A Suitable Model for a Bi-criteria Optimization Approach to Railway Track Allocation

Supported by the Federal Ministry of Economics and Technology (BMWi), grant 19M4031A.

ZIB-Report 08-22 (March 2008)
A Suitable Model for a Bi-criteria Optimization Approach to Railway Track Allocation*

Ralf Borndörfer** Thomas Schlechte**

March 31, 2008

Abstract

Technical restrictions and challenging details let railway traffic become one of the most complex transportation systems. Routing trains in a conflict-free way through a track network is one of the basic scheduling problems for any railway company.

This article focuses on a robust extension of this problem, also known as train timetabling problem (TTP), which consists in finding a schedule, a conflict free set of train routes, of maximum value for a given railway network.

However, timetables are not only required to be profitable. Railway companies are also interested in reliable and robust solutions. Intuitively, we expect a more robust track allocation to be one where disruptions arising from delays are less likely to be propagated causing delays of subsequent trains. This trade-off between an efficient use of railway infrastructure and the prospects of recovery leads us to a bi-criteria optimization approach. On the one hand we want to maximize the profit of a schedule, that is more or less to maximize the number of feasible routed trains. On the other hand if two trains are scheduled as tight as possible after each other it is clear that a delay of the first one always affects the subsequent train.

We present extensions of the integer programming formulation in Borndörfer & Schlechte [2007] for solving (TTP). These models can incorporate both aspects, because of the additional track configuration variables. We discuss how these variables can directly be used to measure a certain type of robustness of a timetable. For these models which can be solved by column generation techniques, we propose so-called scalarization techniques, see Ehrgott [2005], to determine efficient solutions.

Here, an efficient solution is one which does not allow any improvement in profit and robustness at the same time. We prove that the LP-relaxation of the (TTP) including an additional \( \epsilon \)-constraint

---

*This work was funded by the Federal Ministry of Economics and Technology (BMWi), project Trassenbörse, grant 19M4031A.

**Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Takustr. 7, 14195 Berlin-Dahlem, Germany, Email {borndoerfer, schlechte}@zib.de
remains solvable in polynomial time. Finally, we present some preliminary results on macroscopic real-world data of a part of the German long distance railway network.

1 Introduction

Constructing a maximum number of train routes in a conflict-free way through a track network is one of the major scheduling problems a railway company has to face. From a complexity point of view this problem turns out to be \( \mathcal{NP} \)-hard, see Caprara et al. [2002], but nevertheless in the literature several optimization models were discussed which are able to solve real-life instances to near-optimality, see Caprara et al. [2001], Caprara et al. [2002], Borndörfer et al. [2006], Cacchiani et al. [2008] and Cacchiani [2007] for more details.

All of these articles model the track allocation problem in terms of a multi-commodity flow of trains in an appropriate time expanded digraph. Feasibility is ensured by additional packing constraints, which rules out conflicts between the routes.

In Borndöfer & Schlechte [2007] a different approach that handles conflicts not in terms of constraints, but in terms of additional variables, was introduced. Its path formulation is amenable to standard column generation techniques and therefore suited for large-scale computation. The major contribution of this paper is that these additional variables can be used to measure robustness in terms of available buffer times of a timetable.

Robust optimization - that means the incorporation of data uncertainties through mathematical models in its pure definition as proposed by Soyster [1973] are not applicable for large scale optimization problems, yet. Moreover these models will end up with too conservative solutions; resistant against all considered eventualities - but far away from implementable in real world.

However robust optimization becomes a fruitful field in the last years due to the fact that more and more optimization problems can be solved in adequate time. This opens the door to additionally deal with stochastic assumptions instead of only nominal given data. In Ben-Tal & Nemirovski [1998] and El-Ghaoui et al. [1998] less conservative models were introduced, which are able to adjust the robustness of the solution by some protection level parameters. In Bertsimas et al. [2007] a recent survey to robust optimization theory and its applications is given.

In Kroon et al. [2006], Liebchen et al. [2007] and Fischetti et al. [2007] these robust considerations applied to the world of railways. There, an cyclic version of the timetabling problem, modelled as Periodic Event Scheduling Problem, briefly PESP, is investigated and stochastic methodology of Light Robustness is introduced. Aim of all these considerations is to gain more
insights in the trade-off of efficiency and robustness of solutions; finding a so-called ‘price of robustness’.

We are focusing on a pure combinatorial bi-criteria optimization approach, which is somehow related to Ehrgott & Ryan [2002] and Ehrgott et al. [2007], broaching the issue of robustness in airline crew scheduling. Motivated by the fact that robustness (available buffer times, quality of day-to-day operations) and efficiency (used track kilometers, planned capacity utilization) are incomparable entities, we favor a bi-criteria optimization approach to cope with this.

We forbear from supporting this by recent statistics to punctuality and reliability of any railway company. But obviously, decision makers are more and more sensitive to the importance of finding a good compromise between profitable and reliable timetables.

The organization of this article is as follows, in Section 2 we briefly introduce a version of the train timetabling or track allocation problem. In Section 3 we recite corresponding linear programming formulation ACP using arc variables and PCP using path variables, respectively. Then, we extend these models to measure robustness - this leads directly to a bi-criteria optimization approach of the problem.

Section 3.1 discusses details on a straightforward solution approach for these kind of problems. We prove that the LP-relaxation of the PCP including an additional $\epsilon$-constraint remains solvable in polynomial time. To determine efficient solution of the bi-criteria models, we propose so-called scalarization techniques, see Ehrgott [2005].

Finally, we present some preliminary computational results on a part of the German Railway Network in Section 4. Let us point out explicitly that we do not claim these results are already practically significant; we only want to show the potential of our approach on real-world-data.

2 The Track Allocation Problem

The track allocation problem in its single objective version can be formally defined in terms of several digraphs $D = (V, A)$. By $\delta_{in}$ we denote the set of incoming arcs $a \in A$ for $v \in V$, by $\delta_{out}(v)$ the set of outgoing arcs, respectively. Arrivals and departures of trains at a set $S$ of stations at discrete times $T \subseteq \mathbb{Z}$ are represented by the nodes $v \in V$, arcs model activities like runs of trains between or turnovers and dwelling inside stations. For each $v \in V$ we denote by $\sigma(v) \in S$ the station associated with departure or arrival and by $\tau(v) \in T$ the time of this event; we assume $\tau(u) < \tau(v)$ for each arc $uv \in A$ such that $D$ is acyclic; denote by $J = \{\sigma(u)\sigma(v) : uv \in A\}$ the set of all railway tracks.

We are further given a set $I$ of requests to route trains through $D$. More precisely, train $i \in I$ can be routed on a path through some suitably defined
digraph $D_i = (V_i, A_i)$ from a source node $s_i \in V_i$ to a sink node $t_i \in V_i$; let $P_i$ be the set of all routes for train $i \in I$, and $P = \bigcup_{i \in I} P_i$ the set of all train routes (taking the disjoint union). An arc $uv \in A$ blocks the underlying track $\sigma(u)\sigma(v)$ during the time interval $[\tau(u), \tau(v) - 1]$, that two arcs $a, b \in A$ are in conflict if their respective blocking intervals overlap, and that two routes $p, q \in P$ are in conflict if any of their arcs are in conflict.*

A track allocation or timetable is a set of conflict-free routes, at most one for each train. Given arc weights $w_a$, $a \in A$, the weight of route $p \in P$ is $w_p = \sum_{a \in p} w_a$, and the weight of a track allocation $X \subseteq P$ is $w(X) = \sum_{p \in X} w_p$. The optimal track allocation problem (OPTRA) in its single objective version is to find a conflict-free track allocation of maximum weight.

OPTRA can be seen as a multi-commodity flow problem with additional packing constraints - usually ensured by inequalities in the models. In Borndöfer & Schlechte [2007], we proposed an alternative formulation that is based on valid ‘configurations’. These are sets of arcs on the same underlying track that are mutually conflict-free.

Formally, let $A_{st} = \{uv \in A : \sigma(u)\sigma(v) = st\}$ be the set of all arcs associated with some track $st \in J$; a configuration for this track $st$ is a set

*In reality train conflicts are not that easy. This is just a notational simplification avoiding the introduction of headway matrices.

---

<table>
<thead>
<tr>
<th>symbol</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>stations</td>
</tr>
<tr>
<td>$J$</td>
<td>tracks</td>
</tr>
<tr>
<td>$G = (S, J)$</td>
<td>infrastructure digraph</td>
</tr>
<tr>
<td>$I$</td>
<td>train requests</td>
</tr>
<tr>
<td>$w$</td>
<td>arc weights (profit)</td>
</tr>
<tr>
<td>$r$</td>
<td>arc weights (robustness)</td>
</tr>
<tr>
<td>$\sigma : V \mapsto S$</td>
<td>mapping of nodes to stations</td>
</tr>
<tr>
<td>$\tau : V \mapsto \mathbb{Z}$</td>
<td>mapping of nodes to time</td>
</tr>
<tr>
<td>$s_i, t_i$</td>
<td>source, sink of train $i \in I$</td>
</tr>
<tr>
<td>$D_i = (V_i, A_i)$</td>
<td>train digraph of $i \in I$</td>
</tr>
<tr>
<td>$s_{xy}, t_{xy}$</td>
<td>source, sink of track $j = xy \in J$</td>
</tr>
<tr>
<td>$D_j = (V_j, A_j)$</td>
<td>track digraph of $j \in J$</td>
</tr>
<tr>
<td>$P_i$</td>
<td>set of $s_i, t_i$ -paths in $D_i$</td>
</tr>
<tr>
<td>$Q_j$</td>
<td>set of $s_{xy}, t_{xy}$ -paths in $D_j$</td>
</tr>
<tr>
<td>$A_{LR}$</td>
<td>coupling arcs</td>
</tr>
<tr>
<td>$A_{RL}$</td>
<td>backward arcs</td>
</tr>
<tr>
<td>$A := A_I \cup A_J$</td>
<td>all arcs</td>
</tr>
</tbody>
</table>

Tab. 1: Notation for the Track Allocation Problem.
of arcs \( q \subseteq A_{st} \) that are mutually not in conflict, i.e. in our simplified case not overlapping. Let \( Q_j \) denote the set of all configuration associated with track \( j \in J \), and \( Q = \bigcup_{j \in J} Q_j \) the set of all configurations. By \( A_{LR} = \bigcup_{st \in J} A_{st} \), we denote the set of all ‘forward’ arcs.

For the construction of configurations we have to introduce track digraphs \( D_j = (V_j, A_j) \) on each track \( j \in J \). Consider the forward arcs \( A_{xy} = \{ uv \in A : \sigma(u)\sigma(v) = xy \} \) on a track \( j = xy \in J \). Denote by \( L_{xy} := \{ u : uv \in A_{xy} \} \) and \( R_{xy} := \{ v : uv \in A_{xy} \} \) the associated set of departure and arrival nodes. Construct two new, additional nodes \( s_{xy} \) and \( t_{xy} \) by setting \( \sigma(s_{xy}) = y \), \( \tau(s_{xy}) := \min \tau(R_{xy}) - 1 \), and \( \sigma(t_{xy}) = x \), \( \tau(t_{xy}) := \max \tau(R_{xy}) + 1 \), i.e., \( s_{xy} \) marks an artificial source node at station \( y \) before the departure of the earliest trip on \( xy \), and \( t_{xy} \) marks an artificial sink node at station \( x \) after the arrival of the latest trip on \( xy \). Let \( \overline{L}_{xy} := L_{xy} \cup \{ t_{xy} \} \) and \( \overline{R}_{xy} := R_{xy} \cup \{ s_{xy} \} \); note that all arcs in \( A_{xy} \) go from \( \overline{L}_{xy} \) to \( \overline{R}_{xy} \) (actually from \( L_{xy} \) to \( R_{xy} \)).

Now let \( A_{RL} := \{ vu : \tau(v) \leq \tau(u), v \in \overline{R}_{st}, u \in \overline{L}_{st} \} \) be a set of ‘backward’ arcs that go in the opposite direction; they connect the arrival of a trip on \( xy \) (or node \( s_{xy} \)) with all possible follow-on trips (or node \( t_{xy} \)) on that track.

Table 1 summarized the notation and Figure 1 gives some intuition of the construction on a simple example. Finally, it is easy to observe, that per definition

- each train digraph \( D_i \) is acyclic
- each track digraph \( D_j \) is acyclic and bipartite
- each arc \( a \in A_{LR} \) is part of exactly one train digraph \( D_i \) and one track digraph \( D_j \)
- there is an isomorphism between \( Q_j \) and the set of all \( s_{xy}t_{xy} \)-paths in \( D_j \).
3 Bi-criteria Optimization Approach

Introducing 0/1-variables $x_p$, $p \in P$, and $y_q$, $q \in Q$, OPTRA can be stated as the following integer program.

\((PCP)\)

(i) \quad \text{max} \sum_{p \in P} w_p x_p \\
(ii) \quad \sum_{p \in P} x_p \leq 1, \quad \forall i \in I \\
(iii) \quad \sum_{q \in Q} y_q \leq 1, \quad \forall j \in J \\
(iv) \quad \sum_{a \in A} x_p - \sum_{a \in A} y_q \leq 0, \quad \forall a \in A_{LR} \\
(v) \quad x_p, y_q \geq 0, \quad \forall p \in P, q \in Q \\
(vi) \quad x_p, y_q \in \{0, 1\}, \quad \forall p \in P, q \in Q.

The objective PCP (i) maximizes the weight of the track allocation. Constraints (ii) state that a train can run on one route only, constraints (iii) allow at most one configuration for each track. Inequalities (iv) couples train routes and track configurations to guarantee a conflict-free allocation, (v) and (vi) are the non-negativity and integrality constraints. Note that the upper bounds $x_p \leq 1$, $p \in P$, and $y_q \leq 1$, $q \in Q$, are redundant.

An arc based version can be formulated as well. Variables $x_a$, $a \in A_i$, $i \in I$ control the use of trip $a$ in $D_i$ and $y_a$, $a \in A_j$, $j \in J$ in $D_j$, respectively.

\((ACP)\)

(i) \quad \text{max} \sum_{a \in A} w_a x_a \\
(ii) \quad \sum_{a \in \delta^A_{out}(v)} x_a - \sum_{a \in \delta^A_{in}(v)} x_a = 0, \quad \forall i \in I, v \in V_i \setminus \{s_i, t_i\} \\
(iii) \quad \sum_{a \in \delta^A_{in}(s_i)} x_a \leq 1, \quad \forall i \in I \\
(iv) \quad \sum_{a \in \delta^A_{out}(s_j)} y_a - \sum_{a \in \delta^A_{in}(s_j)} y_a = 0, \quad \forall j \in J, v \in V_j \setminus \{s_j, t_j\} \\
(v) \quad \sum_{a \in \delta^A_{in}(s_j)} x_a \leq 1, \quad \forall j \in J \\
(vi) \quad x_a - y_a \leq 0, \quad \forall a \in A_{LR} \\
(vii) \quad x_a, y_a \geq 0, \quad \forall a \in A \\
(viii) \quad x_a, y_a \in \{0, 1\}, \quad \forall a \in A.

As before, the objective, denoted in (ACP) (i), is to maximize the weight of the track allocation. Equalities (ii) and (iii) are well-known flow conservation constraints for all trains $i \in I$, (iv) and (v) for all tracks $j \in J$, 

respectively. Inequalities (vi) link arcs used by train routes and track configurations to ensure a conflict-free allocation, (vii) and (viii) are the non-negativity and the integrality constraints.

![Configuration routing digraph](image)

Figure 1: Configuration routing digraph and from none-robust to robust configuration.

Let us explain the incorporation of robustness on a plain example. By $r_q$ we denote a robustness value for each configuration $q \in Q$. We assume that a high robustness value $r_q$ means configuration $q$ is robust and a smaller the contrary. The only thing, we expect for $r$ is that $r_q = \sum_{a \in q} r_a$.

Figure 1 illustrates the idea on an a single track. Imaging the track digraph on the left is induced by three train requests. Straight maximizing the number of scheduled trains in that easy setting will always lead to a schedule with profit value 3, but as you can see this can result in a lot of varying schedules, in fact all $LR$-paths of length 5, i.e. the three showed in Figure 1. Assume a buffer of length $b \in \mathbb{Z}$ is large enough to ensure robustness for succeeding trains, then the following robustness function $r : \mathbb{R}^{|A|} \rightarrow \mathbb{R}$ with

$$r((u, v)) := \begin{cases} \sqrt{b} & (u, v) \in A_{RL} \text{ and } t(v) - t(u) > b \\ \sqrt{t(v) - t(u)} & (u, v) \in A_{RL} \text{ and } t(v) - t(u) \leq b \\ 0 & \text{otherwise} \end{cases}$$

will measure the available buffers in an appropriate way. The robustness measure $r$ benefits buffer times near to $b$ and somehow balances the partition
of the available buffer times by its concaveness. Assume $b = 2$ in our example in Figure 1. Then the first configuration $q_1$ has value $r_{q_1} = 0$, for the second configuration $r_{q_2}$ is $\sqrt{2}$ and the third one has $r_{q_3} = 2$. (For the sake of completeness we set $r_q$ to a sufficient big $M$ for an empty configuration $q$.)

Figure 2: Function $r(a)$ of a single buffer arc $a \in A_{RL}$.

Figure 3: Robustness function of two buffer arcs.

We can easily extend ACP and PCP to bi-criteria models by taking this second, obviously contradictory, objective into account:

(PCP) (i’) max $\sum_{q \in Q} r_q y_q$ or analogously (ACP) (i’) max $\sum_{a \in A} r_a y_a$.

To find all efficient solutions, we propose a straight-forward combined weighted sum and $\epsilon$-constraint method, as described in Ehrgott [2005]. Considering model PCP this leads to the following objective function with a scalar $\alpha \in [0, 1]$:

$$\max \alpha (\sum_{p \in P} w_p x_p) + (1 - \alpha) (\sum_{q \in Q} r_q y_q)$$

and to an additional constraint on one of the objectives, i.e.

$$\sum_{p \in P} w_p x_p \geq \epsilon \quad \text{or} \quad \sum_{q \in Q} r_q y_q \geq \epsilon.$$

Common way in practice would be to choose an allowed deviation from a profit maximizing schedule. Let $v_{opt}$ be the optimal value derived by solving ACP or PCP. By $\nu$ we denote a given percentage of decrease in the profit function. Choosing $\epsilon = (1 - \nu)v_{opt}$ and adding an $\epsilon$-constraint for the first objective, we receive:
[((α, ε) − PCP)]

\[
\begin{align*}
(i) & \quad \max \ & \sum_{p \in P} w_p x_p + (1 - \alpha) \sum_{q \in Q} r_q y_q \\
(ii) & \quad \sum_{p \in P} x_p & \leq 1, \quad \forall i \in I \\
(iii) & \quad \sum_{q \in Q} y_q & \leq 1, \quad \forall j \in J \\
(iv) & \quad \sum_{a \in P} x_p - \sum_{a \in Q} y_q & \leq 0, \quad \forall a \in A_{LR} \\\n(v) & \quad \sum_{p \in P} w_p x_p & \geq \epsilon, \\
(vi) & \quad x_p, y_q \geq 0, \quad \forall p \in P, q \in Q \\
(vii) & \quad x_p, y_q \in \mathbb{Z}, \quad \forall p \in P, q \in Q.
\end{align*}
\]

3.1 Details on Column Generation

Let us remark that the LP-relaxation PLP of PCP, i.e., PLP = PCP (i)–(v) can be solved efficiently by column generation. Due to the added ε-constraint the structure changed, but only slightly, as we will show. Fortunately, it will turn out that the pricing problems remain solvable in polynomial time (by computing longest paths in acyclic digraphs $D_i$ and $D_j$). To see this, consider the dual DLP of the LP-relaxation of ([(α, ε) − PCP]), i.e. neglecting constraints ([(α, ε) − PCP]) (vii):

(DLP)

\[
\begin{align*}
(i) & \quad \min \ & \sum_{j \in J} \pi_j + \sum_{i \in I} \gamma_i - \epsilon \rho \\
(ii) & \quad \gamma_i + \sum_{a \in P} \lambda_a - w_p \rho & \geq \alpha w_p, \quad \forall p \in P_i, i \in I \\
(iii) & \quad \pi_j - \sum_{a \in Q} \lambda_a & \geq (1 - \alpha) r_q, \quad \forall q \in Q_j, j \in J \\
(iv) & \quad \gamma_i, \pi_j, \lambda_a, \rho & \geq 0 \quad \forall i \in I, j \in J, a \in A_{LR}.
\end{align*}
\]

Here, $\gamma_i, i \in I, \pi_j, j \in J, \lambda_a, a \in A_{LR}$ and $\rho$ are the dual variables associated with constraints ([(α, ε) − PCP]) (ii), (iii),(iv) and (v), respectively. The pricing problem for a route $p \in P_i$ for train $i \in I$ is then:

\[\exists \ p \in P_i : \gamma_i + \sum_{a \in P} \lambda_a - w_p \rho < \alpha w_p \iff \sum_{a \in P} ((\alpha + \rho) w_a - \lambda_a) > \gamma_i.\]

This is the same as finding a longest $s_t$-path in $D_i$ w.r.t. arc weights $(\alpha + \rho) w_a - \lambda_a$; as $D_i$ is acyclic, this problem can be solved in polynomial time.
The pricing problem for a configuration \( q \in Q_j \) for track \( j \in J \) (w.r.t. the additional \( \epsilon \)-constraint (v)) is:

\[
\exists q \in Q_j : \pi_j - \sum_{a \in q} \lambda_a < (1 - \alpha)r_q \iff \sum_{a \in q} (\lambda_a + (1 - \alpha)r_a) > \pi_j.
\]

Using arc weights \( (\lambda_a + (1 - \alpha)r_a) \), \( a \in A_{LR} \), and 0 otherwise, pricing configurations \( Q_j \) is equivalent to finding a longest \( s_j t_j \)-path in \( D_j \). As \( D_j \) is acyclic, this is polynomial. By the polynomial equivalence of separation and optimization, see Grötschel et al. [1988], here applied to the (DLP), we obtain:

**Theorem 3.1.** The LP-relaxation of \(((\alpha, \epsilon) - \text{PCP})\) is solvable in polynomial time.

### 4 Preliminary Computational Results

We consider the Hanover-Kassel-Fulda area of the German long-distance railway network. All our instances are based on the macroscopic infrastructure network provided by our project partners from departments for railway track and operations. All data was produced by suited aggregation to minutes based on detailed mesoscopic simulation results (with a precision of seconds).

The network consists of 37 stations, 120 tracks and 6 different train types (ICE, IC, RE, RB, S, ICG). Because of various possible turnover and driving times for each train type, this produces an infrastructure digraph with 146 nodes, 1480 arcs. For the construction of correct track digraphs, we stick to 4320 realistic headway times.

We tested our model on different scenarios, presenting here the results for one with 146 trains. Based on the 2002 timetable of Deutsche Bahn AG, we considered all trains inside that area in a time interval of about 120 minutes at a normal weekday from 9:00 to 11:00; leading to a representable mix of long distance trains (IC, ICE), synchronized regional and suburban passenger trains (S, RE, RB), and freight trains (ICG). Flexibility to reroute trains is set to departure and arrival time windows of 20 minutes length. Maximize the total number of trains in the schedule is our first objective\(^1\); the second goal is to maximize our defined robustness measure, choosing \( b = 20 \) minutes.

We are only presenting preliminary results for the LP case of ACP by using the barrier method of CPLEX 11.0, see CPLEX [2007]. All computations were made single threaded on a Dell Precision 650 PC with 2GB of

\(^1\)Furthermore, we slightly penalize deviations from certain desired departure and arrival times.
Figure 4: Total profit objective (blue, left axis) and total robustness objective (green, right axis) in dependence on $\alpha$.

main memory and a dual Intel Xeon 3.8 GHz CPU running SUSE Linux 10.1. In Figure 4 the trade-off between robustness and profit function depending on $\alpha$ is exemplary plotted. Whether the results presented in this paper are useful in practice will turn out in the future after some extensive computational studies.

References


