Incorporating passenger waiting times in timetabling

Gert-Jaap Polinder · Marie Schmidt · Dennis Huisman

Abstract The regularity of train services is often associated with the quality of a timetable and thus enforced by synchronization constraints when building a timetable. In this paper we investigate to what extent such synchronization constraints prevent us from finding optimal timetables. The objective we minimize is total passenger travel time, which is composed of in-train time, but also waiting time at the origin station. To this end, we introduce an integer programming model that finds an optimal periodic timetable with respect to our objective function. First computational results on the Dutch railway network show that for some lines, deviating from a regular pattern can improve the total passenger travel time considerably.

Keywords Public Transport · Passenger Routing · Timetabling · Waiting Times

G.J. Polinder
Rotterdam School of Management,
Erasmus Center for Optimization in Public Transport,
Erasmus University, Rotterdam, The Netherlands
Tel.: +3110-4088041
E-mail: polinder@rsm.nl

M. Schmidt
Rotterdam School of Management,
Erasmus Center for Optimization in Public Transport,
Erasmus University, Rotterdam, The Netherlands
E-mail: schmidt2@rsm.nl

D. Huisman
Econometric Institute,
Erasmus Center for Optimization in Public Transport,
Erasmus University, Rotterdam, The Netherlands
Process Quality and Innovation
Netherlands Railways, Utrecht, The Netherlands
E-mail: huisman@esc.eur.nl
1 Introduction

In many countries, passenger railway companies operate according to a periodic timetable, that is, the timetable repeats every time period, usually every hour. In the design of such a timetable, (technical) requirements of the line plan and the underlying infrastructure have to be taken into account. However, often the timetable has to satisfy a certain set of service level requirements as well, which can severely limit the solution space, even to the extent that no feasible solution exists any more. Sometimes, small changes to constraints in the model can repair the infeasibility (Polinder et al, 2017), but often larger changes are needed.

In this paper we investigate to what extent synchronization requirements (among departures of trains within one time period) as well as other service requirements lead to sub-optimal timetables. Service requirements are often added to a timetable model, to improve the quality of a timetable. We investigate the influence of this on the quality of the timetable. In order to do this, we look at the total time passengers spend travelling. This time is composed of two parts, the time they spend waiting at their origin station, and the the actual travel time, plus a possible penalty for making a transfer. The rationale behind this approach is that synchronization constraints for departures of trains of the same line were originally introduced to avoid long waiting times at origin stations. By including waiting times at origin stations into our objective function, we can trade-off slightly longer waiting times for other improvements in the timetable.

The routes that passengers use to reach their destination, depend on the timetable. However, in order to find a good timetable, we need to know the passenger routes, so these two items are depending on each other. In order to find an overall good solution, we integrate the timetabling and passenger routing into one Integer Programming (IP) model.

The Integer Programming (IP) model that we propose integrates timetabling and passenger routing, but with respect to current approaches also explicitly takes waiting times at the origin station into account. We assume, similar to de Palma and Lindsey (2001); Kaspi and Raviv (2013), that passengers arrive at the origin station according to a uniform distribution, and want to reach their destination as soon as possible. In the IP model, passengers are assigned to routes (out of a given set of possible routes), and they are assigned to the routes in such a way that the total time they are travelling is minimized. The travel time of each route is depending on the timetable, as is the waiting time of the passengers. Therefore, this model plans the trains and assigns passengers to them in such a way that the total travel time of all passengers is minimized.

In our computational results, we see that even if we simplify the IP model by only consider direct routes for all passengers, we can find improvements over a synchronized timetable. Furthermore, we investigate what the effect is of using service requirements, or leaving the model completely free. On what parts of the network is synchronization of trains important? And what is the effect of this on the timetable and the time passengers spend traveling?
2 Literature overview

An overview of the models that can be used in periodic timetabling is given in Caimi et al (2017). They say that the state-of-the-art modelling approach in periodic timetabling is Periodic Event Scheduling (Serafini and Ukovich, 1989). This problem can be formulated as follows: Given a set of periodically reoccurring events $V = V^{arr} \times V^{dep}$ (the arrivals $V^{arr}$ and departures $V^{dep}$ of the trains), a common time period $T$, and a set $A$ of constraints, find a periodic schedule $\pi : V \rightarrow [0, T]$ satisfying all constraints. Each constraint states that the difference in time between two events should be within a certain $T$-periodic interval and they can be written in the form

\[ l_{ij} \leq \pi_j - \pi_i \mod T \leq u_{ij}. \]

This formulation is very versatile in the sense that it can model hard timetabling constraints like lower bounds on trip and dwell times and safety requirements, as well as constraints introduced to ensure a certain service level, like upper bounds on driving and dwell times, bounds on transfer times, and regularity of departures of the same train line (within one period). Liebchen and Möhring (2007) give an overview of constraints that can be modelled in the PESP modelling framework. Several algorithms exist to solve general PESP instances (cf. Schrijver and Steenbeek (1993); Großmann et al (2012)). These approaches do not take into account some objective function, although in the SAT context it is possible to include an objective function (cf. Gattermann et al (2016)).

The integration of passenger routing and timetabling has been studied before. There are two main versions:

- Solve the passenger routing and timetabling iteratively (cf. Siebert and Goerigk (2013); Kaspi and Raviv (2013); Robenek et al (2016)). In this approach, a timetable is found based on some precomputed passenger routes. Next, the passenger are rerouted based on this timetable, and the timetable is updated again.

- Solve the passenger routing and timetabling in one single mathematical model. In de Palma and Lindsey (2001), the authors consider a very stylized network, with only one connection, in which passengers arrive according to some continuous distribution at the origin station, and busses have to be scheduled in such a way that the total delay that passengers experience is minimized. For some simplified cases, analytical solution can be provided. The assumption of passengers arriving at the origin station at a rate which is independent of the actual timetable is also taken by Kaspi and Raviv (2013). For non-periodic timetabling, an extensive analysis is given by Schmidt and Schöbel (2015). The authors consider several cases of routing and timetabling, and show for which situations an efficient algorithm exists to solve the problem. Siebert and Goerigk (2013) also provides a model in which the two problems are integrated into one, and can solve it for small instances. Borndörfer et al (2017) formulate a model that integrates passenger routing and timetabling as well in one single model.
They precompute several paths between a departure and an arrival node, of which the passenger has to choose one, and optimize the timetable simultaneously.

3 IP model for integrated timetabling and passenger routing with waiting times

The IP model that we propose, consists of a timetabling part and a passenger routing part. The timetabling part of the model that we are using is based on PESP (Serafini and Ukovich, 1989). We are given a set of events \( V \) and constraints \( E \). This PESP instance can be represented by an Event-Activity-Network EAN. For each event \( v \in V \), we have the timetabling variable \( \pi_v \) and for each activity \( a \in A \) we have the duration variable \( y_a \), defined as

\[
y_a = \pi_j - \pi_i + T\beta_{ij} \quad \forall a = (i, j) \in A,
\]

where \( \beta_{ij} \) is an integer accounting for the shift of one cycle period to the next.

Such an activity can be waiting or travelling with a train, or it can model a transfer or safety constraint between two trains.

In order to properly the way in which we integrate this timetabling model with passenger routing, we introduce some notation.

3.1 Notation

First of all, we are given a set of station \( S \). Passengers want to travel from one station to another, this travel demand is given in the form of an Origin-Destination matrix \( OD \in \mathbb{R}^{\mid S \times S} \). For each OD-pair \( k \in OD \), the demand, i.e. the number of passengers, is denoted by \( d_k \).

For each OD-pair \( k \in OD \), there are several paths available that the passengers can use to get to the destination. For OD-pair \( k \in OD \), this set of possible paths is denoted by \( P^k \). The total set of paths in the network is denoted by \( P \), i.e.,

\[
P = \bigcup_{k \in OD} P^k.
\]

For our IP-model, we assume these sets of paths is given. They can for example be computed based on the given line plan. How this path is generated in our experiments, is explained in the last section.

Each path \( p \in P \) is a route through the network from event to event, through a set of arcs, i.e., \( p = (e_1, a_1, e_2, a_2, \ldots, a_n, e_n) \). The first event of the path is denoted by \( j(p) = e_1 \), which always is a departure event. The total set of departure events that OD-pair \( k \in OD \) can use, is denoted by \( V^k \subseteq V^{dep} \), and can be determined as

\[
V^k = \bigcup_{p \in P^k} j(p).
\]
Each path has a certain duration, which is the time difference between the departure time until the arrival time, or equivalently, it is the sum of all the individual activity durations, plus a possible transfer penalty $\gamma_t$:

$$Y_p = \sum_{a \in P} \gamma_r \cdot y_a + \gamma_t \cdot 1_t(a),$$

with $1_t$ an indicator function:

$$1_t(a) = \begin{cases} 
1 & \text{if } a \text{ is an transfer activity} \\
0 & \text{else} 
\end{cases}.$$  

$\gamma_t$ is the in-train time coefficient.

Each path for OD-pair $k \in \mathcal{OD}$ starts with some departure node $v \in V^k$. Multiple paths can start with the same node, this set of paths is denoted by $P^k_v \subseteq P^k$. We assume that out of these paths, at most one of these paths is the best to take, i.e., not two paths will depart and arrive at the same time. The duration of the actual path that is taken, when departing with departure node $v$, is denoted by $Y^k_v$ and will be the minimum of all $Y_p$, with $p \in P^k_v$, i.e.,

$$Y^k_v = \min_{p \in P^k_v} Y_p.$$  

However, there might be cases in which taking the next departing train is not beneficial, and waiting for a later train is better. In order to take this into account, let $Q_{v,v'} \geq 0$ be the time difference between events $v, v' \in V$, i.e.,

$$Q_{v,v'} = \pi_{v'} - \pi_v + \alpha_{v,v'} T, \quad \alpha_{v,v'} \in \{0, 1\}$$

Next, let $\hat{Y}^k_v$ be the realized travel time, from the event time of node $v \in V^k$ onwards, until the passenger reaches his destination. This can be determined as

$$\hat{Y}^k_v = \min_{v' \in V^k} \left\{ \gamma_w \cdot Q_{v,v'} + \hat{Y}^k_{v'} \right\},$$

where $\gamma_w$ is the waiting time coefficient.

Next, let $A^k_v \in \{0, 1, \ldots, T\}$ be the number of minutes for which node $v \in V^k$ is the next departure node for OD-pair $k \in \mathcal{OD}$. This number can only be determined once a timetable exists, and can then be determined as

$$A^k_v = \min_{v' \in V^k \setminus \{v\}} \{Q_{v',v} \}.$$  

Since we clearly have that

$$\sum_{v \in V^k} A^k_v = T,$$

we can rewrite (10) as

$$A^k_v \leq Q_{v',v} \leq T - 1 \quad \forall v' \in V^k \setminus \{v\}.$$  

The expected time passengers have to wait before departure node $v \in V^k$ is denoted by $W^k_v$ and can be calculated as

$$W^k_v = \frac{1}{2} A^k_v,$$  

as the arrival distribution is uniform.
3.2 The IP model

The complete model we are using can now be summarized as follows:

\[
\text{Minimize } \sum_{k \in OD} \frac{d_k}{T} \sum_{v \in V_k} A_v^k \cdot \left[ \gamma_{w} \cdot W_v^k + \hat{Y}_v^k \right] \tag{14a}
\]

Such that \( Y_p = \sum_{a \in p} \gamma_r \cdot y_a + \gamma_t \cdot 1_t(a) \quad \forall p \in P \) \( (14b) \)

\[
\hat{Y}_v^k \leq Y_p + \gamma_w \cdot Q_{v,v'} \quad \forall k \in OD, v, v' \in V^k, p \in P^k_{v,v'} \quad (14c)
\]

\[
\hat{Y}_v^k \geq Y_p + \gamma_w \cdot Q_{v,v'} - M^k_{v,v'} \cdot (1 - z_{v,v',p}^k) \quad \forall k \in OD, v, v' \in V^k, p \in P^k_{v,v'} \quad (14d)
\]

\[
\sum_{v' \in V^k} \sum_{p \in P^k_{v,v'}} z_{v,v',p}^k = 1 \quad \forall k \in OD, v \in V^k \quad (14e)
\]

\[
y_a = \pi_j - \pi_i + T \beta_{ij} \quad \forall (i, j) \in A \quad (14f)
\]

\[
y_a \geq l_a \quad \forall a \in A \quad (14g)
\]

\[
y_a \leq u_a \quad \forall a \in A \quad (14h)
\]

\[
Q_{v,v'} = \pi_{v'} - \pi_v + T \alpha_{v,v'} \quad \forall k \in OD, v, v' \in V^k \quad (14i)
\]

\[
A_v^k \leq Q_{v',v} \quad \forall k \in OD, v, v' \in V^k \quad (14j)
\]

\[
\sum_{v \in V^k} A_v^k = T \quad \forall k \in OD \quad (14k)
\]

\[
W_v^k = \frac{1}{2} A_v^k \quad \forall k \in OD, v \in V^k \quad (14l)
\]

\[
\pi_v \in \{0, \ldots, T - 1\} \quad \forall v \in V \quad (14m)
\]

\[
\alpha_{v,v'}^k \in \{0, 1\} \quad \forall k \in OD, v, v' \in V^k \quad (14n)
\]

\[
\beta_a \in \mathbb{Z}_{\geq 0} \quad \forall a \in A \quad (14o)
\]

\[
z_{v,v',p}^k \in \{0, 1\} \quad \forall k \in OD, v, v' \in V^k, p \in P^k_{v,v'} \quad (14p)
\]

\[
Q_{v,v'} \in [0, T - 1] \quad \forall k \in OD, v, v' \in V^k \quad (14q)
\]

The objective (14a) minimizes the total travel time of all passengers. The number of minutes for which \( v \) is the next departure node, is multiplied by \( d_v/T \) to determine the number of passengers, which is multiplied by the travel time variable. (14b) determines the actual length of each possible path in the
network. (14c) - (14e) determine the actual path duration from the departure
time of node $v$ onwards. (14f) - (14h) determine the length of each activity and
provide the correct bounds on the duration of the activity. (14i) determines
the actual time between two events. In (14j) and (14k), the correct values for
$A^k_v$ are determined, and linked to the waiting time in (14l).

4 Experiments

We have done numerical experiments on instances from Netherlands Railways
to estimate solution time and scalability of our basic Integer Programming
model. We have an instance ‘small’ (Figure 1), ‘A2-corridor’ (Figure 2) and
‘Full IC-network’ (Figure 3). The referenced figures show the infrastructural
network of these instances.

The first instance contains 4 train lines, 1 from each end-point in the south,
to each end-point in the north. Next to the train lines, the instance contains
12 stations and 132 OD-pairs.

The second instance consists of the so-called A2-corridor, a corridor be-
tween Maastricht and Den Helder in the Netherlands. Here we considered 6
train lines, all with a frequency of 2 times per hour. Recently, tests have been
performed to have six intercity trains per hour serving a part of this corridor,
in order to better satisfy demand.

The last instance is the whole intercity network of the Netherlands as it is
nowadays. It contains many train lines, all with a frequency of at least 2 times
per hour.

The IP-model we are using allows for all sorts of PESP constraints. We
however left out all headway constraints that are used to guarantee safety.
We aim for finding a structure of the timetable, in which infrastructure is not
yet fully known, so the headway constraints are irrelevant here. Furthermore,
the model contains no service requirement. The only aspect of service is the
objective function, which minimizes the total duration of the time passengers
are travelling.
The objective function is still quadratic, but can be easily linearized by introducing variables

\[ x^k_{v,d} = \begin{cases} 
1 & \text{if } A^k_v \geq d \\
0 & \text{else}
\end{cases}. \tag{15} \]

We have ran the IP-model for all three instances, for the situation in which we required trains of the same line to be synchronized, and where they were completely free. Then we compare the solution quality and computation time of the experiments.

For the small instance, the optimal solution in both cases was found within a minute. Requiring synchronisation gave a small speedup, as there is less freedom in the model. The solution showed that trains in regions with a lower passenger demand are not synchronised, in favour of the parts with a higher demand.

A partial time space diagram of this solution is shown in Figure 4. Time is shown on the horizontal axis and space on the vertical. The highest demand regions are in the south, which is shown on the lower part of the axis.

If the synchronisation is enforced at the starting stations of the journeys, an increase of the objective value of 1% is seen. To further analyse this, we distinguish between all trains running from north to south, and back. For the southward direction, enforcing synchronisation leads to a reduction in waiting times of 0.2%, but passenger in-train time is increased by 17.7% in total. This is because the northern region has less passenger demand and therefore trains are delayed on their way to run in a more regular pattern once they reach the higher demand region. For the northward direction, the waiting times were 1.8% larger, because trains are synchronised where this is not beneficial, and passenger in-train times are slightly reduced here. Modifying this instance by inserting additional trains in this high demand region (the south) still lead to a regular pattern there, while also changing the pattern to become more regular in other parts of the country where it used to be less regular.
For the second instance, the timetable without synchronization requirements shows that the best solution is to synchronize the trains in the high demand region. The regions with slightly lower demand, show a more irregular pattern. On the parts with only a few trains, synchronization again becomes important.

Finally, to test the scalability of the model, we computed a timetable for the complete Dutch railway network, based on a current line plan. We have put a time limit of one hour to solve this model and in this time we have obtained a solution with a proven optimality gap which is slightly smaller than 7%. This shows that, also for large networks, we can still solve this model in short time to a reasonable optimality gap.

**Conclusions** So far, our results seem to indicate that in the case of low frequencies, synchronization of train services can lead to sub-optimal timetables with respect to passenger travel time. When frequencies on route segments increase, however, timetables tend to become more regular, as is seen in the first test case.

Even if we consider only direct connections for passengers, we can still achieve even better timetables, compared to the situation where synchronization is enforced. This leads to interesting insights how ‘expensive’ these constraints are for passengers.

With the IP-model that is provided, timetables can be computed that minimize the total time passengers spend waiting and travelling. The results lead to a structure of the timetable that is good for passengers, but might not necessarily be feasible with respect to headway constraints. At this moment, we are working on a method that turns a given timetable into a feasible timetable, based on a method by Cacchiani et al (2010).

**References**


