

O.R. Applications

Developing railway timetables which guarantee a better service

P. Vansteenwegen^{*}, D. Van Oudheusden*Centrum voor Industrieel Beleid, Katholieke Universiteit Leuven, Celestijnenlaan 300A, 3001 Leuven, Belgium*

Received 21 November 2003; accepted 9 December 2004

Available online 10 March 2005

Abstract

In order to improve passenger service, a waiting cost function, weighting different types of waiting times and late arrivals, is designed and minimised. The approach is applied to a small part of the Belgian railway network. In the first phase of the approach, ideal buffer times are calculated to safeguard connections when the arriving train is late. These buffer times are based on the delay distributions of the arriving trains and on the weighting of different types of waiting times. In a second phase, standard linear programming is used to construct an improved timetable with well-scheduled connections and, whenever possible, with ideal buffer times. Simulation compares different timetables and optimises the LP timetable. For the case of the Belgian railway network, the final result is a timetable with well-scheduled connections and a waiting cost that is 40% lower than the current timetable. Since only LP modelling is applied, the proposed technique is very promising for developing better timetables—even for very extensive railway networks.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Robustness; Linear programming; Timetabling; Railway connections; Buffer times

1. Introduction

Ensuring a good service in public transportation is a very important task of the state-owned Belgian Railway Company NMBS (Nationale Maatschappij der Belgische Spoorwegen). In Belgium, a better public transportation is seen as the far most important way to relieve pressing traffic

problems. An effective timetable with trains arriving on schedule is a crucial factor for good passenger service [8]. The current railway timetable of the NMBS is based on theoretical “ideal running times”. The ideal running time is the standard time it takes for a specific train to travel from one station to another; it assumes ideal circumstances. At this moment, hardly any time buffer is implemented against small disturbances and non-ideal circumstances, which frequently occur in practice. This paper presents a technique for creating robust timetables that perform well, even in non-ideal circumstances. Particularly, the technique guarantees

^{*} Corresponding author. Tel.: +32 16 322 700; fax: +32 16 322 986.

E-mail address: pieter.vansteenwegen@cib.kuleuven.ac.be (P. Vansteenwegen).

a very functional timetable during rush hour. Since most passengers take a train during rush hour, many passengers will benefit from this approach.

The time a passenger spends waiting is a very critical element for judging passenger service. Typically, a railway passenger faces different types of waiting due to different causes. For instance, when connections are not properly scheduled, a passenger will have to wait a long time between trains. Trains running behind schedule will also create waiting times. During rush hours most of the trains meet with some considerable delay. Thus, the actual travel time it takes a train during rush hour is typically longer than the ideal running time. Not taking this expected delay into account while scheduling, will generate all kinds of waiting [8]. A missed connection causes the biggest waiting cost. The chances of missing a connection increase when connections are scheduled tightly. To lower the chance of a missed connection, usually buffer times are inserted into the schedule [2,6]. On the other hand, these buffer times will keep waiting seated passengers when the train is on time. Still other types of waiting and corresponding inconveniences are dealt with in this paper (see Section 3). In [3–6] a lot of attention is given to “synchronisation”. The purpose of synchronisation is to determine how long a connecting train should wait when an arriving train is late. In this paper it is assumed that a connecting train always leaves on schedule, if possible. Thus, synchronisation policies can be added to further improve the service. Such policies are, however, outside the scope of this paper.

Based on the delay distribution of a specific train and a weighting of the different waiting times, one can calculate an ideal buffer time for a specific connection. A more detailed explanation of how to calculate ideal buffer times is given in Section 3. In Section 4 a sophisticated approach brings into play the ideal buffer times by means of Linear Programming (LP). More precisely, linear goal programming inserts buffer times into the schedule that approach the ideal buffer times as closely as possible. In this formulation, the ideal buffer times are goals to be met. Going both over and under a goal is penalised in the objective function of the LP. The scheduled buffer times cannot meet the ideal

buffer times exactly because of the conflicting tendencies. Since it is linear programming it does not take much time to solve very large instances and it is easy to perform sensitivity analysis. In the past, integer programming has been frequently used to create timetables [10–12]. Integer problems, however, are much more difficult to solve and to interpret [1]. An important outcome of this paper is that linear programming can be sufficient. In Section 5 discrete event simulation will make possible a detailed comparison between the timetables: the original one and the LP generated timetable. Performance measures such as total minutes of waiting time, total generalised waiting cost, etc., can be considered. The simulation results are analysed and discussed.

2. The railway network

In this section the railway network under study is discussed in detail. It is not convenient to deal at once with the complete Belgian network. Hence we focus on a small part of the Belgian railway network and only consider passenger trains. The network consists of seven stations and four train lines connecting these stations. The network is shown in Fig. 1. Most of the lines start in a station outside this miniature network and enter through one of the stations (cities) of the network. For instance, line C starts in “Antwerpen” and enters the network in “Heist-op-den-Berg”. Every line has trains travelling in two directions, here called the 0 and 1 direction.

Cities in the network: Aarschot, Alken, Hasselt, Heist-op-den-Berg, Landen, Leuven and Sint-Truiden.

Different train lines:

C0 and C1, between Heist-op-den-Berg (“Antwerpen” is the real place of origin) and Hasselt (“Liège”);
E0 and E1, between Leuven (“Brussel”) and Hasselt, through Aarschot;
K0 and K1, between Leuven (“Brussel”) and Hasselt (“Genk”), through Landen;
M0 and M1, between Leuven and Heist-op-den-Berg (“Antwerpen”).

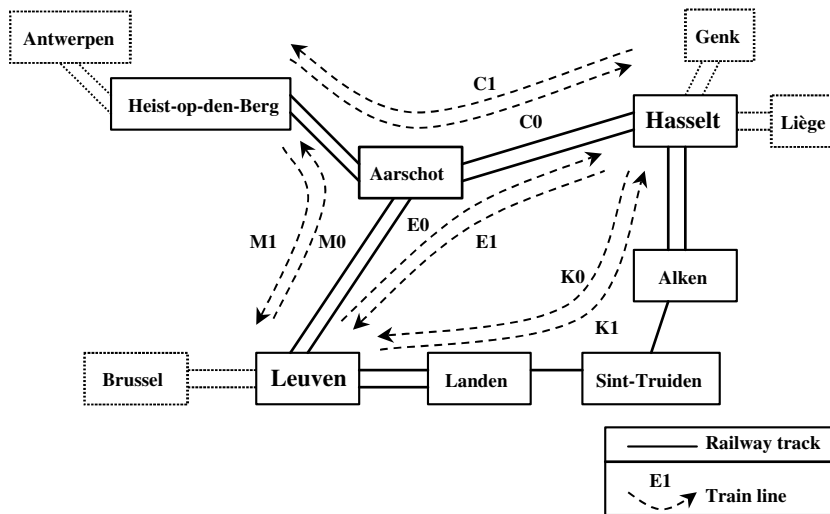


Fig. 1. Railway network.

Lines C and K use Intercity trains (IC) with an average speed of 76 km/h. Line E uses an Inter Regional train (IR) with an average speed of 65 km/h and line M uses a Local train (L) with an average speed 51 km/h [9].

All tracks are double tracks, except between Landen and Sint-Truiden, and between Sint-Truiden and Alken, which are single tracks. This means that, in any direction, there can be only one train at any time. In the station of Sint-Truiden two trains travelling in opposite directions can cross one another.

The NMBS operates a cyclic timetable with a standard period of 60 minutes. This means that the timetable is repeated every hour. If it were acceptable to create different timetables for the morning or evening rush hours or for holiday periods, it would open extra possibilities for improvement. Yet NMBS considers a cyclic timetable necessary for clarity towards the passengers. Shires et al. [16] report on periodicity benefits in general.

For some of the possible trips in this network more than one line is available. For instance between Leuven and Aarschot one can take the M or the E line. It is obvious that customers will be better-off (see also [20]) if the trains of these lines are evenly spread over the period under consider-

ation. In our case, the considered period is 1 hour and, as a consequence, the time between the M and E trips should be approximately 30 minutes. If the lines are not evenly spread, the time between two possible trips can reach up to 55 minutes.

Most of the trains stop in more stations than those mentioned in Fig. 1, but no connections take place in these stations and the number of tracks does not change either. Thus, these stations have no real impact on the timetable problem and are therefore omitted.

In order to minimise “a total generalised waiting cost” for the passengers, it is important to know which connections are frequently used and must be guaranteed, and which connections are unimportant. It appears that only in Leuven and Hasselt connections are crucial. It is furthermore understood that passengers always prefer a direct trip to a trip with one or more connections.

The important connections to be made in Hasselt and Leuven are

Hasselt: from train K1 to train C1: to go from Alken to Aarschot;
 from train K1 to train C0: to go from Alken to Liège;
 from train C0 to train K0: to go from Aarschot to Alken.

Table 1

Current timetable for one period (in minutes after the start of the period)

	Arrival	Departure	A	D	A	D	A	D	A	D	A	D	A	D
	Heist-opden-Berg		Aarschot		Leuven		Landen		Sint-Truiden		Alken		Hasselt	
C0	–	00	09	14	–	–	–	–	–	–	–	–	40	46
M1	–	19	31	32	46	–	–	–	–	–	–	–	–	–
K1	–	–	–	–	–	35	62	64	73	74	82	83	89	–
	Hasselt		Alken		Sint-Truiden		Landen		Leuven		Aarschot		Heist-opden-Berg	
C1	–	20	–	–	–	–	–	–	–	–	46	51	60	–
M0	–	–	–	–	–	–	–	–	–	14	28	29	41	–
K0	–	31	37	38	46	47	56	58	85	87	–	–	–	–
	Leuven		Aarschot		Hasselt									
E0	–	11	23	25	50	–	–	–	–	–	–	–	–	–
	Hasselt		Aarschot		Leuven									
E1	–	10	35	37	49	–	–	–	–	–	–	–	–	–

Leuven: from train K0 to train E0: to go from Landen to Aarschot;
 from train K0 to train M0: to go from Landen to Heist-op-den-Berg;
 from train E1 to train K1: to go from Aarschot to Landen;
 from train M1 to train K1: to go from Heist-op-den-Berg to Landen;
 from train M1 to train K0: to go from Aarschot to Brussel.

The time it takes for a specific train to ride from one station to the next is called the “(ideal) running time”. The ideal running time can be determined by making use of the current timetable for this part of the Belgian network. In Table 1 the current timetable is presented for one period. This table indicates, for example, that train C0 should arrive in Aarschot 9 minutes after every hour, and depart from Aarschot 14 minutes after every hour. In the current timetable, no buffer times are added to the ideal running time. The scheduled transfer times are very large, as one can see in Table 1. For example, since train C0 arrives in Hasselt 40 minutes past the hour and train K0 leaves Hasselt 31 minutes past the hour, a passenger has to wait 48 minute (assuming 3 minutes for actual transfer).

3. Ideal buffer times

The “ideal running time” is the smallest possible running time for the arriving train, from its previous station to the connection station. But the “real running time” of a train will vary, depending on the circumstances. A buffer is determined in order to assure a certain connection; it is only calculated for a connection station. A buffer “B” is added to the ideal running time, to determine the “scheduled running time” and the “scheduled arrival time”. The scheduled times, announced to the public, will thus include these buffer times. The difference between the real running time and the ideal running time is called the delay “*d*” of a train (see Fig. 2). Thus, the buffer increases the train’s opportunity to arrive “on time”. As a result a train will, under ideal circumstances, arrive earlier than scheduled and there will be some extra stopping time in the station. The stopping time is the time a train spends in a station, between arrival and departure. It is assumed that the connecting train always leaves on schedule.

3.1. Waiting costs

Three types of passengers can be distinguished for a train approaching a connection station:

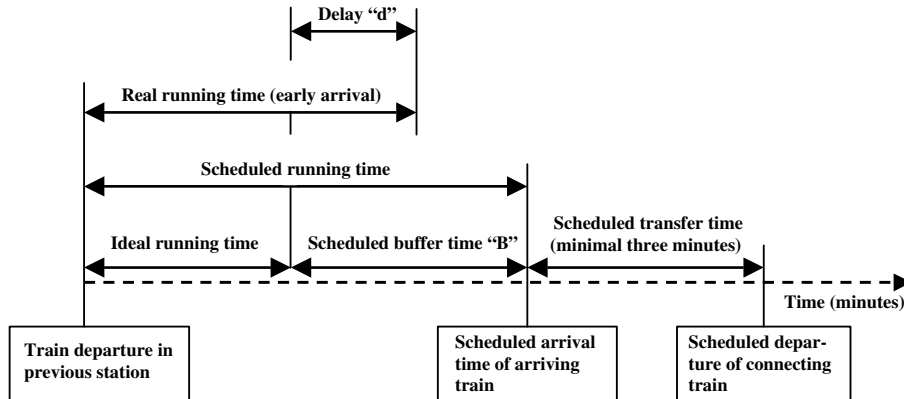


Fig. 2. Definition of concepts.

- transfer passengers, who transfer to another train;
- remaining passengers, who remain on the same train;
- arriving passengers, who reach their end station and leave the train system altogether.

The passengers boarding a train at the start of their journey are not considered in the buffer calculation. If they have to wait because their first train is late, it may inflict a cost on them at the end of their journey. If they still arrive on time there is no problem.

Substantial research has been carried out to determine the value of time (VOT) in different circumstances [5,7,17–19]. In the case of waiting for trains (or public transportation in general), the VOT is almost always expressed relatively to the driving time. A commonly used factor to express the value of waiting time is “two” [18]. This means that passengers rate 1 minute of waiting as equivalent to 2 minutes of driving. In other words, 1 minute of waiting for one passenger will induce a generalised waiting cost of two units.

Adding a buffer B still allows a train to be early or late and four types of waiting can incur. Each type of waiting has its own inconveniences and will be weighed differently. Arriving late will induce a certain cost as well. The different weights will all be in the neighbourhood of the value two.

The transfer passengers face two types of waiting. When they are on time or early, they have to

wait in the station for their connecting train. This could be interpreted as a “normal” type of waiting. The weight used here is exactly two. The transfer passengers face a second type of waiting when they arrive late and miss their connection. This means they have to wait for the next train to their destination. In our case study the duration will be either about 30 or about 60 minutes. The weight for this type of waiting is fixed at 2.2. Indeed, passengers will be more annoyed with this time-wasting than in the previous case. The value 2.2 is relatively low because it has to be multiplied with 30 or 60 to determine the total cost for missing a connection. In all other cases, passengers mostly have to wait for less than 15 minutes.

In previous articles, only the transfer passengers are taken into account [3,5,6]. But remaining and arriving passengers also encounter waiting time. The remaining passengers have to wait when the train arrives too early in a station. They will not comprehend why they stand still, and waiting is always less pleasant than driving. The selected weight here is 1.5. These passengers will arrive on time and can remain seated while waiting; it is much more convenient than standing in a station hall. When the arriving passengers arrive late, they do not have to wait, but they face “a certain cost”. The generalised waiting cost in this case is 2.5 since arriving late is very inconvenient indeed. Arriving late at the end of a journey is what causes most worrying.

Sensitivity analysis on these weights has shown that a small change in the weight values does not make a significant difference for the ideal buffer time calculation.

3.2. Ideal buffer calculation

In order to calculate the ideal buffer, it is necessary to have an idea about the chances of encountering a particular delay. In normal circumstances there will be no delay at all. During rush hour, however, most of the trains can expect a specific delay.

The exponential distribution

$$f(d) = \frac{1}{\lambda} e^{-\frac{1}{\lambda}d} \quad (d \geq 0),$$

$$F(D) = p(D \leq d) = 1 - e^{-\frac{1}{\lambda}D} \quad (d \geq 0),$$

is frequently used to model the delay of a rail journey [4,14]. In this formulation, λ is the expected delay. The data about delays, provided by NMBS, also demonstrates this assumption. For determining buffer times, a generalised waiting cost function can be calculated based on the exponential distribution. Each type of waiting will induce a part of the generalised cost. The procedure applied to calculate the various waiting costs is explained in the next four sections.

3.2.1. Transfer passengers' waiting time

Transfer passengers have to wait when they miss their connection or when the arriving train arrives earlier than scheduled. Passengers will miss their connection when the delay of the train is greater than the scheduled buffer. This probability is

$$p(d > B) = 1 - F(B) = e^{-\frac{1}{\lambda}B},$$

where B is the scheduled buffer size. A missed connection implies that every passenger has to wait 30 or 60 minutes (notation p) for the next connecting train. Actually, the passengers will have to wait $p - (d - B)$ minutes. However, in most cases $d - B$ is relatively small compared to a period p of 30 or 60 minutes, so $d - B$ is ignored. This approximation has no significant influence on the result (see also Section 4.4). The weighted cost because of a missed connection C_{mc} is, per passenger,

$$C_{mc}(B) = 2.2pe^{-\frac{1}{\lambda}B},$$

where the factor 2.2 is the weight for a missed connection.

Transfer passengers also have to wait when the arriving train arrives earlier than scheduled. The chance for this to happen is given by the exponential distribution. The delay d is now smaller than the buffer and the time a passenger has to wait is $x = B - d$.

The density function of d is

$$f(d) = e^{-\frac{1}{\lambda}d} = e^{-\frac{1}{\lambda}(B-x)}.$$

Integration allows us to calculate the cost C_{wc} for waiting, per passenger:

$$C_{wc}(B) = 2.0 \int_0^B x \frac{1}{\lambda} e^{-\frac{1}{\lambda}(B-x)} dx = 2.0 \left[B + \lambda \left(e^{-\frac{1}{\lambda}B} - 1 \right) \right],$$

where 2.0 is the transfer time weight.

3.2.2. Remaining passengers' waiting time

Remaining passengers only have to wait when the arriving train arrives early. The difference with transfer passengers is that remaining passengers are seated while waiting. Only the evaluation weight is different:

$$C_{st}(B) = 1.5 \left[B + \lambda \left(e^{-\frac{1}{\lambda}B} - 1 \right) \right],$$

where 1.5 is the weight for seated passengers' waiting.

3.2.3. Late arrival

In this case the cost depends on the number of minutes arriving passengers arrive later than scheduled. When arriving late $d > B$, and integration allows us to calculate the corresponding cost

$$C_{al}(B) = 2.5 \int_B^\infty (d - B) \frac{1}{\lambda} e^{-\frac{1}{\lambda}d} dd = 2.5 \lambda e^{-\frac{1}{\lambda}B},$$

where 2.5 is the weight for arriving late.

3.2.4. Generalised waiting cost as a function of the buffer time

For the total cost, each component is multiplied by the number of passengers: P_t = transfer passengers; P_r = remaining passengers; P_a = arriving passengers. This produces, for every connection,

the following generalised waiting cost as a function of B :

$$C(B) = 2.2P_t p e^{-\frac{1}{\lambda} B} + (2.0P_t + 1.5P_r) \times \left[B + \lambda \left(e^{-\frac{1}{\lambda} B} - 1 \right) \right] + 2.5P_a \lambda e^{-\frac{1}{\lambda} B}. \quad (1)$$

Fig. 3 shows the costs in (1) for an example where $\lambda = 2$, $p = 60$, $P_t = 119$, $P_r = 1745$ and $P_a = 3490$.

As can be seen from Fig. 3, the generalised waiting cost is minimised with an ideal buffer time close to 4 minutes (the minimum obtained by differential calculus is 3.84 minutes). The cost for transfer passengers is rather low because of their small number. The convex shape of the curve is typical; it is caused by the two types of waiting that transfer passengers may face. A small buffer will result in a large probability of missing a connection and arriving late, while effecting a small extra waiting time for remaining passengers and transfer passengers. A large buffer will result in a small chance of missing a connection or arriving late, but effects a large extra waiting time for remaining and transfer passengers. As can be seen from

Fig. 3, the total cost curve is not symmetrical, but a piecewise linear function can be a good approximation.

3.3. Results for the belgian network

The calculated buffer times are given in Table 2; they are rounded to whole numbers.

The numbers of passengers used in this example are similar to the real numbers, they give an idea of the situation during the morning rush hour. The real numbers cannot be disclosed because of confidentiality. For every connection, the minimum transfer time is set to 3 minutes. When passengers arrive less than 3 minutes before the departure of the connecting train, they will miss their connection, because they need 3 minutes for the actual transfer.

3.4. Summary

Thus, ideal buffer times can be easily calculated for each connection on the basis of following components:

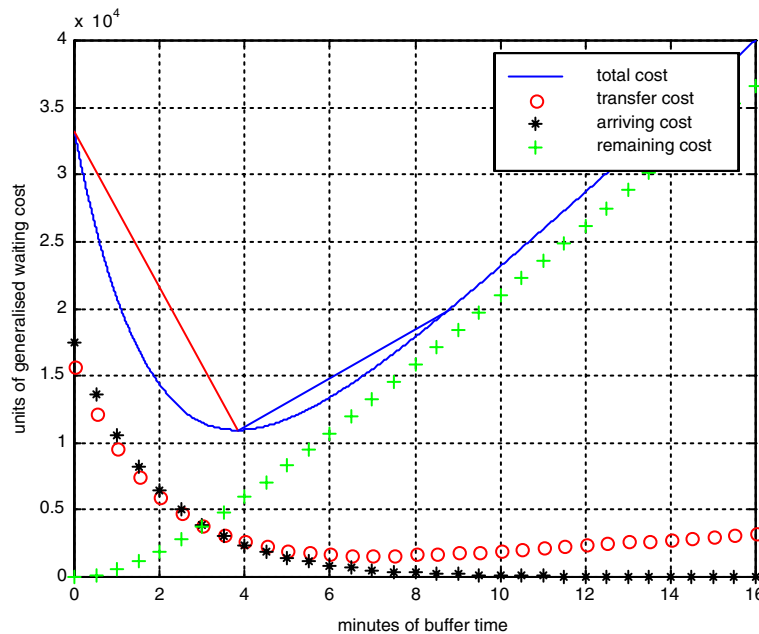


Fig. 3. Generalised waiting cost as a function of buffer time at a connection.

Table 2
Calculated buffer times

Connection	Number of passengers			Expected delay (λ ; minutes)	Minutes until next train	Ideal buffer (minutes)	Rounded buffer (minutes)
	Transfer	Arriving	Remaining				
K1–C1	121	8422	4227	3	30	4.58	5
K1–C0	8422	8422	4227	3	60	8.66	9
C0–K0	119	3491	1745	2	60	3.84	4
K0–E0	152	14,474	43,423	3	60	1.51	2
K0–M0	136	14,474	43,423	3	60	1.50	2
E1–K1	298	7718	23,155	4	30	2.09	2
M1–K1	245	4890	0	5	30	7 ^a	7 ^a
M1–K0	14,671	4890	0	5	30	10.40	10

^a This buffer has been reduced manually. In this special case, there are very few transfer passengers and no remaining passengers. As a result, the buffer time tends to be out of proportion.

- the number of minutes before the next connecting train arrives;
- the different weights for the value of time;
- the expected delay;
- the average number of remaining, transfer and arriving passengers on a regular day.

For each connection, adding the ideal buffer time to the ideal running time will minimise the waiting cost for that connection. However, due to conflicting tendencies, it will not be optimal for the whole network to include these ideal buffer times. Therefore, in the next section, goal programming will be used to meet these ideal buffers as closely as possible and minimise the generalised waiting cost for the whole network at once.

4. New timetable by means of linear programming

Linear goal programming can assist the decision maker in adding to the ideal running times (see Fig. 2) scheduled buffer times that meet the ideal buffer times as closely as possible. In this section, a new timetable is developed by means of LP for the time span of four subsequent hours. During this time window, the entire schedule of three trains, travelling in both directions on every line, can be determined. The start and end of the time window is not important. The timetable is completely cyclic (period of 1 hour) and the 4 hours can be easily extended to a daily, cyclic timetable.

4.1. Preliminary decisions

For the problem to remain linear, some preliminary decisions ought to be made. What will be the sequence of the trains over the single track between Alken and Landen? Also: in Hasselt, will the first K train connect to the first or to the second C train?

For the connection issue, it is necessary to look at the three trains of every line in both directions. For the K train this means: K10–K20–K30, i.e. the 0 direction, and K11–K21–K31, the 1 direction. The following rule is applied to determine which train of each line will connect with which train of the other line:

When a train, that has finished its trip through the network, connects with another train, that has finished its trip through the network, the connection will take place between trains bearing the same number: thus the first in sequence with the first in sequence on the other line (1–1); the second with the second (2–2); etc. (For instance, in Hasselt, train K11 will connect with train C10). If a train that has finished its trip through the network connects with a train that starts its trip, the number of the departing train will be one unit higher (1–2; 2–3). (For instance, in Hasselt, train K11 will connect with train C21.)

With this rule the first trains of each line can start during the first hour of the considered time window. The sequence of connections in Hasselt and Leuven is shown in Table 3.

Table 3
Connections in Hasselt and Leuven

Hasselt	Leuven
K11–C21; K21–C31	K10–E20; K20–E30
K11–C10; K21–C20; K31–C30	K10–M20; K20–M30
C10–K20; C20–K30	E11–K21; E21–K31
	M11–K21; M21–K31
	M11K10; M21K20; M31K30

For the single track issue, either the 0 direction K train precedes the 1 direction K train, or the 1 direction runs before the 0 direction. If necessary, the sequence can be switched in Sint-Truiden. Indeed, in Sint-Truiden station trains can cross one another. In total, there are four possibilities. To determine the best arrangement, all possibilities can be easily checked by linear programming. For more complex situations, 0–1 variables should be introduced. The best results were obtained when the 0 direction runs before the 1 direction; this sequence is not altered in Sint-Truiden.

4.2. Constraints

Table 4 shows, in three columns, the names of the different types of constraints, examples of formulations and additional explanations. The first three types of constraints are straightforward. The fourth type of constraints guarantees a feasible connection between an arriving train and a

connecting train. When the scheduled transfer time is 5 minutes, the connecting train must leave exactly 5 minutes after the scheduled arrival of the arriving train. The fifth spaces out evenly the departure of trains (of different lines) with the same destination. The time distance between two such trains, the spacing time, is set at a value between 20 and 40 minutes. The single track also requires some extra constraints. A particular sequence between the trains in opposite direction is fixed. Furthermore single track buffers (of 5 minutes) are included to ensure that small disturbances do not affect the sequence or cause major delays.

In the LP, the scheduled arrival, departure, transfer, stopping and buffer times are the variables. The ideal running times are fixed by the NMBS and the single track buffer is a parameter to be set.

Other parameters to be set are the minimum and maximum value of the stopping, spacing, transfer and buffer times. The range of values used, are presented in Table 5; they guarantee an appropriate service. Stopping times, for instance, may never be longer than 7 minutes. In order to determine a feasible solution, the scheduled transfer time in some cases should be longer than 20 minutes because of the spacing out requirement of trains. Thus, it is not useful to give an upper bound to transfer times.

Table 4
LP constraints

Constraint type, enforcing	Example	Explanation
Scheduled running times	$AC10Ha = DC10Aa + IRCAaHa + BC10AaHa$	The scheduled arrival of C10 in Hasselt = the scheduled departure of C10 in the previous station + the ideal running time + the scheduled buffer
Cyclic schedule	$DC20He = DC10He + 60$	The scheduled departure of C20 in Heist = the scheduled departure of the previous train (C10) in Heist + 60 minutes
Departure during first hour	$DC10He < 60$	The scheduled departure of C10 in its first station happens in the first 60 minutes
Connections	$DK21Le = AE11Le + TTE11K21$	The scheduled departure of K21 = the scheduled arrival of E11 + the scheduled transfer time between E11 and K21
Time spacing	$AE10Ha > AC10Ha + SMIN$	The scheduled arrival of E10 is later than the scheduled arrival of C10 + minimal spacing
Single track control	$DK11St = DK10Al + ICKAlSt + BK10St$	The scheduled departure of K11 in Sint-Truiden = the scheduled arrival of K10 + a single track buffer

Table 5
Minimum and maximum values of times (minutes)

Type of time	Minimum value	Maximum value
Scheduled buffer time	0	10
Stopping time	1	7
Spacing time	20	40
Scheduled transfer time	3	

Finally, in order to make sure that not too many people miss their connection, some of the transfer times can be given a higher minimum value. This will not affect the total cost too much, but it will provide a better service, because less people miss their connection. A good minimum value can be found by sensitivity analysis and/or by solving the linear programme for different possible values.

4.3. Objective function

The overall generalised waiting cost is minimised. This objective function is NOT identical to expression (1) of 3.2.4, but incorporates all types of waiting costs encountered in all network stations, during the entire time frame. More precisely, it comprises three classes of costs.

The first, most important component of the objective function represents the deviations from the ideal buffer times. Thus both, going over and under the ideal time are penalised. This is the goal programming part, described before. To determine the weight of the deviations, the generalised waiting cost function (1) is approximated linearly around the ideal buffer time (see Fig. 3).

The second component is related to the stopping times. Some trains have to wait for a while before departure. These stopping times are required to obtain a feasible schedule that guarantees all the necessary connections. The weights of these times are determined by the number of remaining passengers (Table 6) and the weight for waiting while seated in a train, 1.5.

Table 6
Number of remaining passengers in intermediate stations

	K0–K1	Station	C0–C1	M0–M1	E0–E1
Landen	5817	Aarschot	980	3002	18,024
Sint-Truiden	5109				
Alken	9054				

The third component concerns the transfer times. Sometimes, while creating a timetable, the scheduled transfer time for a certain connection will be extended to a period longer than the minimum transfer time of 3 minutes. As a result, the waiting cost for the transfer passengers will increase. On the other hand, in the event of a late train arrival, the extended transfer time decreases the chance of a missed connection. It is, however, very difficult to include this benefit in the objective function of the linear decision model. We prefer to omit it and will use simulation (Section 5) to precisely evaluate the generated timetables.

In conclusion, the composition of the LP objective function can be expressed as follows:

Overall generalised waiting cost

$$\begin{aligned}
 &= \text{cost of deviating from the ideal buffer times} \\
 &\quad + \text{cost of waiting in the stations} \\
 &\quad + \text{cost of extended transfer times.} \quad (2)
 \end{aligned}$$

It is certainly possible to formulate other useful linear objective functions. Furthermore, some types of waiting can be forced to remain within a given range of values.

4.4. LP results

The linear programming approach with objective function (2) produces the timetable of Table 7. All the necessary connections are guaranteed. The timetable is cyclic as expected.

The scheduled buffer and transfer times of the generated timetable are shown in Table 8. The buffers deviate not much from the ideal buffer times of Table 2. The transfer times are scheduled much more tightly than in the current timetable (Table 1). For example, to transfer from train C0 to train K0 a passenger only has to wait 6 minutes. Indeed, train C0 arrives in Hasselt 4 minutes past the hour and train K0 leaves Hasselt 13 minutes past the hour (and a passenger needs 3 minutes to transfer). With the current timetable, passengers have to wait 48 minutes. These results are further discussed in Section 5.

Since some approximations have been introduced, the LP approach described in this section

Table 7
The LP timetable

	A	D	A	D	A	D	A	D	A	D	A	D	A	D
	Heist-op-den-Berg		Aarschot		Leuven		Landen		Sint-Truiden		Alken		Hasselt	
C0	–	15	24	28	–	–	–	–	–	–	–	–	64	71
M1	–	35	47	48	63	–	–	–	–	–	–	–	–	–
K1	–	–	–	–	–	10	37	38	47	48	56	57	68	–
	Hasselt		Alken		Sint-Truiden		Landen		Leuven		Aarschot		Heist-op-den-Berg	
C1	–	30	–	–	–	–	–	–	–	–	56	57	66	–
M0	–	–	–	–	–	–	–	–	–	19	33	34	46	–
K0	–	13	19	20	28	29	38	45	74	75	–	–	–	–
	Leuven		Aarschot		Hasselt									
E0	–	46	58	59	84	–	–	–	–	–	–	–	–	–
	Hasselt		Aarschot		Leuven									
E1	–	2	27	28	42	–	–	–	–	–	–	–	–	–

Table 8
Scheduled buffer and transfer times

Connection	K1–C1	K1–C0	C0–K0	K0–E0	K0–M0	E1–K1	M1–K1	M1–K0
Scheduled buffer time	5	5	10	2	2	2	1	1
Scheduled transfer time	22	3	9	32	5	28	7	12

does not completely guarantee an optimal solution with respect to all (generalised) waiting costs.

1. The most important approximation was already mentioned: the potential benefit of an extended transfer time is not included.
2. In the ideal buffer calculation, when a connecting train is missed, it is assumed that the next train will always arrive precisely 30 or 60 minutes later. In the LP formulation (and in any realistic timetable), this is only true in an approximate way. Furthermore, the passengers will not have to wait the whole 30 or 60 minutes, but 30 or 60 minutes minus the late arrival time of the train, as was discussed in 3.2.1.
3. The deviation costs from the ideal buffer costs are approximated by linear functions.

Discrete event simulation will make it possible to dispose of these three types of approximations and to obtain a more accurate picture of all the

waiting costs. Furthermore, simulation allows us to improve the LP formulation. Indeed, after simulation, it may emerge that certain LP constraints should be slightly adjusted or new constraints should be added. Within very short computation time, the modified LP can be solved again. For the small case network, the minimum transfer time for the most important connection (M1–K0) was determined this way. The lowest generalised waiting cost was obtained with a minimum transfer time of 12 minutes rather than with the initially calculated 3 minutes.

5. Evaluation by means of discrete event simulation

One of the main contributions of simulation is the possibility to accurately calculate, in a single run, a variety of performance measures for a proposed railway timetable. Some useful performance criteria can be seen as column headings of Table 9. For other performance criteria: see [4].

Table 9
Performance measures

Measure timetable	Overall generalised waiting cost	Percentage of transfer times longer than 20 minutes	Percentage of missed connections	Total transfer cost	Total delay
Current	4.78E+6	45%	3.2%	3.40E+6	47
LP	2.86E+6	15%	15.5%	1.95E+6	24
LP JIT	2.43E+6	14%	15.5%	1.77E+6	24

A discrete event simulation model has been developed in Matlab[®] [13] for the small Belgian railway case using the weights for the value of time, the delay distributions and the passenger numbers.

During simulation, for every trip to Leuven or Hasselt, the stations where transfers take place, the running time was set equal to the sum of the ideal running time and a simulated delay. For determining fairly precise performance measures, for each timetable, averages were computed over one thousand simulation runs. The computation times were extremely small because of the size of the network.

In the previous sections two timetables were briefly discussed: the current timetable and the LP generated timetable. In the interest of passenger service, it is perhaps useful to introduce a just-in-time (JIT) principle in scheduling and to test it in LP timetables. JIT implies that no train will arrive early. Since one minute waiting is rated as 2 minutes of driving, the total waiting cost may further decrease.

Table 9 shows the key results of the simulation experiments: the scores of three timetables on five different performance criteria. LP JIT is not different from the LP timetable. Only the train arriving mode is different. The superiority of the LP approaches on nearly all the criteria is what strikes one most in Table 9. Notice also that the performance measures viewed apart do not mean much, but together they are very informative.

The overall generalised waiting cost is a measure for the total waiting cost of all passengers travelling through the network. It is an obvious, general service related criterion. The generalised waiting cost consists of the weighted cost of late arrivals for the arriving passengers, the weighted cost of early arrivals for the remaining passengers and the weighted cost of waiting for the transfer

passengers. The improvement of the LP timetable on the current one represents 40%. The most important component contributing to this improvement is the reduced waiting time of the transfer passengers. The connections are scheduled much better and thanks to the buffers, not too many connections are missed. This major improvement clearly illustrates the possibilities of the LP approach for designing and improving customer friendly timetables. Small changes to the weights do not significantly modify the LP timetable or the overall generalised waiting cost score. For instance, changing all weights to the value one (and thus minimising the actual waiting time of all passengers), still produces a timetable with a 41% reduction in cost. One may conclude that the results are rather robust and not very sensitive to weight changes.

The second performance measure of Table 9 is the percentage of transfer passengers who endure a transfer time of at least 20 minutes. This can be either because the scheduled transfer time is longer than 20 minutes or because passengers missed their connection. The scores prove that the LP approach can handle the transfers effectively.

The percentage of missed connections indicates how many passengers, on average, miss their connection and have to wait about 30 or 60 minutes. In the LP timetable the percentage increases to 15.5%, a consequence of the better transfer planning in general. Although 15.5% is small, missing a connection always reflects badly on the railway service level since it is such an unexpected event. Thus, a considerable rise in the percentage of missed connections cannot be defended, even if total service is improved. Solving real-life problems, one will always be forced to make difficult trade-offs between conflicting objectives. In the small network case, obviously a compromise must

be made between a smaller number of missed connections (the current timetable) and an overall improved transfer planning (the LP timetable). The LP modelling frame and simulation are very suitable for making such compromises. To limit explicitly the percentage of missed connections by adding or changing a constraint in the LP formulation is not possible. However, interaction between simulation and LP output will allow the decision maker to reach good compromises between the number of missed connections and an overall reduction of waiting time. The simulation process will furthermore clearly indicate what transfers are particularly important and sensitive. An other possibility to compromise is omitting the upper bound on the running time buffers. This improves the generalised waiting cost and reduces the percentage of missed connections from 15% to 5%. In general, it is easy to manipulate the upper bounds on waiting times and costs or, alternatively, to place more importance on specific costs in the objective function.

The fourth performance measure only considers transfer passengers related costs. Typically, the LP timetable has many small transfer times and, together with the buffers, this improves the transfer cost. The reduction amounts to 43%!

Total delay is the total time, in minutes, that trains arrive later than scheduled. The numbers of passengers on the trains are not considered. Total delay improves by 49% in the case of the LP timetable.

This section demonstrates how simulation can be one of the most useful tools for the evaluation of timetables. LP modelling in combination with simulation appears to be an excellent match. Simulation can also be used to assess the effect of improved train delay distributions and the impact of synchronisation policies; topics outside the scope of this paper.

6. Conclusions and recommendations

The approach presented in this paper appears to produce good results. Firstly, ideal buffer times are calculated for each transfer. Then these buffers are utilised in a linear programme where a general-

ised waiting cost function is minimised. In the small network case, the resulting timetable has a waiting cost 40% lower than the current timetable, a considerable saving. The bulk of this improvement is due to better-scheduled connections. The scheduled buffers assure that not too many of these connections are missed.

The 40% general cost reduction is a very promising result with a view to design acceptable, improved timetables. Indeed, 40% is large enough to allow satisfactory compromises between conflicting features: while the general waiting cost is significantly reduced, other—minor—waiting costs and risks are prevented to escalate. A preliminary study on the Belgian Intercity network [15] makes it apparent that the 40% reduction is not typical for the small network of this paper, but can, most likely, be achieved for the entire Belgian network.

It appears that the proposed LP modelling is very suitable indeed for producing well-balanced, realistic timetables which guarantee an ameliorated service. The approach makes genuine transfer management possible. Suitable LP software is everywhere readily available; LP modelling skills are widely understood and applied; large problem instances can be solved quickly. Input data management, however, may form a problem for large railway networks. For instance, calculating or updating the number of passengers of each type, in every station, can be cumbersome. It is therefore advisable to focus on the most important stations, train lines and connections when extensive systems are studied.

The proposed approach is inherently an interactive one: the coefficients and parameters (such as the maximum stopping time e.g.) are repeatedly modified by the decision maker and several reruns are required. The very flexible features of linear programming and the various possibilities for sensitivity analysis are major assets. Yet, the LP frame cannot deal with all scheduling concerns. As a consequence, LP modelling should go together with discrete event simulation. Simulation easily produces, at a reasonable computation cost, realistic evaluations of proposed timetables.

Further research issues could include: better delay distribution modelling (especially analysing peak and off-peak disturbances); the planning of

extra passenger and freight trains; the impact of synchronisation policies and the assumption that connecting trains always leave on time.

Acknowledgements

Comments by two anonymous referees and by W. Herroelen, professor at the Faculty of Economic and Applied Economic Sciences, Catholic University Leuven, are gratefully acknowledged. This research has been partly financed by the Flemish Scientific Research Foundation (Fonds voor Wetenschappelijk Onderzoek).

References

- [1] P. Bollen, LP-formulation for the routing problem at the Belgian Railway Company, Thesis, Faculty of Economic and Applied Economic Sciences, Catholic University Leuven, Belgium, 1997 (in Dutch).
- [2] A.J. Davenport, C. Gefflot, J.C. Beck, Slack-based techniques for robust schedules, in: *Proceedings of the Sixth European Conference on Planning*, 2001.
- [3] O. Engelhardt-Funke, M. Kolonko, Analysing stability and investments in railway networks using Advanced Evolutionary Algorithms, Technical University Clausthal, Germany, 2002.
- [4] R.M.P. Goverde, Optimal scheduling of connections in railway systems, Paper Presented at the Eighth WCTR, Antwerp, Belgium, 1998.
- [5] R.M.P. Goverde, Synchronization control of scheduled train services to minimize passenger waiting times, in: *Proceedings of the fourth TRAIL Annual Congress, Part 2*, TRAIL Research School, Delft, The Netherlands, 1998.
- [6] R.M.P. Goverde, Transfer stations and synchronization, Delft University of Technology, The Netherlands, 1999.
- [7] H.F. Gunn, J.G. Tuinenga, Y.H.F. Cheung, H.J. Kleijn, Value of Dutch travel time savings in 1997, in: *Proceedings of eighth WCTR*, Antwerp, Belgium, 1998, pp. 515–526.
- [8] A. Higgins, E. Kozan, Modeling train delays in urban networks, *Transportation Science* 32 (4) (1998) 346–357.
- [9] J. Istas, J. Viskovic, Opportunities for improving the reactive scheduling at the Belgian Railway Company, Master Thesis, Centre for Industrial Management, Catholic University Leuven, Belgium, 2003.
- [10] L. Kroon, L. Peeters, A variable trip time model for cyclic railway timetabling, *Transportation Science* 37 (2) (2003) 198–212.
- [11] C. Liebchen, R. Möhring, A case study in periodic timetabling, *Electronic Notes in Theoretical Computer Science* 66 (6) (2002).
- [12] C. Liebchen, L. Peeters, Some practical aspects of periodic timetabling, in: P. Chamoni et al. (Eds.), *Operations Research 2001*, Springer, Germany, 2002.
- [13] Matlab®, Version 5.2, The MathWorks, Inc., 1998.
- [14] Y. Meng, Bemessung von Pufferzeiten in Anschlüssen von Reisezügen, in: W. Schwanhäuß, P. Wolf (Eds.), *Veröffentlichungen des Verkehrswissenschaftlichen Institutes der Rheinischen-Westfälischen Technischen Hochschule Aachen*, Heft 46, Aachen, Deutschland, 1991 (in German).
- [15] H. Pham Ngoc, Intercity train timetable with better passenger service, Master Thesis, Centre for Industrial Management, Catholic University Leuven, Belgium, 2004.
- [16] J. Shires, D. Johnson, C. Nash, C. Tyler, Appraisal framework and results for testing a regular interval rail timetable, Working Paper 578, Institute for Transport Studies, University of Leeds, 2003.
- [17] K.A. Small, Urban transportation economics, *Fundamentals of Pure and Applied Economics*, vol. 51, Harwood Academic Publishers, Chur, Switzerland, 1992.
- [18] M. Wardman, Public transport values of time, ITS Working Paper 564, University of Leeds, UK, 2001.
- [19] M. Wardman, J. Hine, S. Stradling, Interchange and Travel Choice, vol. 1, Scottish Executive Central Research Unit, Scotland, 2001.
- [20] M. Wardman, J. Shires, W. Lythgoe, J. Tyler, Consumer benefits and demand impacts of regular train timetables, *International Journal of Transport Management* 2 (1) (2004) 39–49.