. Since this is usually a narrow interval, the two terms differ more or less by a constant term. In order to compensate for this double counting, the parameter \bar{w} is multiplied by 2.

If minimizing the number of train compositions needed for operating the timetable is also part of the objective, then this can be included in the model as in Section 3.2.

5 Generalization

In the previous sections we used the fact that the rolling stock and passenger connections of a number of trains that arrive in a synchronized way to a number of trains that depart in a synchronized way can be expressed in terms of a number of PESP constraints. In this section this method is described in a more general way.

In particular, for the case that there are n arriving trains and m departing trains in a station, we describe under which conditions it can be guaranteed by PESP constraints that there will be $G = \gcd(m, n)$ connections between the arriving and the departing trains. The latter is the best one can hope for if the arriving and departing trains are more or less evenly spread over the period $[0, T)$. Note that we allow slight deviations from an exact even spread of the trains over the period.

Thus in a certain station we consider n arriving and m departing trains. Furthermore, let $0 \leq a_1 < a_2 < \dots < a_n < T$ be integers, and let the periodic set $A = \{a_1, a_2, \dots, a_n\}_T$ denote the arrival times of the arriving trains. The arrival times are evenly spread within one period $[0, T)$ in the following sense:

$$a_k - a_l \in \frac{(k-l)T}{n} + [-\delta; \delta]_T \quad \forall k, l = 1, \dots, n. \quad (22)$$

The *tolerance* δ satisfies $0 \leq 2n\delta < T$, so that the periodized intervals in condition (22) are disjoint. The periodic set of departure times of the trains $D = \{d_1, d_2, \dots, d_m\}_T$ satisfies

$$d_k - d_l \in \frac{(k-l)T}{m} + [-\eta; \eta]_T \quad \forall k, l = 1, \dots, m, \quad (23)$$

where η is such that $0 \leq 2m\eta < T$. In words, the conditions (22) and (23) ensure, up to some positive tolerances (defined by δ and η), that the arrival and departure times are evenly spread in time with a specified frequency within one period (defined by n and m).

There are other formal conditions than condition 22 that can be used to represent the intuitive requirement of evenly distributing a number of event times over the period. For example, a less restrictive variant of condition (22) reads

$$a_k - a_{k-1} \in \frac{T}{n} + [-\delta; \delta]_T \quad \forall k = 1, \dots, n, \quad (24)$$

where $a_0 := a_n$. In the following, condition (22) defines periodic sets of class $\mathcal{A}_{n,\delta}$, and condition (24) defines periodic sets of class $\mathcal{B}_{n,\delta}$. Lemma 5.1 gives some relations between the classes $\mathcal{A}_{n,\delta}$ and $\mathcal{B}_{n,\delta}$. The proofs are straightforward and omitted therefore.

Lemma 5.1. *The following statements hold for all $\delta \geq 0$:*

$$\begin{aligned} \mathcal{A}_{n,\delta} &\subset \mathcal{B}_{n,\delta}, & n &\geq 1, \\ \mathcal{A}_{n,\delta} &= \mathcal{B}_{n,\delta}, & 1 &\leq n \leq 3, \\ \mathcal{A}_{n,\delta} &\neq \mathcal{B}_{n,\delta}, & n &\geq 4, \\ \mathcal{B}_{n,\delta} &\subset \mathcal{A}_{n,\delta} && \text{if and only if } \eta \geq \lfloor \frac{n}{2} \rfloor \delta. \end{aligned}$$

Lemma 5.1 shows that conditions (22) and (24) are very similar. In the remainder of this section, we will use periodic sets of class $\mathcal{A}_{n,\delta}$. However, mutatis mutandis, the results are also valid for sets of class $\mathcal{B}_{n,\delta}$.

The following constraint on the arriving and departing trains defines a “connection”.

(†) *There is an arriving train with arrival time a_k and a departing train with departure time d_l such that $d_l - a_k \in [\alpha; \alpha + \varepsilon]_T$.*

Here α is a positive value indicating the minimum time required for the connection, and the nonnegative value ε indicates the extent up to which the connection time between the arrival and the departure can be stretched.

This section reformulates a more general form of this connection constraint (†) in terms of PESP constraints. If there are n arriving trains and m departing trains, which are all evenly spread over the period, then the best one may hope for is that there are $\gcd(m, n)$ connections between these trains. Theorem 5.3 is the main result describing when this number of connections can be guaranteed. Proposition 5.4 shows the limitations to the scope of the theorem.

In the following we assume that $A = \{a_1, \dots, a_n\}$ is of class $\mathcal{A}_{n,\delta}$ and $D = \{d_1, \dots, d_m\}$ is of class $\mathcal{A}_{m,\eta}$. We also write $G = \gcd(m, n)$ and $L = \text{lcm}(m, n)$. Furthermore, in the remainder of this section, we assume that the non-negative parameters δ , η , and ε satisfy

$$3\delta + 3\eta + \varepsilon < \frac{T}{L}. \quad (25)$$

Next, we write $I_j(n) = ((j-1)\frac{G}{n}, j\frac{G}{n}]$. Then $\bigcup_{j=1}^G I_j(n) = (0, n]$ provides a disjoint union of the interval $(0, n]$. The following result of Lemma 5.2 is needed to prove Theorem 5.3.

Lemma 5.2. *Suppose that condition (25) is satisfied, let $A = \{a_1, \dots, a_n\}$ be of class $\mathcal{A}_{n,\delta}$, and let $D = \{d_1, \dots, d_m\}$ be of class $\mathcal{A}_{m,\eta}$. Then for each integer p with $1 \leq p \leq L-1$, there exist at most $G = \gcd(m, n)$ pairs (a_k, d_l) , such that*

$$d_l - a_k \in \frac{pT}{L} + [-\delta - \eta + \alpha, \delta + \eta + \alpha + \varepsilon]_T =: W_p$$

Proof. Assume that there exist $G+1$ pairs (k, l) such that $d_l - a_k \in W_p$. This implies that for some j with $1 \leq j \leq G$, there exist two distinct $k, k' \in I_j(n)$ and some $1 \leq l, l' \leq m$ such that $d_l - a_k$ and $d_{l'} - a_{k'} \in W_p$. On one hand, the latter implies that

$$(d_l - a_k) - (d_{l'} - a_{k'}) \in [-2\delta - 2\eta - \varepsilon, 2\delta + 2\eta + \varepsilon]_T. \quad (26)$$

On the other, since the arrival and departure times are of classes $\mathcal{A}_{n,\delta}$ and $\mathcal{A}_{m,\eta}$, we have

$$(d_l - a_k) - (d_{l'} - a_{k'}) = (d_l - d_{l'}) + (a_{k'} - a_k) \in \frac{(l-l')T}{m} + \frac{(k'-k)T}{n} + [-\delta - \eta; \delta + \eta]_T.$$

Using $mn = GL$, the latter set equals

$$\left((l-l')\frac{n}{G} + (k'-k)\frac{m}{G} \right) \times \frac{T}{L} + [-\delta - \eta; \delta + \eta]_T. \quad (27)$$

Observe that $(l-l')\frac{n}{G} + (k'-k)\frac{m}{G}$ is not a multiple of L . Indeed, if $(l-l')\frac{n}{G} + (k'-k)\frac{m}{G} = \sigma L = \sigma\frac{mn}{G}$, then $(k'-k)\frac{m}{G} = (\sigma m - l + l')\frac{n}{G}$. Since $\frac{m}{G}$ and $\frac{n}{G}$ do not have any common prime factors, this would imply that $\frac{n}{G}$ divides $(k'-k)$, contrary to the fact that $0 < |k'-k| < \frac{n}{G}$.

Thus in (27) the set $[-\delta - \eta; \delta + \eta]_T$ is translated by the first term in (27) over at least $\frac{T}{L}$ units in positive or negative direction. Since, by assumption (25), $\frac{T}{L} > 3\delta + 3\eta + \varepsilon$, the two sets in (26) and (27) have a non-empty intersection. This contradicts the fact that the number $(d_l - a_k) - (d_{l'} - a_{k'})$ is an element of both sets. \blacksquare

Theorem 5.3. *Suppose that condition (25) is satisfied, let $A = \{a_1, \dots, a_n\}$ be of class $\mathcal{A}_{n,\delta}$, and let $D = \{d_1, \dots, d_m\}$ be of class $\mathcal{A}_{m,\eta}$. Then the following two statements are equivalent:*

1. *There exists a permutation π on the set $\{1, \dots, G\}$ such that the following holds: For $j = 1, \dots, G$, there exists $k_j \in I_j(n)$ and $l_j \in I_{\pi(j)}(m)$, such that*

$$d_{l_j} - a_{k_j} \in [\alpha; \alpha + \varepsilon]_T.$$

2. *For all k and l with $1 \leq k \leq n$ and $1 \leq l \leq m$, it holds true that*

$$d_l - a_k \in \left(\bigcup_{p=1}^{L-1} \frac{pT}{L} + [-\delta - \eta + \alpha, \delta + \eta + \alpha + \varepsilon]_T \right) \cup [\alpha, \alpha + \varepsilon]_T.$$

Moreover, statement 1 implies statement 2 also if condition (25) is not satisfied.

Proof. In order to prove that the first statement implies the second one, let $1 \leq k \leq n$ and $1 \leq l \leq m$. Now we distinguish the following four cases.

- (i) In the case when $l = l_j$ and $k = k_j$ for some $1 \leq j \leq G$, we have $d_l - a_k \in [\alpha; \alpha + \varepsilon]_T$.
- (ii) In the case when $k = k_j$ and $l \neq l_j$ for some $1 \leq j \leq G$, we can write

$$d_l - a_k = (d_l - d_{l_j}) + (d_{l_j} - a_{k_j}) \in \frac{(l - l_j)T}{m} + [-\eta + \alpha; \eta + \alpha + \varepsilon]_T.$$

Using $mn = LG$, we can write

$$\frac{(l - l_j)T}{m} = \frac{(l - l_j)n}{G} \times \frac{T}{L}.$$

Since $0 < |l - l_j| < m$, we find that $l - l_j$ is not a multiple of m . Hence $\frac{(l-l_j)n}{G}$ is not a multiple of L . Thus

$$\frac{(l - l_j)n}{G} \times \frac{T}{L} = (p + qL) \times \frac{T}{L} = \frac{pT}{L} + qT,$$

with integers q and p satisfying $1 \leq p \leq L - 1$.

- (iii) The case when $l = l_j$ and $k \neq k_j$ for some $1 \leq j \leq G$ is similar to case (ii).

(iv) In the case when $k \neq k_j$ and $l \neq l_j$ for all $1 \leq j \leq G$, we write

$$d_l - a_k = (d_l - d_{l_j}) + (a_{k_j} - a_k) + (d_{l_j} - a_{k_j}) \in \frac{(l - l_j)T}{m} + \frac{(k_j - k)T}{n} + [-\delta - \eta + \alpha; \delta + \eta + \alpha + \varepsilon]_T \quad (28)$$

with $0 < |l - l_j| < m$ and $1 \leq j \leq G$ chosen in such a way that $k \in I_j(n)$, so that $0 < |k - k_j| < \frac{n}{G}$. Note that such a j exists by definition of the sets $I_j(n)$.

Next, we prove, in the same way as in Lemma 5.2, that $(l - l_j)\frac{n}{G} + (k_j - k)\frac{m}{G}$ is not a multiple of L , and hence can be written as $p + qL$ with integers q and p satisfying $1 \leq p \leq L - 1$. Indeed, if $(l - l_j)\frac{n}{G} + (k_j - k)\frac{m}{G}$ would be a multiple of $L = \frac{mn}{G}$, then for some integer σ we would have $(k_j - k)\frac{m}{G} = (\sigma m - l + l_j)\frac{n}{G}$. Since $\frac{m}{G}$ and $\frac{n}{G}$ do not have any common prime factors, this would imply that $\frac{n}{G}$ divides $(k_j - k)$, which contradicts $0 < |k - k_j| < \frac{n}{G}$. Thus, with $1 \leq p \leq L - 1$, we get in (28):

$$\frac{(l - l_j)T}{m} + \frac{(k_j - k)T}{n} = \left((l - l_j)\frac{n}{G} + (k_j - k)\frac{m}{G} \right) \times \frac{T}{L} = \frac{pT}{L} + qT.$$

These four cases prove the first implication. Note that we did not use condition (25) in this proof. Hence this proof is also valid if this condition is not satisfied.

Next we prove the converse implication. There are $L - 1$ distinct sets W_p . By Lemma 5.2, for $p = 1, \dots, L - 1$, there are at most G pairs (k, l) with $d_l - a_k \in W_p$. Thus there are at most $(L - 1)G = LG - G = mn - G$ pairs (k, l) with $d_l - a_k$ in the union $\bigcup_{p=1}^{L-1} W_p$.

Using the same arguments as in the proof of Lemma 5.2, we can show that there are at most G pairs (k, l) with $d_l - a_k \in [\alpha; \alpha + \varepsilon]_T$. By exhaustion, each set W_p contains exactly G differences, and the set $[\alpha; \alpha + \varepsilon]_T$ contains exactly G differences as well.

Thus we can write $(k_j, l_j) \in [\alpha; \alpha + \varepsilon]_T$ with $k_j \in I_j(n)$ and $l_j \in I_i(m)$ for $j = 1, \dots, n$. Each $I_i(m)$ contains at most, and hence exactly one l_j , so the mapping π , defined through $\pi(j) = i$, is a permutation on the set $\{1, \dots, G\}$. This proves the theorem. \blacksquare

Going back to the example described in Section 3.1, we have $m = n = 2$, $\delta = \eta = 0$, $\alpha = 10$ and $\varepsilon = 15$. Condition (25) is clearly satisfied in this case. We first have the synchronization constraints:

$$a_2 - a_1 \in [30; 30]_T \quad d_2 - d_1 \in [30; 30]_T$$

According to Theorem 5.3, the constraints that must be satisfied here in order to have two connections between the arriving and the departing trains are the following:

$$\begin{aligned} d_1 - a_1 &\in [10; 25]_T \cup [40; 55]_T & d_2 - a_1 &\in [10; 25]_T \cup [40; 55]_T \\ d_1 - a_2 &\in [10; 25]_T \cup [40; 55]_T & d_2 - a_2 &\in [10; 25]_T \cup [40; 55]_T \end{aligned}$$

These disjunctive constraints can be replaced in the usual way by two conjunctive constraints, giving the same constraints that were described in the example in Section 3.1.

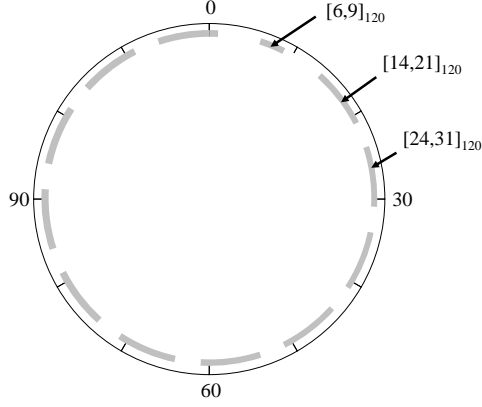


Figure 11: Periodic time intervals

The case when $T = 120$, $m = 4$, $n = 6$, $\delta = \eta = 1$, $\alpha = 6$ and $\varepsilon = 3$ provides the most simple example where $G = \gcd(m, n) = 2 > 1$ and m is not a multiple of n , nor vice-versa. Observe that the conditions of the theorem are satisfied. There are 11 distinct and disjoint intervals of the form $10p + [4, 11]_{120}$ for $1 \leq p \leq 11$, all disjoint from $[6, 9]_{120}$, see Figure 11. From the theorem we conclude that, if all differences $d_l - a_k$ are in the union of the aforementioned intervals, there are two connections, i.e., differences in $[6, 9]_{120}$.

If $\delta = \eta = 0$, then condition (25) is easily satisfied. Moreover, this condition is close to being necessary for the statement of Theorem 5.3, as the following Proposition 5.4 shows. Here two integers are called *co-prime* if their greatest common divisor is equal to one.

Proposition 5.4. *Assume that the positive integers m and n are co-prime. Assume further that the integers α , δ , η and ε satisfy: $0 \leq 2n\delta < T$, $0 \leq 2m\eta < T$, $0 \leq \varepsilon \leq \eta$, $\delta + 2\eta + \varepsilon < \frac{T}{mn}$, and $2\delta + 2\eta + \varepsilon \geq \frac{T}{mn}$. Then the sets*

$$A := \left\{ \frac{kT}{n} - \delta - \eta - \varepsilon, \frac{(mn-1)T}{mn} + \delta + \eta \mid 1 \leq k \leq n-1 \right\}_T,$$

and

$$D := \left\{ \alpha, \alpha + \frac{lT}{m} - \varepsilon, \mid 1 \leq l \leq m-1 \right\}_T$$

satisfy $A \in \mathcal{A}_{n,\delta}$, $D \in \mathcal{A}_{m,\eta}$. Moreover,

$$D - A \subset \left[\frac{T}{mn} - \delta - \eta + \alpha; T - \frac{T}{mn} + \delta + \eta + \alpha + \varepsilon \right]_T. \quad (29)$$

In this case, statement 2 in Theorem 5.3 holds, but statement 1 does not.

Proof. We first prove that $A \in \mathcal{A}_{n,\delta}$. Write $a_k = \frac{kT}{n} - \delta - \eta - \varepsilon$ for $1 \leq k \leq n-1$ and $a_n = \frac{(mn-1)T}{mn} + \delta + \eta$. Then $a_l - a_k = \frac{(l-k)T}{n}$ for $1 \leq k \leq l < n$. Moreover,

$$a_n - a_k = \frac{(mn-1)T}{mn} - \frac{kT}{n} + 2\delta + 2\eta + \varepsilon = \frac{(n-k)T}{n} + 2\delta + 2\eta + \varepsilon - \frac{T}{mn}.$$

Observe that, by assumption, $0 \leq 2\delta + 2\eta + \varepsilon - \frac{T}{mn} < \delta$. This proves that $A \in \mathcal{A}_{n,\delta}$. The proof that $D \in \mathcal{A}_{m,\eta}$ is straightforward and is omitted here.

Next we prove that $D - A$ is contained in the right hand side of (29). First, note that

$$D - A = \left\{ \alpha - \frac{kT}{n} + \delta + \eta + \varepsilon, \alpha + \frac{T}{mn} - \delta - \eta, \alpha + \frac{lT}{m} - \frac{kT}{n} + \delta + \eta, \right. \\ \left. \alpha + \frac{lT}{m} + \frac{T}{mn} - \delta - \eta - \varepsilon \mid 1 \leq k \leq n-1, 1 \leq l \leq m-1 \right\}_T.$$

Furthermore, the assumption $2\delta + 2\eta + \varepsilon \geq \frac{T}{mn}$ implies that

$$\mathcal{U} := \bigcup_{k=1}^{mn-1} \frac{kT}{mn} + [-\delta - \eta + \alpha; \delta + \eta + \alpha + \varepsilon]_T = \left[\frac{T}{mn} - \delta - \eta + \alpha; T - \frac{T}{mn} + \delta + \eta + \alpha + \varepsilon \right]_T.$$

Thus, since $\delta + \eta + \varepsilon < \frac{T}{mn}$, the sets \mathcal{U} and $[\alpha; \alpha + \varepsilon]_T$ have an empty intersection. Using the fact that $ln - km$ is not a multiple of mn for $1 \leq k \leq n-1$ and $1 \leq l \leq m-1$, it is not difficult to verify that $D - A \subset \mathcal{U}$. Thus the latter implies that statement 2 in Theorem 5.3 holds, but statement 1 does not. \blacksquare

Observe that the case with $T = 60$, $m = 2$, $n = 3$, $\alpha = 3$, and $\delta = \eta = \varepsilon = 2$ is covered by Proposition 5.4. According to the description in Proposition 5.4, we have $A = \{14, 34, 54\}$ and $D = \{3, 31\}$ in this case. Thus there is no appropriate connection with a time interval in $[3, 5]$ from an arriving train to a departing train. This is caused by the fact that the flexibility represented by the parameters δ , η and ε is too large.

6 Application

In this section we illustrate the advantages of flexible connections by an application of the PESP model to three connected intercity lines of Netherlands Railways. The Dutch railway timetable is based on a cyclic timetable with a period of $T = 60$ minutes (Kroon et al., 2009). Most lines are operated with a frequency of two trains per hour, but there are also lines, for example freight lines, that are operated only once per hour. Therefore, it is not possible to consider the timetable as having a period of 30 minutes.

In this example, we consider the intercity lines 800, 1900, and 2600. These lines are all operated twice per hour in both directions between the stations Alkmaar (Amr) and Maastricht (Mt) (800 line), between The Hague (Gvc) and Venlo (Vl) (1900 line), and between Amsterdam (Asd) and The Hague (Gvc) (2600 line), see Figure 12. In Amsterdam, The Hague, and Eindhoven (Ehv) the lines have common stations, where

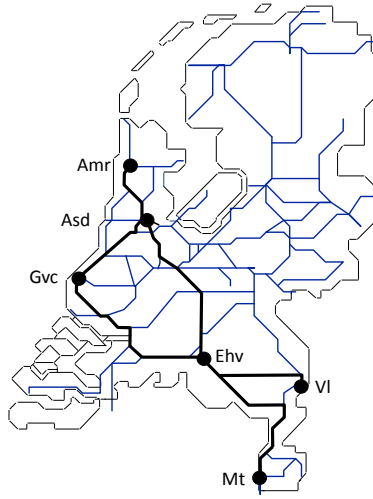


Figure 12: The Dutch railway network with the 800, 1900, and 2600 lines

passenger connections in both directions are created. In Eindhoven, the current (2012) timetable has cross-platform connections in both directions between the 800 line and the 1900 line. This means that passengers have a connection time in both directions equal to three minutes. The maximum dwell time in Eindhoven is also equal to three minutes, so the trains, respectively in the direction of Maastricht (800 line) and Venlo (1900 line), or in the direction of Alkmaar (800 line) and The Hague (1900 line), have to arrive and depart at exactly the same time in Eindhoven.

In the PESP model, each line is required to be operated with between 28 and 32 minutes between successive trains in the same direction, and the dwell times have to be between 1 and 3 minutes at each intermediate station. We also use the rule that for each trip the running time in one direction is the same as in the reverse direction. The return times at the terminal stations are required to be between 10 and 25 minutes.

The passenger connection times in Amsterdam and The Hague are required to be between 3 and 15 minutes, and in Eindhoven they must be between 3 and 20 minutes, as will be explained later. All passenger connections in Amsterdam, The Hague and Eindhoven are modeled with flexible connections. The rolling stock connections in the *northern* terminal stations of the lines (Alkmaar for the 800 line, The Hague for the 1900 line, and Amsterdam for the 2600 line) are modeled with flexible connections. Recall that it does not help to use flexible rolling stock connections at *both* terminal stations of a line.

We implemented the PESP model with flexible connections in OPL Studio 6.3. The number of constraints of the model equals 474, the number of non-zeros equals 1596, and the number of variables equals 470, with 120 binary variables, 156 integer variables and 194 continuous variables. So the instance is relatively small, but we use it mainly to illustrate the advantages of the flexible connections. In a larger instance, these advantages will be less obvious due to all kinds of other interactions and relationships.

We used the MIP solver CPLEX 12.1.0 to find optimal solutions, thereby minimizing:

- the number of required train compositions (see Section 3.2), or
- the passenger connection times in Amsterdam, The Hague, and Eindhoven (see Section 4.2), or
- the sum of the running times and the dwell times.

The third objective of the sum of the running times and the dwell times can be defined by referring to (2). To that end, let A_r be the set of arcs representing a trip of a train from one station to another, and let A_d be the set of arcs representing a train dwelling in a station. Then the sum of the running times and the dwell times is computed as:

$$\sum_{a=(j,k) \in A_r \cup A_d} (\pi_k - \pi_j + T \times p_a)$$

A combination of the mentioned objectives can be handled by giving each objective a certain weight. Furthermore, when minimizing either of the three objectives, the other objectives are taken into account with a small weight in order to minimize them, given the minimal value of the main objective.

6.1 Results

When using flexible connections, it turns out that the required number of train compositions is either 27 or 28. However when some rolling stock or passenger connections are fixed a priori, more compositions may be needed, see Section 6.2. With flexible connections, using more than 28 compositions does not lead to any improvement in the objectives, so we fixed the number of compositions to either 27 or 28, and then minimized the sum of the running and dwell times, or the total passenger connection time.

	Comp.	Connect.	Running	Dwell	R+D	Return
R+D	27	152	1380	32	1412	208
R+D	28	168	1400	32	1432	248
Connection	27	100	1424	40	1464	156
Connection	28	48	1466	60	1526	154

Table 1: Results obtained in successive optimization runs

Table 1 shows per row the results obtained in each run, while minimizing either the sum of the running and dwell times (rows “R+D”) or the total passenger connection time (rows “Connection”), and at the same time fixing the number of compositions at 27 or 28.

Each row shows in horizontal direction the obtained results in each run: the total number of compositions (“Comp.”), the total passenger connection time (“Connect.”), the total running time (“Running”), the total dwell time (“Dwell”), the sum of the running and dwell times (“R+D”), and the total return time (“Return”). Note that in all cases the sum of the running, dwell and return times equals 60 times the number of train compositions, as could be expected. Furthermore, in all cases the computation times are shorter than 5 seconds, therefore we did not show them in the table.

6.1.1 Running and dwell times

As can be seen in Table 1, when minimizing the running and dwell times, the running times and the return times are larger in the case with 28 compositions than in the case with 27 compositions. In the timetable with 27 compositions, see Figure 13, all running and dwell times are at their lower bounds. This timetable has one combined rolling stock circulation (line 800). The timetable with 28 compositions, see Figure 14, has two combined rolling stock circulations (lines 800 and 1900).

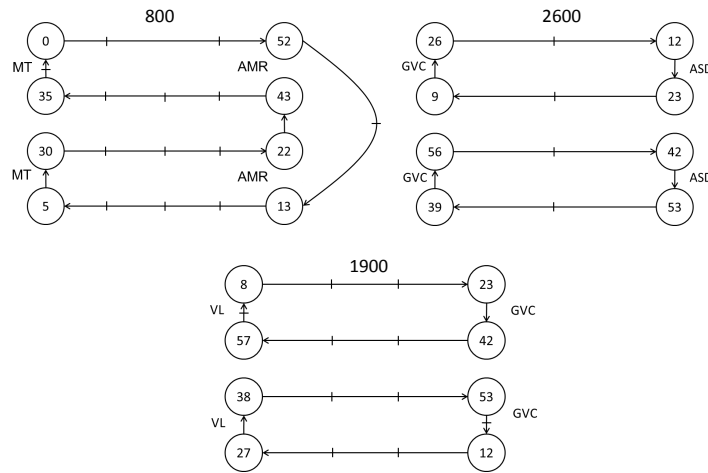


Figure 13: Results when minimizing R+D times with 27 compositions

When going from 27 to 28 compositions, the extra composition has to be added to exactly one of the three lines. This line will be the one where adding the additional composition leads to the least increase in running and dwell times. This is the timetable that can absorb the largest part of the added 60 minutes in the return times. Note that for each line an upper bound on the total return time is $4 \times 25 = 100$ minutes.

In the timetable with 27 compositions, the 800 line has a total return time of 92 minutes (8 spare minutes), the 1900 line has a total return time of 60 minutes (40 spare minutes), and the 2600 line has a total return time of 56 minutes (44 spare minutes). Thus at first sight it seems optimal if the extra composition would be added to the 2600 line. However, due to the passenger connection constraints, the 2600 line is not able to use all 44 spare minutes in the return times to absorb the added 60 minutes of the extra composition. Therefore, the extra composition is added to the 1900 line. This line is able to use all 40 spare minutes in its return times. The other 20 minutes are added to the running times.

Note that the rolling stock circulations of the 800 line and the 2600 line have the same structure in the timetables with 27 and 28 compositions. The rolling stock circulation of the 1900 line flips from two separate circulations to one combined circulation.

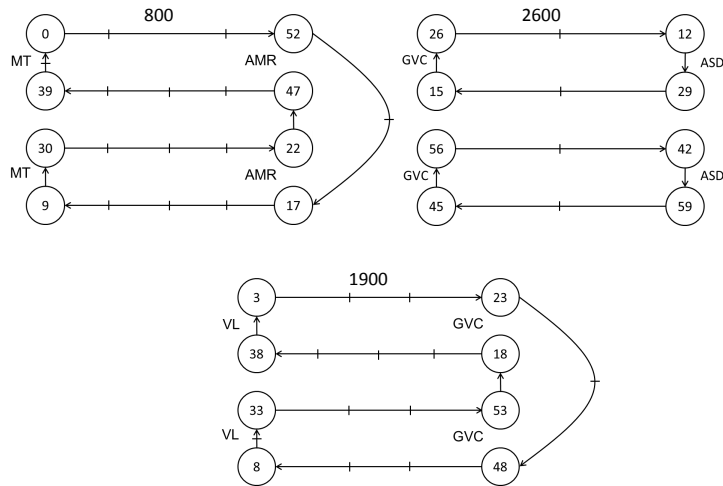


Figure 14: Results when minimizing R+D times with 28 compositions

6.1.2 Passenger connection times

When minimizing the passenger connection times, the obtained timetables differ mostly with respect to the cross-platform connections in Eindhoven. Remember that, in order to have a cross-platform connection, two trains of the two different lines have to arrive and depart at exactly the same time in Eindhoven.

Tables 2 and 3 show the arrival and departure times of the trains in Eindhoven when minimizing the passenger connection times with 27 compositions and 28 compositions, respectively. Here “ID” is the identification number for the two trains of the same line.

With 28 compositions, the cross-platform connections can be made in both directions. However, with 27 compositions, a cross-platform connection can be made only in the northern direction. Thus the main trade-off here is the one between the number of required compositions and the total connection time.

If a cross-platform connection can be made in one direction, then the total connection time in these connections is $2 \times 2 \times 3 = 12$ minutes. Furthermore, with 27 compositions, the total passenger connection time in the direction of Maastricht and Venlo equals 64 minutes ($48 \rightarrow 0 = 12$; $18 \rightarrow 30 = 12$; $59 \rightarrow 19 = 20$; $29 \rightarrow 49 = 20$).

Note that, if a cross-platform connection cannot be made in one direction, then the total passenger connection time in the resulting connections equals $2 \times ((x+1) + (31-x)) = 64$ minutes. Here x is the time between the arrival of a train of the first line and the first successive arrival of a train of the other line. Thus $10 \leq x \leq 20$, due to the maximum connection time of 20 minutes in Eindhoven. The additional single minutes in the terms $(x+1)$ and $31 = 30 + 1$ are due to the dwell times in Eindhoven.

This also explains why the upper bound for the connection times in Eindhoven is 20 minutes instead of 15 minutes, as in the other locations. If the upper bound would be 15,

Line	Heading	Direction	ID	Arrival time	Departure time
800	North	Amr	1	0	3
800	North	Amr	2	30	33
1900	North	Gvc	1	30	33
1900	North	Gvc	2	0	3
800	South	Mt	1	48	49
800	South	Mt	2	18	19
1900	South	Vl	1	59	0
1900	South	Vl	2	29	30

Table 2: Connections at Eindhoven with 27 compositions

Line	Heading	Direction	ID	Arrival time	Departure time
800	North	Amr	1	59	2
800	North	Amr	2	29	32
1900	North	Gvc	1	29	32
1900	North	Gvc	2	59	2
800	South	Mt	1	47	50
800	South	Mt	2	17	20
1900	South	Vl	1	47	50
1900	South	Vl	2	17	20

Table 3: Connections at Eindhoven with 28 compositions

then the cross-platform connections would be forced, which is too restrictive.

As was mentioned before, the main managerial question with respect to the passenger connections in Eindhoven is whether the cross-platform connections are worth an additional train composition. In the current Dutch railway system, this question is answered in a positive way, since the cross-platform connections are used by many passengers. In fact, Eindhoven is an important hub in the connection between the southern part of the Netherlands and the rest of the country.

Therefore, Table 4 shows the results after fixing the cross-platform connections in Eindhoven and minimizing the passenger connection times (row “Connection”) or the running and dwell times (row “R+D”). As can be seen, the number of compositions equals 28 in both cases. Thus, if the cross-platform connections in Eindhoven are fixed, then the next trade-off that must be made for the rest of the timetable is the one between low passenger connection times and low running and dwell times.

	Comp.	Connect.	Running	Dwell	R+D	Return
Connection	28	48	1466	60	1526	154
R+D	28	110	1396	48	1444	236

Table 4: Results obtained when fixing the cross-platform connections

6.2 A priori fixing

In order to show the advantages of the flexible connections, we will compare the results of the flexible model with the results of three different models where the connections are fixed a priori. In all models, the sum of the running and dwell times is minimized. In the first model, the rolling stock connections are fixed such that there are three separate rolling stock circulations. Secondly, a model is considered where the passenger connections at all common stations are fixed such that the first train of a line has a passenger connection with the first train of the other line. Finally, a model is considered where both the rolling stock connections and the passenger connections are fixed in the above way.

Table 5 shows the results for the four cases. The vertical direction shows the mentioned cases, where “Flexible” is the case where flexible connections are used, “Rolling stock” is the case where only the rolling stock connections are fixed, “Passenger” is the case where only the passenger connections are fixed, and “Pass + Roll” is the case where both the passenger connections and the rolling stock connections are fixed. In the horizontal direction the obtained results are shown. As can be seen, fixing only the rolling stock connections already leads to worse results than using flexible connections on all objectives. Fixing the passenger connections leads to even worse results for the running and dwell times, and only the passenger connection times improved (as could be expected). Fixing both passenger and rolling stock connections leads to infeasibility. These results clearly demonstrate the advantages of the flexible connections.

	Comp.	Connect.	Running	Dwell	R+D	Return
Flexible	27	154	1380	32	1412	208
Rolling stock	28	194	1418	46	1464	216
Passenger	29	77	1468	76	1544	196
Pass + Roll	-	-	-	-	-	-

Table 5: Fixing connections a priori

6.3 Sensitivity and Robustness

As was illustrated in the earlier sections, the flexible connections allow the existence of timetables with rather different structures. For example, depending on the rolling stock connections, a single combined rolling stock circulation may be operated on a train line, or two separate rolling stock connections. Which of the two options is the better one depends on the details of the input data. Changing the input data may flip the structure of the timetable. Thus the timetable is rather sensitive with respect to the input data.

The sensitivity due to the passenger connections is most striking for the cross-platform connections in Eindhoven. As was indicated, if the cross-platform connections can be made in one direction, then the total connection time in these connections is $2 \times 2 \times 3 = 12$ minutes per period. If the cross platform connections cannot be made in one direction, then the total connection time in these connections is $2 \times ((x + 1) + (31 - x)) = 64$ minutes per period, where $10 \leq x \leq 20$. Again, whether or not the cross-platform connections can

be made depends on the details of the input data. Changing the input data may flip the structure of the timetable from a timetable *with* cross-platform connections to one *without* cross-platform connections, or vice versa.

For example, if the minimum return time is decreased from 10 minutes to 9 minutes, then it turns out that the cross-platform connections in Eindhoven can be made in two directions with only 27 train compositions. In that case, one has a low number of compositions and short passenger connection times, but one possible jeopardizes the robustness of the timetable. This is yet another trade-off than the ones mentioned before.

Indeed, if some process times (running and dwell times, return times, headway times) are planned at (or even below) their technical minimum, then the resulting timetable will not be robust. That is, small disturbances in the operations will lead to delays, and delays cannot be recovered. Therefore, the robustness of a timetable is usually improved by adding certain buffer times to the technically minimum process times. However, determining appropriate buffer times is out of the scope of the current paper. For a description of a stochastic optimization model for determining appropriate buffer times in a cyclic timetable, we refer to Kroon et al. (2008). The robustness of a timetable can be evaluated by applying simulation models, see e.g. Hürlimann (2001).

7 Summary and conclusions

In this paper we illustrated that a priori fixing rolling stock or passenger connections in a PESP model may lead to suboptimal solutions if the involved trains are operated more than once per hour. It may even lead to infeasibility for instances where a feasible solution would exist if the connections were selected in the right way. These results were illustrated by a case based on three intercity lines of Netherlands Railways, the largest operator of passenger trains in the Netherlands.

Using the well-known method for translating pairs of disjunctive constraints in a PESP model into pairs of conjunctive constraints, we described the rolling stock and passenger connections in a flexible way, so that the model can choose between the different options. For the cases of two and three trains per hour in both directions, we expressed the number of train compositions needed to operate the timetable in terms of the cycle variables of certain cycles in the constraint graph.

We also briefly indicated that, under certain conditions, the results for three trains per hour are also valid for any prime number of trains per hour. For a larger composite number n of trains per hour, which hardly ever occurs in practice, the situation is different, due to the fact that in that case also other circulations exist than (i) one large circulation containing n arriving and n departing trains, or (ii) n separate circulations of one arriving and one departing trains.

Finally, for the case that there are n arriving trains and m departing trains, we described under which conditions it can be guaranteed by PESP constraints that there will be $G = \text{gcd}(m, n)$ connections between the arriving and the departing trains. This is the best one can hope for if the arriving and departing trains are evenly spread in time.

The described flexible connections have been implemented in the DONS system. DONS

is the automated timetabling system used by Netherlands Railways and ProRail for generating periodic timetables for the Dutch railway system (Hooghiemstra et al., 1999; Kroon et al., 2009). ProRail is the manager of the Dutch railway infrastructure. CADANS, the solver of the underlying PESP model, was developed by Schrijver and Steenbeek (1993). The flexible connections in DONS are mainly passenger connections, but flexible rolling stock connections can be applied as well. Due to the flexible connections, the underlying PESP model turns out to be feasible more often, and the obtained solutions are usually better than the ones obtained with a priori fixed connections.

References

- Borndörfer, R. and C. Liebchen (2008). When periodic timetables are suboptimal. In J. Kalcsics and S. Nickel (Eds.), *Operations Research Proceedings 2007*, pp. 449–454. Springer Verlag, Berlin.
- Hooghiemstra, J.S., L.G. Kroon, M.A. Odijk, M. Salomon, and P.J. Zwaneveld (1999). Decision support systems support the search for win-win solutions in railway network design. *Interfaces* 29(2).
- Hürlimann, D. (2001). *Object oriented modeling of infrastructure elements and business processes in railways*. Ph. D. thesis, ETH Zürich.
- Kroon, L.G., D. Huisman, E. Abbink, P.J. Fioole, M. Fischetti, G. Maróti, A. Schrijver, A. Steenbeek, and R. Ybema (2009). The New Dutch Timetable: The OR Revolution. *Interfaces* 39(1), 6–17.
- Kroon, L.G., G. Maróti, M. Retel Helmrich, M.J.C.M. Vromans, and R. Dekker (2008). Stochastic Improvement of Cyclic Railway Timetables. *Transportation Research Part B: Methodological* 42(6), 553–570.
- Kroon, L.G. and L.W.P. Peeters (2003). A variable trip time model for cyclic railway timetabling. *Transportation Science* 37, 198–212.
- Liebchen, C. (2004). Symmetry for railway periodic timetables. *Electronic Notes in Theoretical Computer Science* 92, 34–51.
- Liebchen, C. (2006). *Periodic Timetable Optimization in Public Transport*. Dissertation.de – Verlag im Internet.
- Liebchen, C. and L.W.P. Peeters (2009). Integral cycle bases for cyclic timetabling. *Discrete Optimization* 6(1), 98–109.
- Nielsen, M.N., B. Hove, and J. Clausen (2006). Constructing periodic timetables using MIP - a case study from DSB S-train. *International Journal of Operational Research* 1(3), 213–227.

- Odijk, M.A. (1996). A constraint generation algorithm for the construction of periodic railway timetables. *Transportation Research B* 30(6), 455–464.
- Peeters, L.W.P. (2003). *Cyclic Railway Timetable Optimization*. Ph.D. thesis, Erasmus Universiteit Rotterdam.
- Schrijver, A. and A.G. Steenbeek (1993). Dienstregelingontwikkeling voor Nederlandse Spoorwegen N.S. Rapport Fase 1, Centrum voor Wiskunde en Informatica. In Dutch.
- Serafini, P. and W. Ukovich (1989). A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics* 2(4), 550–581.

Publications in the Report Series Research* in Management

ERIM Research Program: “Business Processes, Logistics and Information Systems”

2012

Solving Weighted Voting Game Design Problems Optimally: Representations, Synthesis, and Enumeration

Bart de Keijzer, Tomas B. Klos, and Yingqian Zhang

ERS-2012-006-LIS

<http://hdl.handle.net/1765/32170>

Flexible Connections in PESP Models for Cyclic Passenger Railway Timetabling

Leo G. Kroon Leon, W.P. Peeters, Joris C. Wagenaar, and Rob A. Zuidwijk

ERS-2012-008-LIS

<http://hdl.handle.net/1765/32996>

* A complete overview of the ERIM Report Series Research in Management:

<https://ep.eur.nl/handle/1765/1>

ERIM Research Programs:

LIS Business Processes, Logistics and Information Systems

ORG Organizing for Performance

MKT Marketing

F&A Finance and Accounting

STR Strategy and Entrepreneurship