An Optimization Model for Integrated Timetable Based Design of Railway Infrastructure

Andreas Schöbel[1], Günther Raidl[2], Igor Grujicic[2],
Gerhard Besau[3], Gottfried Schuster[3]

[1] Institute of Transportation, Research Centre for Railway Engineering,
Vienna University of Technology,
Karlsplatz 13/230-2, 1040 Vienna, Austria
e-mail: andreas.schoebel@tuwien.ac.at

[2] Institute of Computer Graphics and Algorithms, Algorithms and Data Structures Group,
Vienna University of Technology,
Favoritenstraße 9/1861, 1040 Vienna, Austria
e-mail: {raidl|grujicic}@ads.tuwien.ac.at

[3] ÖBB Infrastruktur AG,
Praterstern 3, 1020 Vienna, Austria
e-mail: {gerhard.besau|gottfried.schuster}@oebb.at

Abstract
The design of new railway infrastructure is a complex planning process today in most European countries due to several requirements. From an operational point of view new infrastructure basically has to fulfil the requirements defined by the later customers who are the railway undertakings. Hereby passenger traffic is often organised in a periodic timetable with well defined arrival and departure times in the hubs. So far there is no automated tool available to help in determining a minimum cost infrastructure fulfilling all the requirements defined by a timetable and the operation of the railway system. Instead, this task is typically carried out manually based on graphical design, human experience, and also intuition. This paper presents a first formalization of this task as a combinatorial optimization problem trying to capture the most essential aspects. For solving it promising algorithmic concepts based on mathematical programming techniques and metaheuristics are sketched.

Keywords
Railway Infrastructure Design, Integrated Timetables, Combinatorial Optimization

1 Introduction

The design of new railway infrastructure is nowadays strongly guided by pre-specified integrated timetables that have been derived from expected traffic to be served [11]. Integrated timetables synchronize the traffic in major nodes (hubs, e.g., main railway stations in major cities) at regular time intervals, ensure connectivity between different lines with minimum waiting times, and allow passengers to easily remember the regular departure and arrival times. In many European countries, integrated timetables have been successfully introduced in the last years and could prove their substantial advantages.

Implementing the concept of integrated timetables, however, imposes major challenges and constraints, see e.g. [8]. In fact, the almost simultaneous arrival of the most relevant trains at a station and the strongly regulated travel times between stations, which must be multiples of a basic cycle interval, frequently demand extensions of existing railway infrastructure. Substantial financial investments are typically necessary for building further tracks, platforms, and other elements in order to be able to realize an integrated timetable and benefit from its long-term efficiency and higher flexibility.

Considering an integrated timetable as a central dogma when building new or extending existing railway infrastructure has a major impact on the design process. Unfortunately till today decisions like where and how to strengthen an existing single route frequently are done from a predominantly local perspective considering the specific route's properties and demands almost only, and timetables are adapted thereafter. Nowadays with an integrated timetable, dependencies between routes of different trains are much stronger, and impacts of certain design decisions have a more global influence. A more systematic optimization approach is thus required in the design in order to achieve a cost-effective solution that guarantees the constraints imposed by the integrated timetable.

Today's state of the art in developing the infrastructure layout is given by graphical procedures upon the required arrival and departure times [19] and validation by microscopic simulation of railway operation [9]. In this paper we present a concrete graph-theoretic/combinatorial approach for modeling the basic problem. It considers existing railway infrastructure as well as various extension possibilities in a fine-grained track-segment based way, speed limits on segments in dependence of trains and chosen routes (entry situation in stations), and various kinds of costs for installing new elements of infrastructure (tracks and switches) in order to be able to realize the connections as specified by the integrated timetable. The model is flexible in the sense that it can be relatively easily adapted to further, more special or alternative requirements.

We start by reviewing the so far used graphical estimation process in Section 2. Section 3 then presents the new formal optimization model to determine a minimum-cost demand for the infrastructure. In Section 4, we discuss the problem from a theoretical point-of-view, also showing that it belongs to the class of computationally difficult NP-hard problems. This implies that we cannot expect to find an efficient (i.e., polynomial time) exact algorithm for solving any instance to optimality. Nevertheless, exact methods based on mixed integer linear programming (MIP) appear highly promising for addressing smaller instances in practice, and metaheuristic approaches as well as MIP/metaheuristic hybrids seem well suited to approximately approach larger scenarios. Section 5 briefly sketches first concepts. Concrete implementations and experimental evaluations using artificial test instances derived from real-world scenarios are work in progress. Finally, Section 6 concludes this article.

2 Graphical Estimation

Having once defined the hubs for an integrated timetable in a railway network, the railway lines between the nodes have to be designed in accordance. Starting with a single track line, the demand for crossing opportunities can be simply identified in the graphical timetable. Figure 1 shows the situation when two trains with predefined arrival and departure times in the neighboring nodes want to cross each other on a single track line. Taking into consid-

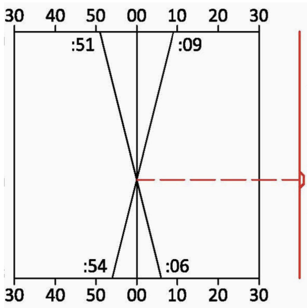

Figure 1: Graphical timetable and required infrastructure layout for two trains running on a line in opposite directions.

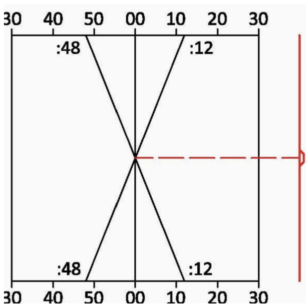

Figure 2: Graphical timetable for two additional trains and their required infrastructure.

Figure 3: Infrastructure layout for the combined operation of the four trains.

eration the maximum length of the involved trains, the minimum length of the second track can be also simply calculated. Of course, the second track should be designed at track speed to face no restrictions of speed and thereby additional running time.

More complicated becomes the situation when additionally two other trains are supposed to run on this single track section in the same time slot, cf. Figure 2. This might be the case when two fast trains and two local trains have to be operated to have constant intervals. For the local services also another meeting point could be defined by adjusting the arrival and/or departure times in the neighboring nodes. If both services have to be operated in the same time slot, the combination of all trains specifies the infrastructure layout.

The combination of both services requires at least an infrastructure which is shown in Figure 3. As a generic rule the minimum number of required tracks is defined by the number of trains in an infrastructure section. The graphical solution has some short comings as e.g. train dynamics are not considered exactly but at least it delivers a rough estimation of the minimum infrastructure layout required in simple scenarios. In more complex scenarios, this approach quickly becomes inefficient.

3 Combinatorial Optimization Model

We now model the *Integrated Timetable Based Design of Railway Infrastructure* (TTBDRI) as a combinatorial optimization problem, trying to consider the most relevant real-world aspects on the one hand while performing simplifications and discretizations where it appears reasonable and meaningful on the other hand.

The TTBDRI problem has given the following input data.

- The undirected graph $G = (V, E)$ represents the existing railway infrastructure plus all possible installable extensions on a detailed level. Vertices V correspond to atomic track segments (or more precisely their respective center points), edges E represent these segments' connections, possibly with switches inbetween. Mutually exclusive extensions (alternatives) are modeled in the graph by independent connected components from which only one may finally be used. Multiple parallel tracks are represented by multiple paths. Each edge $e \in E$ has associated a length $l_e \geq 0$ corresponding to the real distance of the centers of the track segments referred to by the connected vertices. Figure 4 shows an example.

- Let the subgraph $G^0 = (V^0, E^0)$, with $V^0 \subseteq V$ and $E^0 \subset E$, correspond to the already existing infrastructure, and the graph $G' = (V, E')$ with $E' = E \setminus E^0$ represent the additionally possible infrastructure by which the existing infrastructure may be extended. Nodes $v \in V$ have associated installation costs $c_v \geq 0$ with $c_v = 0$ for $v \in V^0$.

- Let $R \subseteq V$ be the set of vertices where signaling stations are (to be) located. Paths starting and ending at such vertices and otherwise containing only vertices from $V \setminus R$ are called *compound routes*. Connected components of the network that are delimited by nodes in R may at any time be accessed by a single train only. A train may enter a compound route only after it received a reservation, and this reservation stays active until the train leaves the compound route.

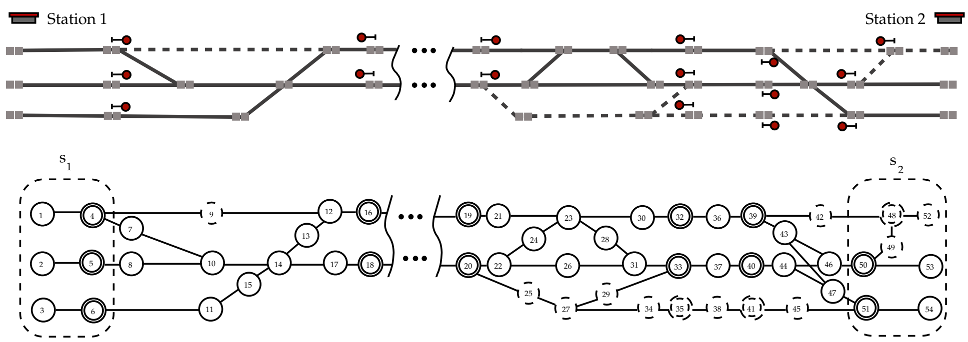

Figure 4: A railway system connecting two stations and the corresponding graph G on which the TTBDRI problem is defined. Dashed nodes and arcs indicate potential new infrastructure.

- Set S represents the major railway stations (hubs) considered in the integrated timetable. Each railway station $s \in S$ has associated a vertex set $V(s) \subset V$ corresponding to the tracks at the platforms for boarding/disembarking trains in station s.

- Let $G^{\mathrm{D}} = (V, A)$ be the directed version of graph G, where we have for each edge $(u, v) \in E$ two corresponding oppositely directed arcs $(u, v), (v, u) \in A$ with lengths $l_{(u,v)}$.

- An integrated timetable specifies a set of connections C to be realized, where a connection $t \in C$ is a tuple $(s_t^{\text{start}}, s_t^{\text{end}}, T_t^{\text{start}}, T_t^{\text{end}}, G_t^{\mathrm{D}}, train_t, l_t)$ with $s_t^{\text{start}}, s_t^{\text{end}} \in S$ being start and destination stations and T_t^{start} and T_t^{end} the times when the train may leave station s_t^{start} and has to arrive at station s_t^{end} latest, respectively. The connection has to be realized by a path in the limited subgraph $G_t^{\mathrm{D}} = (V_t, A_t)$ with $V_t \subseteq V$ and $A_t \subseteq A$. It can safely be assumed that G_t^{D} is acyclic. Finally, $train_t$ indicates the used train's ID. Typically, a train is used for a series of connections. Let $l(train_t)$ refer to the train's length.

- Values $maxspeed_{t,a} \geq 0$ indicate the maximum allowed average speed by which the train realizing connection $t \in C$ may go from the source to the target vertex of arc $a \in A_t$.

A solution consists of:

- A subgraph $G'' = (V'', E'')$ with $V'' \subset V$ and $E'' \subseteq E'$ indicating the infrastructure to be installed.
  Let $G^{\mathrm{e}} = (V^{\mathrm{e}}, E^{\mathrm{e}})$ represent the complete augmented infrastructure, i.e., $V^{\mathrm{e}} = V^0 \cup V''$ and $E^{\mathrm{e}} = E^0 \cup E''$.

- For each connection $t \in C$ a directed path $P_t \subseteq A_t$ starting at a vertex from $V(s_t^{\text{start}})$ and ending at a vertex from $V(s_t^{\text{end}})$.
  Considering the signaling stations R as separators, P_t can be partitioned into a list of successive compound routes $L_t = (P_{t,1}, \ldots, P_{t,\lambda_t})$.
  The length of a compound route $P_{t,i}$, $i = 1, \ldots, \lambda_t$, is $l(P_{t,i}) = \sum_{a \in P_{t,i}} l_a$.

- For each arc $a = (u, v) \in P_t$, $t \in C$, an actual (average) speed $0 < speed_{t,a} < maxspeed_{t,a}$. To keep the model a combinatorial optimization problem, we may restrict ourself to an appropriately chosen discrete set of allowed average speeds σ, i.e., $speed_{t,a} \in \sigma \wedge maxspeed_{t,a}$.
  Consequently, the train takes time $T_{t,a} = l_a / speed_{t,a}$ for going from the center of track segment u to the center of segment v.

- For each route $P_{t,i}$, $i = 1, \ldots, \lambda_t$, $t \in C$, a reservation time slot $(T_{t,i}^{\text{enter}}, T_{t,i}^{\text{exit}})$ in which the train will safely be able to pass this route.
  Safety margins are added: When the train takes total time

$$T_{t,i} = \sum_{a \in P_{t,i}} T_{t,a} \tag{1}$$

  for passing the route, the reservation time slot has duration

$$T_{t,i}^{\text{exit}} - T_{t,i}^{\text{enter}} = (1 + 2\alpha) \cdot T_{t,i}, \tag{2}$$

i.e., margins $\alpha \cdot T_{t,i}$ are added before the train is expected to enter and after the train is expected to leave, respectively.

To be feasible, a solution must satisfy:

- For each connection $t \in C$: $\forall (u,v) \in P_t \rightarrow u, v \in V^{\mathrm{e}} \wedge (u,v) \in E^{\mathrm{e}}$, i.e., the infrastructure used in the chosen paths must exist or be installed.
- All constraints for realizing possible extensions (e.g., mutual exclusivity of some alternatives) must be adhered.
- The time slots of consecutive routes of a connection overlap exactly by the corresponding safety margins.
- For each connection $t \in C$, the earliest start and latest arrival times T_t^{start} and T_t^{end} are adhered, respectively.
- At each time, each edge $e \in E^{\mathrm{e}}$ may only be used in at most one reserved route.
- If the same train is used for two successive connections, its arrival vertex at the first connection's target station must be the same as the vertex where it leaves from in the second connection.

The objective is to find a feasible solution with minimum total costs

$$\sum_{v \in V''} c_v + \sum_{t \in C} \sum_{a \in P_t} c(t, a, \mathit{speed}_{t,a}), \tag{3}$$

where the first term refers to the costs of the infrastructure to be installed and function $c(t, a, \mathit{speed}_{t,a})$ represents costs for the train passing arc a with speed $\mathit{speed}_{t,a}$. Assuming that higher speeds and in particular unnecessary speed changes are less desirable, this function may in the simplest case be

$$c(t, a, \mathit{speed}_{t,a}) = \varepsilon \cdot \mathit{speed}_{t,a}^{\gamma}, \tag{4}$$

where ε is a small constant and $\gamma > 1$.

4 Computational Complexity of TTBDRI and Related Work

The stated problem is in general difficult to solve from a theoretical as well as practical point of view. This is documented by the fact that TTBDRI obviously generalizes several well-known NP-hard combinatorial optimization problems, such as the classical multi-commodity flow problem with resource constraints.

With respect to railway optimization, TTBDRI contains a variant of the diverse train scheduling/timetabling problems that have already received considerable attention in the literature, see e.g. [20, 5, 12, 4]. In particular, Caprara et al. [5] give an explicit proof showing that any polynomial-time approximation algorithm with worst-case performance guarantee is hopeless for their problem variant unless P=NP. The reduction is from the notoriously NP-hard maximum independent set problem and can be adapted to our case.

The main goal in our work, however, lies in the infrastructure design aspect, and in this respect we are not aware of any other directly comparable work, besides the already mentioned graphical solution approach [19].

More generally, TTBDRI has similarities to a variety of network design problems as they appear in telecommunication applications, see e.g. [3].

5 Solution Approaches for TTBDRI

We approach the problem from two sides: On the one hand we are working on a mixed integer programming (MIP) based method to solve smaller instances to proven optimality. As it is unrealistic to assume that we will be able to solve large real-world scenarios in this way in reasonable time, we also consider (meta-)heuristics and hybrid optimization techniques to obtain approximate solutions on the other hand.

Our MIP approach utilizes the concept of a space-time network as described in [4]. It is obtained by extending graph G with the dimension of time: The nodes of the network correspond to pairs of space and time, they are multiple copies of V indexed by a discrete selection of times, e.g. every 10 seconds over the whole time interval considered in the timetable. The network is acyclic and all arcs are directed from a node indexed with a lower time to a node indexed with a higher time. Arcs are established for all possibilities of a train going from one node in V no an adjacent node, considering all possible times for leaving the source node and arriving at the destination when going with each of the possible speeds σ. Graphs G_t^D considered for the individual connections $t \in C$ induce corresponding subnetworks.

Using this space-time network, it is straight-forward to express TTBDRI by a multi-commodity flow formulation with additional constraints. For each train we define an individual commodity and enforce subsets of nodes that must be passed in accordance to the connections specified in the timetable. Similarly to [5] we express the aspect that compound routes may only be accessed exclusively by one train with conflict sets.

Unfortunately, the space-time network will become extremely large when a meaningful set of discrete times as well as a larger railway infrastructure is considered. Thus, the multi-commodity flow model can in practice not directly be applied in the sketched way. Instead, we are considering an adaptive approach: We start with a relatively small network considering only few discrete times and an abstract, macroscopic representation of the infrastructure. The flow model is then iteratively solved on this network, which is incrementally extended to get better and more precise solutions. We refer in this respect to our adaptive layers framework [18, 17] which works well in the context of delay-constrained minimum tree problems.

Furthermore, a Dantzig-Wolfe Decomposition of the flow formulation yields an alternative MIP model that is based on path variables, i.e., a model that considers variables for all possible train paths P_t for realizing the connections. While this model cannot be solved directly due to the exponential number of variables, column generation and branch-and-price [6] provide promising approaches for practically efficient algorithms. Their key-idea is to start with a small set of paths, solving the linear programming relaxation and augmenting the model iteratively with additionally considered path variables which are identified by solving a pricing subproblem. Borndörfer et al. [4] sketch a similar approach. Fischer and Helmberg [7] further described a related dynamic graph generation to solve pricing problems for very large graphs. As column generation based methods are frequently prone to problems with degeneracy, the consideration of stabilization techniques plays a major role. We expect that we will be able to adapt our stabilization approach described in [10] for constrained tree network design problems. It it should speed up column generation for TTBDRI considerably.

With respect to heuristic methods, we work on a variable neighborhood search approach [14] that utilizes a set of specifically designed neighborhood structures, among them also very large scale neighborhood search techniques [1]. One basic strategy, for example, is to remove a small number of connections and reinsert them in an optimal (or almost optimal) way. This smaller subproblem can be solved by either the multi-commodity flow approach described above or, if it comes to single connections, by dynamic programming, a label-correcting algorithm, or A^* search. See [15] for a general overview on such hybrid metaheuristics. As partly already finding any feasible solution is a challenging task due to the many constraints, we further consider constraint programming [16] to be very useful in this context. Last but not least, we intend to combine the column generation approach with the variable neighborhood search in a way similar to the SearchCol framework from [2]. This hybrid method is expected to combine the benefits of the column generation approach with the better scalability of the metaheuristics.

6 Conclusions

The design of new railway infrastructure in order to meet a given demand specified by a periodic timetable has so far been mainly done manually or by relatively simple graphical tools. This paper presents a first formal model to express this problem as a combinatorial optimization problem. It turned out that this problem is (loosely) related to several other well known combinatorial optimization problems and obviously is NP-hard. Subsequently we sketched our work in progress: Based on a space-time network, a multi-commodity flow MIP formulation can be derived. Following a Dantzig-Wolfe Decomposition a path-formulation can be obtained, which can be approached by column generation and branch-and-price. Besides these MIP methods, we work on a variable neighborhood search that utilizes very large scale neighborhood structures in order to approximately solve larger problem instances in practice in reasonable time. Particularly promising seem to by hybrid approaches that combine MIP and metaheuristic techniques. A special challenge are appropriate techniques to keep the space-time network as small as possible or to avoid its explicit creation.

We finally note that a sparser, more economical infrastructure might be more at risk of running into deadlocks when trains are scheduled too naively. Therefore, deadlock prevention becomes more important, cf. [13].

References

[1] R. K. Ahuja, Ö. Ergun, J. B. Orlin, and A. P. Punnen. A survey of very large-scale neighborhood search techniques. *Discrete Applied Mathematics*, 123(1-3):75–102, 2002.

[2] F. Alvelos, A. de Sousa, and D. Santos. Combining column generation and metaheuristics. In *Hybrid Metaheuristics*, volume 434 of *Studies in Computational Intelligence*, pages 285–334. Springer, 2013.

[3] G. Anandalingam and S. Raghavan. *Telecommunications Network Design and Management*. Kluwer Academic Publishers, 2002.

[4] R. Borndörfer, T. Schlechte, and E. Swarat. Railway track allocation – simulation, aggregation, and optimization. In *Proceedings of the 1st International Workshop on High-Speed and Intercity Railways*, volume 148, pages 53–69. Springer, 2012.

[5] A. Caprara, M. Fischetti, and P. Toth. Modeling and solving the train timetabling problem. *Operations Research*, 50(5):851–861, 2002.

[6] G. Desaulniers, J. Desrosiers, and M. M. Solomon, editors. *Column Generation*. Springer, 2005.

[7] F. Fischer and C. Helmberg. Dynamic graph generation and rolling horizon techniques in large scale train timetabling. In *Proceedings of the 10th Workshop on ALgorithmic Approaches for Transportation Modelling, Optimization, and Systems*, volume 14 of *Open Access Series in Informatics*, pages 45–60, Dagstuhl, Germany, 2010.

[8] C. Gabrio, L. Marco, S. Kaspar, W. Stefan, and F. Martin. The periodic service intention as a conceptual framework for generating timetables with partial periodicity. *Transportation Planning and Technology*, 34(4):323–339, 2011.

[9] D. Hürlimann. *Objektorientierte Modellierung von Infrastrukturelementen und Betriebsvorgängen im Eisenbahnwesen*. PhD thesis, ETH Zürich, 2001.

[10] M. Leitner, M. Ruthmair, and G. R. Raidl. On stabilizing branch-and-price for constrained tree problems. *Networks*, to appear 2013.

[11] M. Lichtenegger. Der Integrierte Taktfahrplan: Abbildung und Konstruktion mit Hilfe der Graphentheorie, Minimierung der Realisierungskosten. *Eisenbahntechnische Rundschau*, 40:171–175, 1991.

[12] R. Lusby, J. Larsen, M. Ehrgott, and D. Ryan. Railway track allocation: Models and methods. *OR Spectrum*, 33:843–883, 2011.

[13] R. Mittermayr, J. Blieberger, and A. Schöbel. Kronecker algebra-based deadlock analysis for railway systems. *Promet – Traffic & Transportation*, 24(5):359–369, 2012.

[14] N. Mladenović and P. Hansen. Variable neighborhood search. *Computers and Operations Research*, 24(11):1097–1100, 1997.

[15] G. R. Raidl, J. Puchinger, and C. Blum. Metaheuristic hybrids. In M. Gendreau and J. Y. Potvin, editors, *Handbook of Metaheuristics*, volume 146 of *International Series in Operations Research & Management Science*, pages 469–496. Springer, 2nd edition, 2010.

[16] F. Rossi, P. V. Beek, and T. Walsh, editors. *Handbook of Constraint Programming*. Elsevier, 2006.

[17] M. Ruthmair. *On Solving Constrained Tree Problems and an Adaptive Layers Framework*. PhD thesis, Vienna University of Technology, Institute of Computer Graphics and Algorithms, Vienna, Austria, May 2012.

[18] M. Ruthmair and G. R. Raidl. A Layered Graph Model and an Adaptive Layers Framework to Solve Delay-Constrained Minimum Tree Problems. In O. Günlük and G. Woeginger, editors, *Fifteenth Conference on Integer Programming and Combinatorial Optimization (IPCO XV)*, volume 6655 of *LNCS*, pages 376–388. Springer, Heidelberg, 2011.

[19] A. Schöbel and G. Besau. Timetable Based Design of Railway Infrastructure. In *EURO-ZEL 20th International Symposium*, June 2012.

[20] B. Szpigel. Optimal train scheduling on a single track railway. In *Operations Research 72*, pages 343–351, Amsterdam, The Netherlands, 1972.