Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

RALF BORNDÖRFER[1] MARTIN GRÖTSCHEL[1] SASCHA LUKAC[2] KAY MITUSCH[2] THOMAS SCHLECHTE[1] SÖREN SCHULTZ[3] ANDREAS TANNER[2]

An Auctioning Approach to Railway Slot Allocation

[1]Konrad-Zuse-Zentrum für Informationstechnik Berlin, Takustr. 7, 14195 Berlin, Germany; Email:{borndoerfer, groetschel, schlechte}@zib.de

[2]Workgroup for Infrastructure Policy (WIP), Strasse des 17. Juni 135, 10623 Berlin, Germany; Email:{sl, km, ant}@wip.tu-berlin.de

[3]Department of Railway Track and Operations (SFWBB), Salzufer 17 - 19, 10587 Berlin, Germany; Email:{sschultz@railways.tu-berlin.de}

ZIB-Report 05-45 (October 2005)

An Auctioning Approach to Railway Slot Allocation

Ralf Borndörfer Martin Grötschel Sascha Lukac
Kay Mitusch Thomas Schlechte Sören Schultz
Andreas Tanner

Abstract

We present an approach to implement an auction of railway slots. Railway network, train driving characteristics, and safety requirements are described by a simplified, but still complex macroscopic model. In this environment, slots are modelled as combinations of scheduled track segments. The auction design builds on the iterative combinatorial auction. However, combinatorial bids are restricted to some types of slot bundles that realize positive synergies between slots. We present a bidding language that allows bidding for these slot bundles. An integer programming approach is proposed to solve the winner determination problem of our auction. Computational results for auction simulations in the Hannover-Fulda-Kassel area of the German railway network give evidence that auction approaches can induce a more efficient use of railway capacity.

1 Introduction

During the last decades many countries have experimented with open access policies to their railway networks. Particularly the EU advocates non-discriminatory open access as a part of its general policy towards market opening and on-track competition, and as an instrument to foster European railway integration, namely, by allowing national train operating companies to access the networks of neighbor states.

The usefulness of open access for railways is widely discussed about. The main argument is whether on-track competition can be a viable alternative to fully integrated or regional franchise systems or not. One of the main objections is that on-track competition wouldn't really work on a full scale, due to the problems associated with coordination and network timetabling. Indeed, there does at present not exist a satisfactory allocation mechanism for network slots under conditions of non-discriminatory open access and on-track competition.

This paper outlines an approach to design and develop such a mechanism.[1] The basic idea is to perform slot auctions, in which scarce network

[1] The paper is based on research project "Trassenbörse" ("Railway slot exchange"),

capacities are allocated to the most valuable uses, and not by inherited 'rights' or by simple priority rules. This idea is not new. In fact, auctioning procedures have already been established in several European railway markets. For example, in Germany, slots are allocated according to priority rules, however, railway law and regulation also provide for the application of a 'highest price procedure' in case of a conflict between equally ranked slot requests.[2] This 'highest price procedure' is being criticized since it does not take account of interdependencies.The argument is that someone who wins a particular bottleneck slot in a bidding procedure might block other scarce parts of the network and drive out demands that were not involved in the particular bidding procedure. This can lead to inefficient allocations. Isolated bidding for bottlenecks is therefore not appropriate for railway network allocation.

The way to overcome this objection is to consider a so-called *combinatorial auction* that allocates a multitude of interdependent slots simultaneously. Currently, theory and practice of such combinatorial auctions is rapidly developing[3], the main applications being logistics and spectrum auctions in telecommunications.[4] Railway as well as airport slot allocation are also often cited as possible applications for combinatorial auctions, however, as far as we know, no specific auction design has yet been developed that would meet the particular needs and problems of railway systems. There probably doesn't exist one for airport slots either, and even if it did, it is unclear how one could transfer results from this area to a railway setting, because the analogy between railway and airport allocation is limited. In fact, for air planes, only take-off and landing slots have to be planned in advance, while routing in air takes place spontaneously. In railways, in contrast, issues like track capacities, overtaking, and signalling systems are essential ([Pachl 2003, p.228]). Therefore, the efficient use of a railway network hinges not only on departures and arrivals, but also on the combination of routes and speeds taken by individual trains in the network. Another difference to airlines is the larger importance of regular service patterns in timetables. Constraints such as these make railway capacity allocation much more complex than that of airport slots.

In this paper we will describe our basic approach to combinatorial railway

funded by the German Ministry of Education and Research (BMBF), 19 M 4031A. It is joint research of the following institutions: TU Berlin, Workgroup for Infrastructure Policy (WIP) and Department of Railway Track and Operations (SFWBB). Konrad-Zuse-Zentrum Berlin (ZIB) for mathematical optimization. Institut für Verkehrswesen, Eisenbahnbau und -betrieb (IVE), University Hannover. Rail Management Consultants GmbH, Hannover. Consultants Ilgmann & Miethner. IABG Munich. Dr. Udo Ziegenhorn.

[2]In Great Britain some kind of auction mechanism for the allocation of individual train paths was considered during the rail reform process but finally not implemented; see Affuso [2003] and Affuso & Newbery [2004].

[3]See for example Milgrom [Milgrom 2004, Chapter 8].

[4]See Ledyard et al. [2002] or Caplice & Sheffi [2003].

slot auctioning. We (i) model the railway network in a suitably simplified form, (ii) design a multi-round auction including an appropriate bidding language for train operating companies to express their bids, and (iii) solve the combinatorial track allocation problems that arise in each round. At present, the auction itself is quite simple. We will sketch our auction rules (i.e., the determination of winners and actual payments etc.) only briefly, and do not address bidding strategies at all; these topics are for further research.

Our approach allows train operating companies (TOCs) to express bids that entail some flexibility with respect to departure and arrival times, to speed, and to the exact route taken (where deviations from the most preferred schedule lead to reductions of the monetary bid). In each auction round a combinatorial optimization problem, a so-called *optimal track allocation problem* ("OPTRA"), will be solved to pick the combination of slot requests that maximizes network proceeds. In this way, requests that would drive out too many other high-valued requests will not be allocated. More importantly, the structure of allocated slots will reflect the overall scarcity of slots in the network. For example, it might happen that a high-speed train is slowed down by the overall optimization procedure in order to leave more room for slower, but highly valued trains; such a result would normally not be achieved by hierarchical planning. In accordance with EU criteria, we assume that the base prices for slots are fairly low. Thus, the true aim of the auction mechanism is an *optimal use* of the network, not literally the maximization of network proceeds.[5]

Mathematical optimization of an entire network timetable is completely new to the industry. Until now, network timetables are hand-optimized: inherited timetables are improved selectively in a trial and error manner. The main reason for this way of planning seems to be the complexity that arises from the manifold important technical components and details of a railway network with its tracks, switches, signals, and train running dynamics. Indeed, constructing an algorithm to optimize the network use on a *microscopic* basis seems to be hopeless. We will therefore resort to simplifications: a coarser classification of train types replaces individual train running dynamics and the network is described by a simplified 'macroscopic' model. On this basis, our mathematical optimization module, as it stands, works for a small subset of the German network. We are, of course, aware that it remains to be shown that our approach can be scaled to larger scenarios and, perhaps even more important, that a coarse timetable of the type that we propose can be refined to a detailed, 'tractable' plan in a routine way without loosing its characteristics.

The optimization of the *entire* network timetable in one single auction

[5]Between auction dates it will be possible to acquire slots on short notice for a short period of time, but this is not our main interest.

is not a practical aim anyway. One of the reasons for considering smaller scenarios are terms of validity. Generally, the time interval between two auctions should be shorter (say, one year) than the expiration time of the rights to use a slot (say, four years). As a consequence, only part of the network capacity will be on sale in any single auction. Such a partial auctioning will reduce the complexity of the optimization problem, but it will also reduce the optimization potential. Nevertheless, it is hoped that the overall timetable will improve over time if every slot will eventually return to the market. In any case, if an auction mechanism will be put into practice one day, it will be introduced on small, confined allocation problems (like a region or an important corridor) to test its power and applicability.

Irrespective of the organizational structure of the railway system and its marketing, a powerful track allocation method will be of interest in its own right. A fully integrated company would be subdivided into different transportation departments (like regional and long-distance passenger services, freight services) and a rational allocation of network slots to the independent profit centers requires a fair and non-discriminatory allotment procedure. Such a company might very well want to do an internal auction of slots. Moreover, since open access is required by EU directives and competition is politically appreciated, the presence of such an integrated company would certainly call for a fair and non-discriminative allocation mechanism. Of course, a network manager in a vertically separated system would be even more interested in a fair, revenue-maximizing mechanism.

The paper is outlined as follows. After a general literature overview in Section 2, Section 3 describes our railway model. The main topic here is the tradeoff between the level of detail required by a realistic model and the requirement to reduce complexity for the sake of tractability. Section 4 proceeds with our auction design including an economic bidding language that enables train operating companies to express their bids in a satisfactory and flexible way. Our mathematical optimization approach to the track allocation problems arising in each round of the auction is explained in Section 5. Section 6 illustrates its capabilities by some computational results. The final Section 7 outlines the route for further research.

2 Literature Overview

Much of the early research on auctioning railroad slots was inspired by the Swedish Parliament's move to search for market-based mechanisms for track allocation. Brewer & Plott [1996] propose in this context an auction-based mechanism for track allocation provided that slots have the *binary exclusion* property, that is, a set of slots is consistent if any two of its elements are. This condition, however, does not hold for a general network of tracks with alternative connections available between origin-destination pairs. Nilsson

[2002] follows up and introduces the triangular shape of the operator's utility functions that we will also use. He notes that some scheduled trains may have complementary value (e.g., by securing connections), and suggests that bidding for *bundles* of trains may be allowed. To handle negative interdependencies by closely scheduled trains competing for customers, the author suggests a *regret rule* that allows bidders to withdraw their last bid. Also, the author quotes a result of Ellis and Silva that states that in long-term equilibrium, there will be only one company serving each market segment, and concludes that considering negative interdependencies is not strictly necessary. Authors don't give a final statement on the necessity of the regret rule. Isacsson & Nilsson [2003] report an experimental game where participants compete over one rail segment, and compare bidding behaviour under first- and second price auctions and two types of stopping rules.

Parkes & Ungar [2001] present an auction-based track allocation mechanism for the case that single-track, double-track, and yard segments have to be concatenated to form a single line. (Double-track segments allow trains to pass each other, yards allow passes and meets of trains.) Parkes assumes that train operators schedule trains between a given origin destination pair, and uses operator utility functions similar to those of Nilsson. Parkes assumes that every line segment is *managed autonomously* by a dispatcher. All dispatchers run simultaneous but otherwise independent auctions to find the optimal allocation of slots for their track segment. Train operators must bid for every segment their train passes on its way, i.e., they bid for the right to enter and leave a segment at specified times where times can be either fixed or flexible. For flexible bids, the bid amount decreases linearly with the allocated times deviating from the requested times. Composition of XOR-connected combinatorial bids is allowed. The rules of the auction follow Parkes i-bundle (Parkes [1999]) auction design.

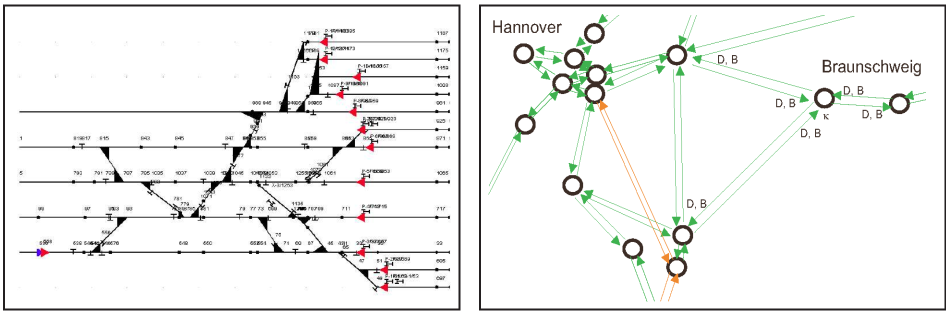
In Parkes's approach, the only form of coordination between different segments is that all auctions end simultaneously. His auction therefore can be seen as a combination of a combinatorial auction with a simultaneous ascending auction: the combination of segments to a complete train line is done via simultaneous ascending auctions for each required segment, while for the timing of segment access, XOR-connected bids are allowed. It is well-known (see e.g. Milgrom [2004] chapter 7) that simultaneous ascending auctions work well for goods that are substitutes, while for complementary values, such as the segments of a line (Parkes uses a linear network, i.e., without cycles), there is in general no stable bidding strategy. Parkes also claims no theoretical result regarding efficieny. He rather reports on experiments, in which all trains travel from end-to-end over the network, i.e., all trains have identical complemetarities, and with players using a myotic best response strategy. In this setting, he reports high efficiencies.

Bassanini et al. [2002] give a game-theoretic model of train operators' competition about track usage. They don't model train conflicts, but rather consider *congestion* which slows down affected trains. The minimal track occupation time between two stations is the sum of the free-running time and a congestion-dependent delay. A schedule then is feasible if every scheduled train in it is feasible (that is, respects the speed limit of the train, a minimal dwelling time at stations, etc.), and if the constraints imposed by congestion are satisfied. The game is played separately for every *line* (which is composed of line segments). The authors suggest that if multiple lines are considered, the mechanism should be applied to all lines subsequently in order of decreasing traffic intensity of this line. They would also allow requests for connections from branch lines. The authors have compared empirical data from the Rome-Milan line with numerical simulations of their model and found a good correspondence. However, the model is too coarse to generate a valid traffic diagram, and therefore not applicable for our purpose.

Brnnlund et al. [1998] propose an integer programming approach to the optimal track allocation problem similar to, but less detailed, than ours. They consider a railway network with a block security system, in which all conflicts between trains arise from simultaneous allocations of one block to more than one train at a time, and fixed departure and arrival times. Based on a path-based integer programming model, they generate a feasible schedule using Langrangean-relaxation techniques and predefined prioritiy rules for the considered train types. For a linear network between Uppsala and Borlänge with 17 stations and 30 trains over a time horizon of 17 hours, they can compute timetables with optimality gaps of at most 3,8%.

Caprara et al. [2002] introduce an integer programming approach to what they call the train timetabling problem; this problem is equivalent to our optimal track allocation problem without combinatorial bidding constraints. They propose an arc-based multi-commodity flow model and attack it with a heuristic based on Lagrangean relaxation. The computational experiments study real-world problems provided by Italian railway companies. The instances deal with a single, one-way track linking two major stations, with a number of intermediate stations in between. The networks considered contain up to 73 stations and 500 trains during a time horizon of one day, which was discretized in steps of one minute. In these instances, almost all trains could be scheduled wtih optimality gaps below 2%. They also tested more congested instances, for which only a part of the trains can be scheduled. In these scenarios, the optimality gap increases up to 10-20%, the solution time increases as well. In a subsequent publication, Caprara et al. [2001] improved their train timetabling model by taking into account several additional constraints that arise in real-world applications. Computational results for the above mentioned, modified scenarios with up to 221 train requests are reported.

Overall, the question of whether auctions can be used for efficient rail path allocation, and how an implementation would need to work, can't be



(a) A station head in microscopic view, with signals (red triangles) and switches (black triangles) — far too many details for a mathematical optimization

(b) Macroscopic view of a subnet, with nodes and arcs (high-speed in red). For node Braunschweig, station capacity κ and the driving-time vectors D and headway matrices B of the adjacent arcs are indicated.

Figure 1: Microscopic and macroscopic view of railway networks

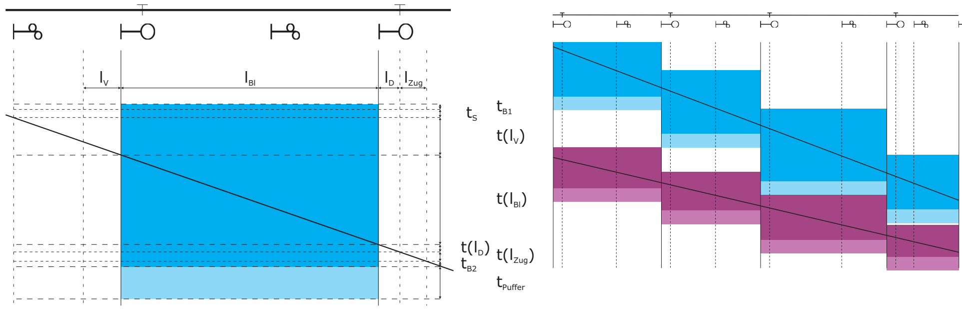
answered from the existing literature. As far as we know, the literature does not propose models and solution concepts for track allocation problems of practically relevant sizes, that would adequately reflect the complexity of a real-world railway network. In fact, most of the literature considers simplified, non-branching lines, or even single segments.

3 Railway Model

Railway scheduling has always been considered a difficult problem. Current schedules are the result of a year-long process of incremental changes that reflect experience and needs to adapt to changing demand and technology. Schedules are, however, not optimized in a mathematical sense, at least not on a national level.

3.1 Blocking Times

Classical railway scheduling is based on a *microscopic model* of the railway network, i.e., a model that includes all tracks, switches, and signals, see Figure 1(a). Such a model is suited for planning on the principle of *exclusive use of block sections*. This means that a block section, roughly spoken the line segment between any two main signals, may be occupied by at most one train at a time. Driving characteristics of individual trains, such as maximum speed, acceleration and braking distance, as well as the characteristics of the respective track segments, such as allowed speed and signalling system, determine the blocking time, i.e., the time that a block section is occupied, and the headway, i.e., the minimum time interval be-



(a) One block (b) A sequence of blocks

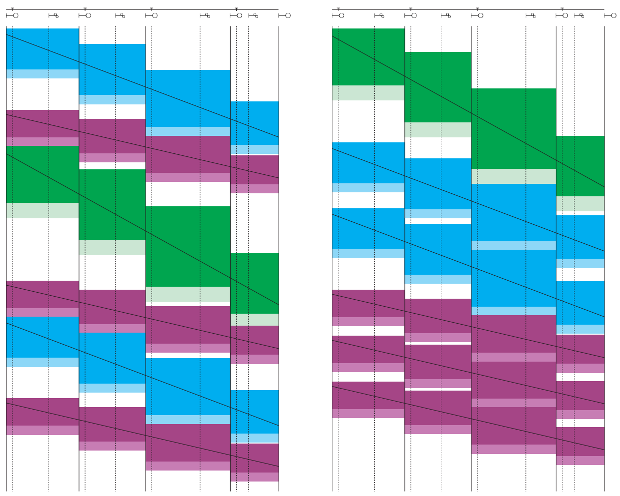
t_{B1}	time to set up a route	t_{Puffer}	buffer time
t_S	time to perceive the signal aspect	l_V	warning distance
$t(l_v)$	approach time - travel time between distant and main signal	l_D	slip distance
$t(l_{Bl})$	driving time within the block section	l_{Bl}	length of a block section
$t(l_D)$	drive time for slip distance	l_{Zug}	train length
$t(l_{\text{Zug}})$	route clearing time		main signal
t_{B2}	route release time		distant signal

Figure 2: Blocking time stairway

tween two trains; see Pachl [2002], Brnger [1996], Hauptmann [2000] and Hrlimann [2001] for more information on railway scheduling.

Block section exclusivity can be visualized in a *blocking time diagram*, see Figure 2(a) for an example. The horizontal axis of such a diagram represents a given track with its signals; the vertical axis measures time (downwards). The black line represents a train traversing through a block section. The blue area marks the blocking time interval where access to the block is locked for any other train; its computation takes into account the length of the block, the distance from the distant signal to the start of the block, the length of the slip distance, the train length, and the train driving time. Considering the blocking times for a sequence of consecutive block sections gives rise to the characteristic *blocking time stairway*, see Figure 2(b). Here, shaded areas mark buffer times added to the minimal headways. These additional buffer times give operational stability to a planned schedule in the case of unforeseen delays. To determine them a balance has to be found between desired robustness of a schedule and the maximum usage of track capacity.

Blocking time stairways can also illustrate the influence of the signalling system and the sequence of trains on the capacity of a railway network. Figure 3 illustrates how the throughput of a line is increased by arranging trains in such a way that trains of the same type, i.e., with blocking time stairways of identical slope, follow each other. It can be seen that the minimal headways between trains of the same type is typically smaller than those between trains of different types. This so-called *bundling* of trains is



(a) Before bundling (b) After bundling

Figure 3: A set of trains with blocking times, before and after bundling.

a popular technique of manual schedule optimization. Bundling is a combinatorial problem. Besides train selection, bundling is the main source of potential in any optimization approach to railway scheduling; we have also implemented bundling in our track allocation module OPTRA.

3.2 Macroscopic Infrastructure Model

A combinatorial optimization of railway scheduling based on a microscopic model is not possible with the current computing and mathematical technology; it is, in fact, already difficult to *simulate* rail traffic at this level of detail. We therefore propose a *macroscopic model* that is detailed enough to capture the essential characteristics of a railway network and coarse enough to be amenable to mathematical optimization approaches.

The first component of our model is an infrastructure *network* N with nodes and arcs, i.e., a (directed) graph, Figure 1(b). The *nodes* of this graph represent places where trains can pass each other: stations and line crossings. Note that complex stations can be handled by refining the model appropriately, introducing sufficient numbers of additional nodes. We denote nodes by lower case letters like a, b or v_1, v_2, v_3.

Nodes are connected by *arcs* that model railway line segments. We shall assume that arcs are dedicated to a specific driving direction; bidirectional tracks can be modelled by introducing a pair of antiparallel arcs. I.e., a physical track segment between stations a and b, which can be used in both directions, is represented by two directed arcs (a, b) and (b, a), oriented in opposite directions, while a physical line segment, that can only be used in one direction, corresponds to exactly one arc. We assume that it is not possible to overtake on an arc. Rather, overtaking is modelled using an additional node, namely, a simple line crossing. We denote by $N = (V_N, A_N)$ the infrastructure network consisting of the set of all nodes V_N and the set of all connecting arcs A_N.

Figure 1(b) illustrates this macroscopic railway description (labels B, D, and κ will be explained below). We currently work with a subnet of the German railway network that is bounded by the lines connecting Hannover, Kassel, Fulda and Braunschweig.

3.3 Train Types

As a second component, we group trains into *train types* rather than considering an unlimited variety of individual driving dynamics. The classification of train types is based on the following properties:

- similar driving characteristics
- equivalent train protection system
- similar intermediate stops within the arcs
- equivalent service type (passenger vs. freight, local vs. long distance).

Examples for train types, that we currently use, are: InterCityExpress *ICE*, InterCity *IC*, RegionalExpress, RegionalTrain, and InterCargoTrain. The set of all train types is denoted Y, a single train type by $y \in Y$.

We associate with each train type parameters that describe standardized characteristics such as maximum length, acceleration, and maximum speed. We must, of course, make sure that every train can meet the characteristics of its type. For example, the broader category LongDistancePassengerTrain, a pooled description for *ICE* and *IC*, is determined by the length and acceleration of an ICE and the maximum speed of an IC. Similar to the infrastructure, the train type model can be refined to an arbitrary level of detail by introducing additional train types.

We allow a *mimicking* of train types by other train types with superior driving characteristics. For example, it is allowed to slow down an *ICE* train on high speed line segment to the speed of an *IC* train, i.e., an *ICE* train can mimic an *IC* train (but not the other way round). Mimicking capabilities are stated in terms of a function $F(\cdot) : Y \to 2^Y$, which states that a train type y can mimic the behavior of all train types $z \in F(y)$.

3.4 Operational Constraints

We label the nodes and arcs of our infrastructure network with operational data that models a complete set of rules and constraints for all possible track allocations.

The individual arcs are labeled with the following data:

- $D_Y^{A_N} = (d_y^e) \rightarrow$ the vector of driving times[6] for all train types
- $B_{Y\times Y}^{A_N\times A_N} = (b_{y,z}^{e,f}) \rightarrow$ the matrix of the scheduled headways[7] for any *ordered pair* of train types for each pair of arcs $(e, f) \in A_N \times A_N$:
  - an identical pair of arcs describing the sequence of trains for one direction (i.e. one train follows another)
  - arcs corresponding to opposite directions on single tracks (i.e., opposing arcs)
  - arcs corresponding to same-level line crossings[8]
  - arcs corresponding to similar conflict cases (e.g., narrow tunnels).

In addition to the standardized driving style of the involved train types, the computation of driving times and headway matrices is based on a number of physical properties of the track that the corresponding arc represents, among them the *maximum allowed speed*, the *gradient profile of the track*, and the *track safety equipment*.

The nodes of the infrastructure network are also labeled with *capacities*, i.e., the number of trains that can pass and/or stop at a node at any point in time. We denote the track capacity of a node $v \in V_N$ by $\kappa_v \in \mathbb{N}$, and the vector of all such capacities by κ. We remark that such a label is clearly only a first step to model the operational constraints inside a station.

We summarize the macroscopic model that we have developed so far as consisting of an infrastructure network $N = (V_N, A_N)$, a set of train types Y, a type mimicking function F, a driving time matrix D_{Y,A_N}, a headway matrix $B_{Y\times Y}^{A_N\times A_N}$, and a node capacity vector κ.

3.5 Time Expanded Model

We finally expand our model along a discretized time axis to model timetables. Using a discretization of one minute over a time horizon, we construct multiple copies of the infrastructure node set, one node set for each minute. The arcs of the infrastructure network are also copied, connecting nodes in time layers that fit with the driving times. The result is a space-time network in which railway slots correspond to directed paths, proceeding in time.

[6]The driving time is the time needed to traverse the arc using a standardized driving style for the respective train type.

[7]The diagonal entries denote the headways for single arcs, i.e., $b_{y,z}^{e,e}$ is the headway for a train of type z following a train of type y on arc e, while the off-diagonal entries record the headways for ordered pairs of arcs, i.e., $b_{y,z}^{e,f}$ is the headway for a train of type z entering arc f after a train of type y entered arc e.

[8]There is one such instance in our sample network, the so-called "Hildesheimer Kurve".

The formal construction is as follows. We denote the time horizon by $T = \{t_0, \ldots, t_{\max}\} \subseteq \mathbb{Z}$, i.e., t_0 is the first minute of the time horizon and $t_{\max}$ the last. The set of time-nodes is $V_H = V_N \times T = \{(v,t) : v \in V_N, t \in T\}$, i.e., (v,t) is the copy of infrastructure node v at minute t. Waiting at node $v \in V_N$ is modeled by a time-arc $((v,t),(v,t+1),y)$ for each train type $y \in Y$ and for all $t \in \{t_0, \ldots, t_{\max} - 1\}$. We connect two time-nodes (u,τ) and (v,t) by a time-arc $((u,\tau),(v,t),y)$ of type y if nodes u and v are connected by an arc uv in the infrastructure network and if the driving time d_y^{uv} from u to v for a train of type y is equal to $t - \tau$, i.e., if $d_y^{uv} = t - \tau$; we denote the set of all such arcs by A_H. The space-time network consisting of all such time-nodes and time-arcs is denoted by $H = (V_H, A_H)$. We finally denote by $A_H^y = \{((u,\tau),(v,t),z) \in A_H : z \in F(y)\}$ the set of all time-arcs that a train of type y can use, driving as a train of its own type or mimicking some inferior type $z \in F(y)$, and by $H^y = (V_H, A_H^y)$ the space-time graph restricted to this set of arcs.

A first important property of this construction is that it captures the complete information that is needed to construct a feasible route for an individual train through our macroscopic infrastructure network. Namely, a *slot* for a train of type y corresponds to a directed path in the network H^y (and vice versa, i.e., macroscopic slots and space-time paths are in one-to-one correspondence). We will denote such a slot as a sequence $s = (y, (v_1,t_1), (v_2,t_2), \ldots, (v_n,t_n))$; it is understood that the time-arcs $((v_i,t_i),(v_{i+1},t_{i+1}),y)$ underlying a slot must exist.[9] Note that $v_{i+1} = v_i$ means that a train is waiting at v_i.

In a similar way, the space-time network can also be used to make all potential *conflicts* between two or more train slots explicit. In fact, each conflict corresponds to a so-called packing constraint, stating that a conflict free set of train slots can use only a certain maximum number of arcs out of an appropriately chosen set of space-time arcs. This works as follows. For a headway conflict, consider two train slots of type y_1 and y_2 running into the same infrastructure node v via arcs a_1 and a_2, arriving at times t_1 and t_2, respectively; let $t_1 \leq t_2$. There is a headway conflict between these slots if $t_2 < t_1 + b_{y_1,y_2}^{a_1,a_2}$. This conflict can be ruled out by stipulating the constraint that a conflict free set of slots can use only one of the arcs a_1 and a_2. Doing this for all pairs of conflicting arcs enforces correct minimum headways. For a capacity conflict, consider train slots s_i, $i = 1, \ldots, k$, entering a time-node (v,t) of capacity κ_v. The capacity of node v at time t is exceeded if more than κ_v trains arrive at this station at time t. This conflict can be ruled out in a similar way as before by stipulating the constraint that a conflict free set of slots can use at most κ_v of the arcs that enter node v at time t. We

[9] For clarity we point out that a (planned) slot of an actual timetable s must be distinguished from a *slot request* r, to be explained in the next section. The latter one will also specify a monetary bid; on the other hand, the time-nodes of a slot request need not be fully specified so as to leave some room for the optimization procedure to arrange slots.

denote such a conflict by c, the set of all conflicts by $\mathcal{C}$, the set of conflict arcs associated to conflict c by A_c, and the maximum number of arcs from A_c that a conflict-free set of slots can use by κ_c.

We can now define that two slots s and s' are in conflict if they contain more than κ_c arcs from some conflict arc set A_c; otherwise, s and s' are conflict-free. A slot or track allocation is a set of mutually conflict-free slots.

With these defintions, our final, time-expanded macroscopic railway model is quite simple to state: It consists of a space-time network $H = (V_H, A_H)$ over a time horizon $T = \{t_0, \ldots, t_{\max}\}$, with subnetworks $H^y = (V_H, A_H^y)$ for each train type y out of some set of train types Y, and a set of conflicts $\mathcal{C}$ with associated limits κ_c for each $c \in \mathcal{C}$.

4 Auction Design

Auctions can be understood as *formalized negotiation procedures.* The auction design describes the rules which these negotiations follow. It can be crucial about the auction's success or failure. Two aspects characterize a negotiation: the *economic good* that is negotiated about, and the *negotiation procedure*, that is, the sequence of offers and counter offers and their varying levels of commitment. In an auction, the *bidding language* describes the subjects or goods that are auctioned, and the *auction procedure* describes the rules for submission of bids and how final contracts, including payment obligations, are constructed.

Operating trains involves, like any other enterprize, risks. The philosophy of our auction design is that train operating companies (TOCs) should bear those risks that they possess means to control: it is in the operator's core competence to have some estimation about demand for a certain connection, and consequently, a train operator should bear the risk involved by such a prognosis. Similarly, a train operator has to take into consideration the competitive situation. Therefore, we will allocate as many slots as feasible rather than selling exclusive rights for certain lines.

On the other hand, the risk of getting slots that are technically infeasible to operate should not be put on operators. No TOC can generate social welfare from, say, the right to use an isolated track segment without any connecting segments. We have designed our auction language with the goal of minimizing these risks of "technical nature": operators submitting a bid for a tour can be certain that it will be technically feasible to run their trains as planned, if they provide for suitable rolling stock. We have also included possibilities to bid for connections and regular services in order to allow that positive network effects can be realized.

4.1 Bidding Language

The economic good that is auctioned in the railway slot exchange is *the right to run scheduled trains*. The bidding language allows train operators to describe their preferences on running their business. Its design is based on the following principles:

- *Flexibility and tractability.* Train operators are often indifferent about the exact route of a train between two scheduled stops (as long as it is consistent with the specified rolling stock) or about the exact timing of a train. For network optimization it is very important to know about bidders' degrees of flexibility. In principle, it would be possible to express such flexibility as exclusive "either-or" bids ("XOR"). However, for tractability, bidders should be given the opportunity to express flexibility about route and time in a more concise way. At the same time, flexibility in other, less essential aspects will be restricted in order to reduce the complexity of bidding and optimization. For example, we will not allow a bid for "either a slot from A to B, or one from C to D". In a similar vein we will also restrict the ability to make "AND" bids for different slots; we will consider bids that allow for rolling stock management and for some important interdependencies between different trains, but not for arbitrary interdependencies.

- *Bids that allow rolling stock management.* For instance, an operator may specify that scheduled trains from A to B and back are served by a shuttle. This has the consequence that the arrival in B must take place before the departure from there, plus some time for turn-around.

- *Bids can express specified interdependencies between trains.* For instance, an operator may request that a train arriving in B has connection to C for passengers, or an operator prefers that his train is part of a regular service timetable.

- *Only "socially desirable" interdependencies should be expressible.* A dominant firm may be tempted to make all bids contingent on each other, such that they would have to be accepted en block. In order to avoid this, rules should be set up that will confine interdependencies of bids.

In what follows, the bidding language will be outlined and discussed, but not formalized. In fact, instead of mathematical formulation the bidding language will be implemented as software package. Flexibility of bids, as well as its limits, will take the form of limited options on a user (bidder) interface.

4.1.1 Slot requests and tours

A *slot request* describes a certain scheduled train. A slot request must specify (at least): a monetary base bid, a train type, a route, and time-value specifications. The *monetary base bid* is a sum of b_s€ the operator is willing to pay for the slot s, aside any additions or reduction. The *train type* specifies the technical characteristics like maximum speed, acceleration, braking distance, weight etc. of a scheduled train; see section 3.2. The *route* of the train is given by a sequence of stations, denoted $v_1, v_2, \ldots, v_n$. A station can be *mandatory* or *optional*. In the latter case, an *additional monetary value* is assigned to the bid if the allocation allows serving the station.

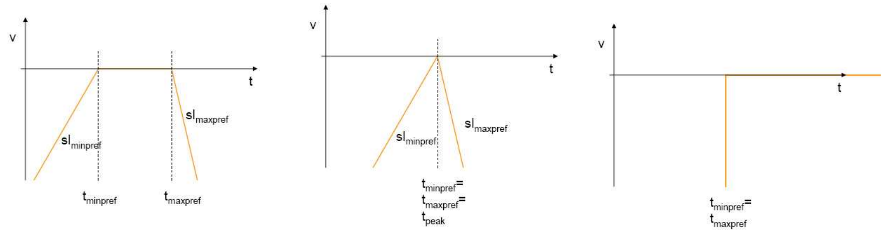
Modelling time flexibility. Railway schedules are precise up to the minute. Train operators, however, will allow some flexibility regarding departure and arrival times of their trains: it is generally not important whether a train is scheduled for 8:03 or 8:13. Flexibility can be expressed via *time-value specification* which describes a preference, expressed in monetary terms, for an arrival or departure time. A time-value specification can be either *absolute* ("9:00 in the morning") or *relative* to some other time ("one hour after departure"). We assume that the functional dependency between value and time is piecewise linear and can be expressed by

$$v(t) = \begin{cases} -sl_{\text{minpref}}(t_{\text{minpref}} - t) & \text{if } t < t_{\text{minpref}} \\ 0 \text{ if } t_{\text{minpref}} \leq t \leq t_{\text{maxpref}} & \\ -sl_{\text{maxpref}}(t - t_{\text{maxpref}}) & \text{if } t > t_{\text{maxpref}} \end{cases} \tag{1}$$

for a tuple $(t_{\text{minpref}}, t_{\text{maxpref}}, sl_{\text{minpref}}, sl_{\text{maxpref}})$. This means *indifference* between t_{minpref} and t_{maxpref}, and a linearly decreasing value, with rate sl ("slope"), for deviation outside these boundaries. Figure 4 shows some typical time value specifications: Figure 4(a) shows the "general" type, figure 4(b) a preference for a peak time with flexibility to both sides, and figure 4(c) shows a value for "anytime after".

Example 4.1 (An *ICE* train). An operator who wishes to run a long-distance high-speed train between two major cities A and B may formulate a slot request as follows:

- depart at A between 7:45 and 8:15; with every minute earlier or later involves a 100€ penalty
- arrive at B 3 hours after departure, with every additional minute of journey time involves a 200€ penalty
- optionally stop at C for an additional value of 1000€



(a) General

(b) Peak ($t_{\text{minpref}} = t_{\text{maxpref}}$)

(c) Anytime after ($sl_{\text{minpref}} = \infty$, $sl_{\text{maxpref}} = 0$, $t_{\text{minpref}} = t_{\text{maxpref}}$)

Figure 4: Example time-value specifications

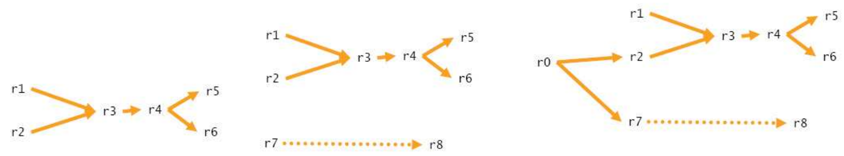
- use rolling stock type *ICE-Series 402* (with well-known maximum speed, acceleration, etc.
- base bid 13000€

Example 4.2 (An overnight cargo express). An operator who wishes to run an overnight cargo express that connects a port P and a factory F may formulate the following slot request:

- depart at P anytime after 20:00, no flexibility for early departure
- arrive at F anytime before 8:00, with every minute late involving a 1000€ penalty
- use rolling stock type *cargo, weight 1000 tons*
- base bid 8000€

Tours. Suppose a train operator wants to run a train from A to B and split it in B with one part continuing to C and another part continuing to D. He could submit three independent slot requests from A to B, from B to C, and from B to D. If there is flexibility in the arrival time at B, the operator would have to schedule his trains departing from B with an excessive buffer time to make sure that no matter when arrival at B takes place there is sufficient time for splitting the train and getting off with the two parts (not knowing the exact pieces of track that are used for the arriving and departing trains). This would lead to inefficient schedules.

Therefore, we allow train operators to explicitly group slot requests to *tours*. A tour is a group of slot requests connected by the $\rightarrow$ relation. Figure 5(a) shows a tour where $r1$ and $r2$ are merged into train $r3$, $r4$ reuses stock from $r3$ and is split into $r5$ and $r6$. To avoid that operators arbitrarily



(a) An example tour (b) Not a tour (c) Corrected

Figure 5: Tour bids

group slot requests to tour bids, we require that *all* slot requests in a tour are connected via the $\rightarrow$ relation, that is, any two slots have a common predecessor or successor in regard to $\rightarrow$. Figure 5(b) shows an example where this condition is violated: Indeed, $r7$ and $r8$ don't share stock with $r1$ to $r6$, and these slot requests should not be combined into one tour. The situation changes in figure 5(c): now $r2$ and $r7$ have a common predecessor $r0$ that they share stock with.

Note that a tour bid specifies re-arrangement of rolling stock, but does *not* specify the shunting moves that may be necessary to temporarily remove stock from a platform. We assume that slots necessary for shunting moves are reserved from the side of the auction mechanism and are included in the payment for the bid. The general business terms will give details on reasonable shunting moves in regard of fuel consumption and staff requirements.

So the economic goods that train operators bid for are *tours*. A tour bid consists of a bid amount and a set of slot requests, including time-value specifications to describe flexibilities, grouped to a tour via a $\rightarrow$ relation. The overall size of tours will, however, be restricted.

4.1.2 Operator-neutral connection requests

A major strength of railways is the ability to offer interlining connections for passengers with lower transfer times than busses or planes. In comparison to busses, rail schedules are more reliable and thus less buffer time is required than for bus connections. Transfer between planes is subject to security considerations: passengers can't walk across the airport from one plane to another; they have to pass through arrival and departure gates. Of course, also the handling of luggage takes time, while rail passengers generally carry their luggage from train to train or use porters.

Optimizing connections is part of the schedule design process. We can't expect that the same quality will result as an outcome of an auction for independent point-to-point connections. On the other hand, a dominating market player should not be allowed to make all his bids contingent on each other by connection requests. Therefore, connection requests will be *operator-neutral*, that is, ask for destinations rather than for specific trains. As a consequence, a connection request may be satisfied by the same operator asking for it or by any other train operating company.

A *connection request* for a certain slot consists of the following parts:

- a *value* that is added to the bid if the request is satisfied,
- a *type* which is either *connection-from*, or *connection-to*,
- a *branching station*, and
- a *connection target station* with a *connection time-value specification*.

The branching station determines where the transfer takes place: for connection-from request, transfer can take place at or before the branching station, and for connection-to requests, transfer can take place at or after the branching station. The connection time-value specification describes the preference about when the connecting train departs (for connection-from requests), or arrives (for connection-to requests) at the target station.

Example 4.3 (A connection-from request)*.* For a slot request running from V_1 through $V_2, \ldots, V_5$ to V_6, a connection-from request may require that

- passengers from O (the target station) may reach the train at or before station V_4 (the branching station)
- having to leave O no earlier than 8:00, with every minute earlier involving a 100€ penalty (this is expressed in the connection time-value specification)
- if the request is satisfied, an additional 1000€ is added to the tour bid

Example 4.4 (A connection-to request)*.* For a slot request running from V_1 through $V_2, \ldots, V_5$ to V_6, a connection-to request may require that

- at or after station V_2, passengers may transfer to a train for D
- reaching D three hours after departure at V_1, with every late minute involving a 200€ penalty
- if the request is satisfied, an additional 500€ is added to the tour bid

4.1.3 Regular service conditions

Regular (basic interval) services give travellers (or users of freight services) predictability on the availability of services during the day. Although not undisputed, there is evidence that, at least for mass transit, availability of regular services through the whole day is attractive to passengers. Train operators can compose regular services by placing bids with *regular service conditions*. As in the case of the connection requests, regular service conditions are operator-neutral. It is perfectly feasible that some operator has a service to B leaving station A 12 minutes after the full hour from 8:12 to 17:12, except for the 9:12 service that is served by a competitor who outbade him.

A *regular service condition* for a slot request for stations $V_1, \ldots, V_n$ has the following components:

- a subset of the stations $V_1, \ldots, V_n$ that is targeted by the condition,
- the offset of the first and last services required,
- the service interval,
- an elasticity term that states how much value is lost for a minute deviation from the rule (per station and service).

Example 4.5 (A slot request with regular service conditions). An operator wants to run hourly service between stations V_1 and V_4. Stations V_2 and V_3 should be served alternatingly, that is, in two-hour intervals. Service should commence 8:00 in the morning and terminate about 20:00. The operator will submit bids for *all* services involved. Even if only part of his bids are accepted, he may be sure that all stations are served as requested. The bid for the 9:00 (i.e. the second) service may look as follows:

- slot request for $V_1, V_2(optional), V_3, V_4$. Departure at V_1 between 7:55 and 8:05, reaching V_4 at most 1 hour later
- regular service condition for V_1, V_2, V_4 with offset -1 and +13 for first and last services, 120 minutes service interval and elasticity 100€ per minute
- regular service condition for A_1, A_3, A_4 with offset 0 and +12 for first and last services, 120 minutes service interval and elasticity 100€ per minute

4.1.4 Interpretation of the language and sector-specific views of the bidding language

To understand the bidding language set out above, it is important to note that the language is complemented by a set of *interpreting rules* that ensure that train operators use the acquired slots in the intended way. These rules may, among other things,

- state conditions on returning acquired slots and possibly penalties for unused slots
- set default times for the duration of intermediate stops for the different train types (examples are standard stopping times at stations for passenger trains or standard switching times between two slots of a tour for passenger resp. freight trains),
- define restrictions for expressing tour bids, connection requests, or regular service for different train types

There are two reasons to restrict bids in regard of tour size and the number of connection requests and regular service rules:

- Nested bids put strain on the optimization procedure, both by increasing its complexity and reducing its flexibility (these are, in fact, two kinds of potential negative externalities of bid interdependency on the overall allocation).
- Some bidders may have strategic reasons to tie bids that are in reality unconnected or only loosely connected.

However, rather than blanket restrictions for all train operators, we define *sector-specific views* of the bidding language. A view here is a subset of the expressive power of the unrestricted language.

For instance, in passenger services, a slot is intended to be a simple run of an unchanged train from A to B. Tours for passenger trains typically connect two slots to implement a trip and its return, while the number of branches and merges along the route is small. Connection requests and regular service conditions play an important role in passenger services. Thus a passenger train operator's view of the language could allow only a few slots per tour bid while being as generously as computation power permits in regard of connection requests and regular service rules.

On the other hand, single wagon load transport might require more complex tours for those slots that are scheduled periodically (the slots that are acquired spontaneously on a day-to-day basis are outside the scope of this paper). The corresponding view would allow quite complex construction of tours but disallow connection requests and regular service rules.

4.1.5 No exclusive bids

Apart from the tours, connection requests, and regular service conditions, no other interdependencies of slots will be allowed. In particular, "exclusive bids" are ruled out. This has two aspects. First, it is not possible to bid for non-connectivity of a competitor's train. In contrast, a competitor's quest for connectivity will be honored. Thus, our bidding language fosters both competition and cooperation between competitors. It is left to the TOCs whether they will offer some kind of through-ticketing to passengers or not.[10] A second aspect of non-exclusivity is that TOCs cannot make their bids contingent on being the sole provider on a line (or more general, on the number of competitors). As in other markets, the risks of competition should not be removed from firms.

This may lead to excessive entry on particular lines up to the point that TOCs will not break even. Consider, for example, a line in a sparsely populated area. It might be that only one passenger train every two hours will be able to cover costs. However, there are many more slots available, so it is possible that two TOCs enter and make losses. Such phenomena are well known from other markets, and firms usually find ways to coordinate (for example, once a firm is in, others will stay out). However, due to the simultaneity of an auction there is danger that too many bidders get allotted. To avoid such accidents, bidders may be allowed to announce and partly coordinate (by cheap talk) their plans in advance. One might also allow some kind of corrections *after* the auction.

4.2 Auction procedure

The procedure envisioned for the railway slot auction is an *iterative combinatorial auction* in the flavor of Parkes' "ibundle" auction. It will only be sketched here. Figure 6 gives an overview. The auction takes place in a sequel of *rounds*. A round consists of two stages:

- In the first stage, operators simultaneously submit each a set of bids. The set of newly submitted bids together with the set of *standing* bids from the previous round forms the set of *live* bids.

- The allocator applies his optimization machinery on the set of live bids and computes the set of bids that are accepted in this round[11], that is, the set of standing bids for this round. Rejected bids are considered *dead*, that is, they are not considered in further rounds, and the submitter has no commitment to them anymore. Every participant gets full information about all submitted bids and their state.

The auction *ends* if for a fixed number of rounds, the total proceed does not change[12]. The procedure of optimization implies that the proceed is

[10] There might be some general rules calling for standardized tickets or through-ticketing. This is not our main field of interest, though.

[11] We assume that in case of ties, the allocator prefers to leave the set of standing bids unchanged.

[12] Note that to make this happen, it is sufficient but not necessary that the set of standing bids changes.

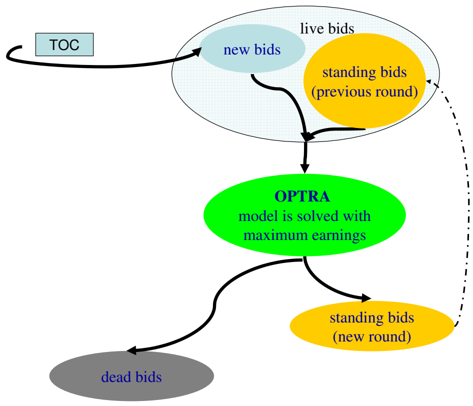


Figure 6: Auction procedure overview

non-decreasing from round to round.

5 Optimal Track Allocation

In every round of our rail track auction the auctioneer has collected a set of bids for slots and must determine a conflict-free slot schedule that maximizes the network proceeds. This winner determination problem, that we denote as the *optimal track allocation problem* (OPTRA), is a combinatorial optimization problem, namely, a so-called *multi-commodity flow problem* with additional constraints. Such problems can be solved with integer programming techniques. In a concrete auctioning approach, a generic track allocation model will be extended in various ways to handle the particular types of bids that are allowed and the auction rules that are used. We will first introduce a basic model for a simple setting, and then discuss an extended version that we are currently using. Even more complex models along the lines of Section 4 can be constructed in a similar way.

5.1 Basic Integer Programming Model

Consider a basic rail track auction setting that allows only simple bids for individual slots of the following form: bid i specifies a train type y^i, a monetary value b^i, a departure station v_1^i and time t_1^i, and an arrival station v_2^i and time t_2^i; denote by $d^i := (v_1^i, t_1^i)$ and $a^i := (v_1^i, t_1^i)$, the time-nodes associated with the departure and the arrival of a slot associated with bid i, respectively. Let I be the set of all bids. If bid i gets assigned, the auctioneer must provide a conflict-free slot according to the bid specifications, and the value of the bid is b^i. The degrees of freedom in this setting are bid assignment and slot routing in space and time.

To formulate the track allocation problem for each round of such an auction as an integer program, we introduce a zero-one variable x_a^i (i.e., a variable that is allowed to take values 0 and 1 only) for each bid i and each arc a of the space-time network. If x_a^i takes a value of 1 in an OPTRA solution, this means that a slot associated with bid i passes through arc a; clearly, this implies that bid i has been assigned. $x_a^i = 0$, on the other hand, means that arc a is not used by a slot associated with bid i, independent of whether bid i is assigned or not. We further introduce proceedings values p_a^i for each bid i and each arc a in order to account for the overall proceedings of a track allocation. We set $p_a^i = b^i$ for each arc of the form $((v_1^i, t_1^i), (w,t), z)$ with $z \in F(y^i)$, i.e., each arc that qualifies as a first arc (an arc leaving the starting node v_1^i at time t_1^i) in a slot associated with bid i. On all other arcs, we set $p_a^i = 0$. In a slot associated with bid i, the first arc will contribute a proceedings value of b^i, while all other arcs contribute 0. Hence, summing over all arcs of a slot yields exactly the value of the associated bid.

Let us finally denote by $\delta_{\text{in}}^i(v,t) := \{((u,s),(v,t),z) \in A^{y^i} : z \in F(y^i)\}$ the set of all arcs entering a time-node (v,t) that are compatible with train type y^i. Similarly, let $\delta_{\text{out}}^i(v,s) := \{((v,s),(w,t),z) \in A^{y^i} : z \in F(y^i)\}$ be the respective set of arcs leaving time-node (v,t).

With these definitions the track allocation problem can be formulated as the following integer program:

$$
\begin{array}{llrll}
\text{(OPTRA)} & \text{(i)} & \max \displaystyle\sum_{i \in I} \sum_{a \in A_H} p_a^i x_a^i & & \\
& \text{(ii)} & \displaystyle\sum_{a \in \delta_{\text{out}}^i(d^i)} x_a^i & \leq 1, & \forall i \in I \\
& \text{(iii)} & \displaystyle\sum_{a \in \delta_{\text{in}}^i(a^i)} x_a^i & \leq 1, & \forall i \in I \\
& \text{(iv)} & \displaystyle\sum_{a \in \delta_{\text{out}}^i(v)} x_a^i - \sum_{a \in \delta_{\text{in}}^i(v)} x_a^i & = 0, & \forall i \in I, v \in V_H \setminus \{d^i, a^i\} \\
& \text{(v)} & \displaystyle\sum_{i \in I} \sum_{a \in A_c} x_a^i & \leq \kappa_c, & \forall c \in \mathcal{C} \\
& \text{(vi)} & x_a^i & \in \{0,1\}, & \forall i \in I, a \in A_H
\end{array}
$$

In this model, the integrality constraints (vi) state that the arc variables take only values of 0 and 1. Constraints (ii)–(iv) are flow constraints for each bid i; together, they guarantee that, in any solution of the problem, the arc variables associated with bid i are either set to 1 if and only if they lie on a path from the departure to the arrival time-node in the space-time network, i.e., they describe a feasible slot associated with bid i, or, otherwise, they are all set to 0, i.e., no slot is assigned to bid i[13]. Constraints (v) rule out headway and node capacity conflicts as discussed in Section 3. The objective function (i) maximizes total network proceedings by summing over all arcs proceedings.

5.2 Extended Integer Programming Model

A main point in the discussion on combinatorial railway auctions is whether it is possible to deal with complex technical and economical constraints in a real-world setting or not. We do, of course, not claim that we can give a real answer to this question, but we want to give an example of a more realistic scenario to indicate that our approach has potential in this direction. To this purpose, we discuss a setting that extends the previous one by allowing for bids with stops at intermediate stations, time windows and penalties for deviations from desired departure and arrival times, and combinatorial AND and XOR bids. This extended model is the one that we use in the computational experiments in Section 6.

With these extensions, it is possible to model most features of the bidding language described in section 4.1. Bids for complete tours can be expressed as AND-connected bids, and an optional stop can be expressed as a XOR-connection of requests for slots with and without this stop. However, it is at present not possible to express an operator-neutral bid for connections or regular service. We hope to remove these limitations in the future.

Let bid i for a single slot specify a train type y^i, a basic monetary value b^i, a departure station v_1^i, an ordered sequence of intermediate stations $v_2^i, \ldots, v_{k_i-1}^i$, possibly empty, and an arrival station $v_{k_i}^i$, a departure time window $[s_1^i, t_1^i]$, halting time windows $[s_j^i, t_j^i]$ and halting durations h_j^i at the intermediate stations $j = 2, \ldots, k_i - 1$, an arrival time window $[s_{k_i}^i, t_{k_i}^i]$, a penalty for α^i for each minute of travel time stretching, and a penalty β^i for each minute of delay at the departure station. A slot corresponding to such a bid must start at the departure station, pass through and stop at the intermediate stations, and stop at the arrival station. The slot must provide for a stop of at least h_j^i minutes at each intermediate station v_j^i within the time interval $[s_j^i, t_j^i]$. It is allowed to depart and arrive later within the arrival and the departure time windows, respectively. However, for each minute of lateness with respect to the earliest departure and arrival time, a penalty of α^i and β^i is subtracted from the bid value b^i, i.e., a slot for bid i that departs δ^i minutes later than the earliest departure time s_1^i, and arrives γ^i minutes later than the earliest arrival time $s_{k_i}^i$, costs only $b^i - \alpha^i\delta^i - \beta^i\gamma^i$ instead of b^i.

Let a combinatorial bid j refer to some subset $I_j \subseteq I$ of bids for single

[13]Note that the underlying space-time network H is acyclic such that no directed cycles can come up.

slots; it may either be an AND or an XOR bid. An AND-bid stipulates that either all single slot bids in I_j must be assignend or none of them. A XOR-bid states that at most one of the bids in the set I_j can be assigned. Let J_{AND} denote the set of AND bids, and J_{XOR} the set of XOR bids.

To model this situation, it is convenient to extend the space-time network H by nodes and arcs that model departures and arrivals of slots explicitly. For every bid i, we introduce corresponding departure and arrival time-nodes $d^i := (v_d^i, s_1^i)$ and $a^i := (v_a^i, t_{k_1}^i)$, departure arcs $((v_d^i, s_1^i), (v_1^i, t), y^i)$ for $s_1^i \leq t \leq t_1^i$, and arrival arcs $((v_{k_i}^i, t), (v_a^i, t_{k_i}^i), y^i)$ for $s_{k_i}^i \leq t \leq t_{k_i}^i$; let A_d^i and A_a^i denote the set of such departure and arrival arcs for bid i, respectively. We also introduce corresponding zero-one variables x_a^i for $a \in A_d^i \cup A_a^i$. Let moreover $A_j^i := \{((v_j^i, t), (v_j^i, t+1), z) : s_j^i \leq t \leq t+1 \leq t_j^i, z \in F(y^i)\}$ denote the set of arcs that correspond to a stop of a slot associated with bid i at the intermediate station j, $j = 2, \ldots, k_i - 1$.

Arc proceedings are defined as follows:

$$p_a^i := \begin{cases} b^i - \beta^i(t - s_1^i) + \alpha^i(s_{k_i}^i - t), & \forall a = ((v_d^i, s_1^i), (v_1^i, t), y^i), \\ -\alpha^i(t - s), & \forall a = ((u, s), (v, t), z) \in A_H, z \in F(y^i), \\ 0, & \forall ((v_{k_i}^i, t), (v_a^i, t_{k_i}^i), y^i). \end{cases}$$

Departure arcs bear the bid value, minus a delay penalty, plus a bonus term for the driving time up to the earliest arrival time, i.e., the time difference between the earliest arrival and departure times stated in the bid. This bonus is reduced by summing over the network arcs associated with a slot; at the earliest stated arrival time $s_{k_i}^i$, the bonus has vanished, and every additional minute of delay is penalized.

Finally, we introduce a zero-one variable z^i for each bid i that is 1 if bid i is assigned and 0 else; these variables are useful in dealing with combinatorial bids.

With these definitons, our extended track allocation model reads:

$$
\begin{array}{llrll}
\text{(xOPTRA)} & \text{(i)} & \max \quad \displaystyle\sum_{i\in I}\sum_{a\in A_H} p_a^i x_a^i + \sum_{i\in I}\sum_{a\in A_d^i} p_a^i x_a^i & & \\
& \text{(ii)} & \displaystyle\sum_{a\in\delta^i_{\text{out}}(d^i)} x_a^i & = z^i, & \forall i\in I \\
& \text{(iii)} & \displaystyle\sum_{a\in\delta^i_{\text{in}}(a^i)} x_a^i & \le 1, & \forall i\in I \\
& \text{(iv)} & \displaystyle\sum_{a\in\delta^i_{\text{out}}(v)} x_a^i - \sum_{a\in\delta^i_{\text{in}}(v)} x_a^i & = 0, & \forall i\in I, v\in V_H \\
& \text{(v)} & \displaystyle\sum_{a\in A^i_j} x_a^i & \ge d^i_j z^i, & \forall i\in I, j=2,\dots,k_i-1 \\
& \text{(vi)} & \displaystyle\sum_{i\in I}\sum_{a\in A_c} x_a^i & \le \kappa_c, & \forall c\in\mathcal{C} \\
& \text{(vii)} & z^i - z^k & = 0, & \forall j\in J_{\text{AND}}, i,k\in I_j \\
& \text{(viii)} & \displaystyle\sum_{i\in I_j} z^i & \le 1, & \forall j\in J_{\text{XOR}} \\
& \text{(ix)} & x_a^i & \in\{0,1\}, & \forall i\in I, a\in A_H\cup A^i_d\cup A^i_a \\
& \text{(x)} & z^i & \in\{0,1\}, & \forall i\in I
\end{array}
$$

Similar as before, the flow constraints (ii)–(iv) ensure that if bid i is assigned, its associated slot corresponds to a single path between the stated departure and arrivals nodes, this time in the extended space-time network; note that constraints (ii) couple the bid and the slot variables. Constraints (v) force appropriate stops at intermediate nodes, (vi) rules out headway and capacity conflicts. Constraints (vii) and (viii) enforce combinatorial AND and XOR bids. The final constraints (ix) and (x) are the integrality constraints.

5.3 Solution Method

It is well known that multi-commodity flow problems such as the optimal track allocation problem and its variants are difficult in a mathematically precise sense: they belong to a class of so-called $\mathcal{NP}$-hard optimization problems ([Garey & Johnson 1979, ND38]). This means that nobody can *guarantee* to solve an arbitrary instance of such a problem in reasonble time to *proven optimality*. This does, however, not mean that one can not tackle such problems; it only means that it may happen that one encounters very hard instances. How often this will occur in practice, and how large and complex the problems one can solve will be is a matter of computational experimentation and, of course, skill.

We have implemented a program to solve instances of the extended track allocation model (xOPTRA) using state-of-the-art mathematical programming tools. In our track allocation system, the integer programming model (xOPTRA) is set up using the publicly available model generator `zimpl`, see Koch [2001] for details. Such a model can be solved using any of several available powerful commercial integer programming solvers; we use CPLEX, see CPLEX [2005]. Even larger scenarios can probably be solved using special purpose methods; this will be a topic of future research.

6 Computational Results

We have performed a number of simulation experiments to put our auctioning approach to the test. The experiments investigate (i) the size and complexity of track allocation problem that can/can not be handeld at present, (ii) the practicability of our rail track auction, and (iii) the existence of econonomic potentials for a better utilization of the railway infrastructure, see, e.g., Fig. 7.

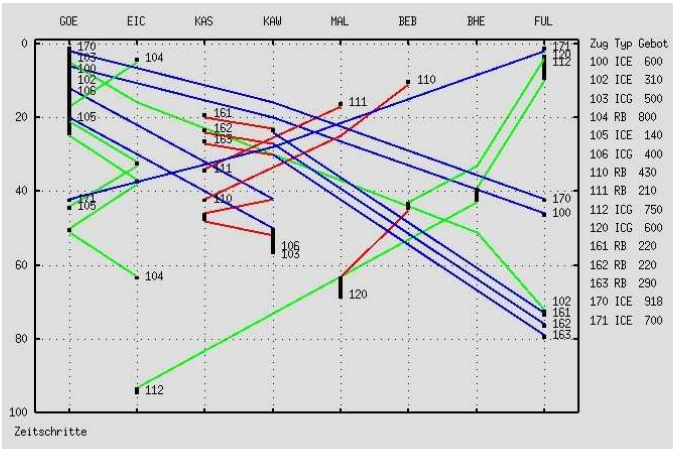


Figure 7: Bundling effects in a rail track auction.

All of our experiments are based on a macroscopic model of a subnetwork of the German long-distance railway network in the area around Hannover-Kassel-Fulda as described in Section 3. This network contains 31 station nodes and 90 arcs corresponding to 45 line segments. We consider the operation of this network over a time horizon of six hours (09:00–15:00) on a week-day containing 310 trains basing on the 2002 timetable of the Deutsche Bahn AG.

The computations themselves were made single threaded on a Dell Precision 650 PC with 2GB of main memory and a dual Intel Xeon 2.6 GHz CPU running SUSE Linux 9.3. The track allocation problems were solved with CPLEX 9.1.

6.1 Tripling

Our first test is a simple artificial experiment that looks for capacity reserves in the current timetable. We took the 310 train slots of the DB timetable and tripled them by adding two copies for each slot, the first 15 minutes later, and the second 30 minutes later. Adding some more trains that we knew had run earlier in our test scenario produced a set of 946 candidate train slots.

We turned these slots into bids by fixing the departure and the arrival station, no intermediate stations were specified. The departure time window for each bid was set to $[t, t+d]$, where t is the departure time of the underlying slot and d ist a parameter that defines the size of the departure window (in minutes); arrival time windows were set to $[t, t_{\max}]$, were t is the arrival time of the underlying slot; the penalites for late departure and arrival were set to 0. Setting parameter $d = 0, 1, 2, 3, 4, 5$ uniformly for all bids produces five sets of bids with increasing flexibility. A willingness to pay was assigned to the bids at random from a uniform distribution.

Table 1: Results of the tripling experiment.

d	computing time	# scheduled trains	objective value
0	6 sec.	420	52.066
1	8 sec.	496	60.612
4	1 day	617	67.069
5	3 days	737	67.975

Table 1 lists the results of a corresponding 1-round auction. We are aware that the tripling experiment is crude and do not want to overinterpret the results. We nevertheless find the increase in the number of scheduled slots as well as in network proceedings remarkable and encouraging. We also note that the computation time increases rapidly if more degrees of freedom are introduced; this is due to more complex track allocation problems. The scenario for $d = 5$ marks about the limits of size and complexity that our current implementation can handle.

6.2 Competition

Our second experiment is based on a *bid generator* that tries to anticipate bids of potential future participants in a railway auction, such as point-to-point low cost passenger and cargo carriers, regional carriers, etc. Space limitations prevent us from giving a comprehensive description here; suffice it to say that the generator uses a gravitational demand model, based on population and production statistics, to estimate costs and sales, and to produce 'reasonable' bids using several line generation techniques; a thorough description can be found in Reuter [2005]. The generator supports two ways of assigning a willingness to pay to a bid: by expected revenues as sketched above, and by applying the track price system TPS 2005 of Deutsche Bahn AG; the track prices were retrieved from the internet.

Our experiments add bids for trains of various types to the bids associated with the current timetable, namely, bids for individual IC and ICE trains (IC/ICE i.), synchronized IC and ICE trains (IC/ICE s.), individual regional trains and commuter trains (R*/S i.), synchronized regional and commuter trains (R*/S s.), cargo trains (ICG), and all of these types together (*); this allows to study displacement effects for different types of railway traffic. Synchronized bids are submitted as a combinatorial AND-bid.

All auctions were implemented as an automated iterative proxy auction along the lines of Section 4.2. At the beginning of the auction, the proxy agents submit minimum prices for each bid; for AND-bids, the minimum price is the sum of the minimum prices of the individual bids. The auction proceeds in rounds. In each round, the track allocation problem associated with the current prices is solved and the winning bids are assigned. The proxy agents increase all non-assigned bids by a minimum increment of 20% up to the maximum willingness to pay, and the next round begins, until no more bids are incremented. The auction finishes with the last set of assigned bids.

(a) Bids prices according to German track price system

Scenario	ICE ind.	ICE syn.	IC ind.	IC syn.	RE ind.	RE syn.	RB ind.	RB syn.	S ind.	S syn.	ICG ind.	#R [1]	#T [2]
Timetable	27	0	27	0	38	19	87	23	0	61	28	0	310
+24 IC/ICE i.	30	0	29	0	38	19	85	23	0	61	25	18	310
+24 IC/ICE s.	24	9	27	9	36	19	83	19	0	58	26	22	322
+27 R*/S i.	27	0	25	0	44	19	89	23	5	58	27	20	326
+27 R*/S s.	27	0	27	0	36	19	83	32	0	62	27	30	337
+15 ICG i.	27	0	25	3	38	19	87	23	0	61	42	19	343
+66 *	28	0	25	3	38	25	85	29	2	55	31	29	322

(b) Bid prices according to expected revenue

Scenario	ICE ind.	ICE syn.	IC ind.	IC syn.	RE ind.	RE syn.	RB ind.	RB syn.	S ind.	S syn.	ICG ind.	#R [1]	#T [2]
Timetable	27	0	27	0	34	19	83	23	0	61	28	0	302 [3]
+24 IC/ICE i.	32	0	29	0	32	19	79	23	0	61	24	40	299
+24 IC/ICE s.	24	9	27	9	32	19	79	19	0	58	26	43	302
+27 R*/S i.	27	0	26	0	41	19	84	23	7	50	27	23	304
+27 R*/S s.	27	0	27	0	34	25	81	29	0	66	27	25	316
+15 ICG i.	27	0	27	0	34	19	83	23	0	61	42	9	316
+33 *	28	0	26	0	37	22	82	23	1	52	30	47	301

[1] Number of rounds until auction was finished.
[2] Number of trains scheduled in the solution of the last auction round.
[3] Because of the revenue oriented willingness to pay some trains had a negative profit value and were not scheduled.

Table 2: Results of bid generator-based experiments.

Tables 1(a) and 1(b) show the results of the corresponding multi-round proxy auctions, more details can again be found in Reuter [2005]. We observe that synchronized traffic of any type (IC*,R*,S) seems to be stable in our experiments. Competition is much more relevant for individual bids. In particular, it turns out that capacities are available for flexible bids for long distance or cargo traffic. This is the result that one would like to see in a real-world scenario.

7 Outlook

The vision of our project is to supply a package of auction design and software implementation that is practically ready for use. Our discussion so far has not completely settled a couple of prerequisites for this ambitious goal:

- Our bidding language is at this point incomplete in that we don't have details on the sector-specific views on the full language. We will fill these gaps.
- We have raised a couple of arguments to support our design of the bidding language and auction procedure. We expect that the auction is untractable for analysis of fully optimized strategic behavior of bidders. Nevertheless we need more evidence for a superior performance of the auction. We hope to provide this evidence by simulations with bidding automats in a simplified economical environment, supplemented with real experiments where people can play some of the bidders.
- So far, the optimization module is not powerful enough to deal with the complete national railway network. Mathematical algorithms will be further developed to enhance the solving capabilities for the optimal allocation problem.
- An important aspect of the network description task is to ensure real implementability of the scheduled trains. For practical purposes this means that a timetable generated on the macroscopic network model can always be transferred to a microscopic network description. We are working on this step. It requires not only a suitable data transfer and adaptation, but also adjustment of buffer times and possibly a refinement of the network description in order to guarantee the implementation of the timetable on a microscopic level and, in a next step, the operational stability in case of unforeseen disturbances. network model can always be transferred to a microscopic network description. We are working on this step. It requires not only a suitable data transfer and adaptation, but also adjustment of buffer times (and possibly refinement of network description) in order to guarantee microscopic implementation and, in a next step, network stability with respect to unforeseen interruptions.

References

Affuso (2003). Auctions of rail capacity? *Utilities Policy 11, 43–46.*

Affuso & Newbery (2004). Case Study: The Provision of Rail Services. In M. Janssen (Ed.), *Auctioning Public Assets : Analysis and Alternatives.* Cambridge University Press.

Bassanini, La Bella & Nastasi (2002). Allocation of Railroad Capacity under Competition: a Game Theoretic Approach to Track Time Pricing. In *Transportation and Network Analysis* pp. 1–17. Kluwer.

Brewer & Plott (1996). A Binary Conflict Ascending Price (BICAP) Mechanism for the Decentralized Allocation of the Right to Use Railroad Tracks. *International Journal of Industrial Organization 14(6), 857–886.*

Brnnlund, Lindberg, Nou & Nilsson (1998). Railway Timetabling using Langangian Relaxation. *Transportation Science 32(4), 358–369.*

Brnger (1996). "Konzeption einer Rechnerunterstttzung fr die Feinkonstruktion von Eisenbahnfahrplnen". Dissertation, RWTH Aachen.

Caplice & Sheffi (2003). Optimization-based Procurement for Transportation Services. *Journal of Business Logistics 24(2), 109ff.*

Caprara, Fischetti, Guida, Monaci, Sacco & Toth (2001). Solution of Real-World Train Timetabling Problems. In *HICSS 34*. IEEE Computer Society Press.

Caprara, Fischetti & Toth (2002). Modeling and Solving the Train Timetabling Problem. *Operations Research 50(5), 851–861.*

CPLEX (2005). *User-Manual CPLEX 9.1.* ILOG CPLEX Division.

Garey & Johnson (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman and Company, New York.

Hauptmann (2000). "Automatische und diskriminierungsfreie Ermittlung von Fahrplantrassen in beliebig groen Netzen spurgefhrter Verkehrssysteme". Dissertation, TU Hannover.

Hrlimann (2001). "Objektorientierte Modellierung von Infrastrukturelementen und Betriebsvorgngen im Eisenbahnwesen". Dissertation, ETH Zrich.

Isacsson & Nilsson (2003). An Experimental Comparison of Track Allocation Mechanisms in the Railway Industry. *Journal of Transport Economics and Policy 37(3), 353–382.*

Koch (2001). ZIMPL-User-Guide.

Ledyard, Olson, Porter, Swanson & Torma (2002). The First Use of a Combined-Value Auction for Transportation Services. *Interfaces 32(5), 4–12.*

Milgrom (2004). *Putting Auction Theory to Work.* Cambridge University Press.

Nilsson (2002). Towards a Welfare Enhancing Process to Manage Railway Infrastructure Access. *Transportation Research 36(5), 419–436*.

Pachl (2002). *Systemtechnik des Schienenverkehrs* (3 ed.). Teubner Verlag, Stuttgart/Leipzig/Wiesbaden.

Pachl (2003). Auswirkungen der Bahnreform auf Infrastruktur und Bahnbetrieb. In *Die Bahnreform – eine kritische Sichtung*. Ritzau KG.

Parkes (1999). *i*Bundle: An Efficient Ascending Price Bundle Auction. In *Proc. 1st ACM Conf. on Electronic Commerce (EC-99)*, pp. 148–157.

Parkes & Ungar (2001). An Auction-Based Method for Decentralized Train Scheduling. In *Proc. 5th International Conference on Autonomous Agents (AGENTS-01)*, pp. 43–50.

Reuter (2005). "Kombinatorische Auktionen und ihre Anwendungen im Schienenverkehr". Diplomarbeit, Technische Universitt Berlin.