A Service Design Problem for a Railway Network

Alberto Caprara Enrico Malaguti Paolo Toth

Dipartimento di Elettronica, Informatica e Sistemistica, University of Bologna
Viale Risorgimento, 2 - 40136 - Bologna (Italy)

Abstract

We study the problem of designing a set of highly-profitable freight routes in a railway network, taking into account the capacity of the network and the level of service requested by different goods; in particular the profit achieved by transporting a good is a nonlinear function of the associated transit time. We propose an ILP model which is solved heuristically by column generation and fixing techniques. Computational results on a real network crossing eleven European countries are reported, showing that we can find solutions that are provably close to optimal. Given the large size of our instances, a key issue of our approach is to avoid finding an optimal solution of the continuous relaxation of our model, stopping as soon as near-optimal primal and dual solutions are available.

Keywords: *Routing, Service Design, Freight Transportation, Railway Applications.*

1 Introduction

In this paper we consider the problem of designing a freight transportation service on a rail network, addressed within the REORIENT project, a European Union research project whose tasks included the definition and study of a trans-European freight corridor in Center-East Europe. The corridor, connecting the Polish ports on the Baltic sea to the ports of Constanza on the Black sea and Thessaloniki on the Aegean sea, and crossing eleven center-eastern European countries, represents an alternative freight route to the classical one through Germany.

The problem that we study asks for the design of paths on this railway corridor, where one of the two main directions North-South and South-North is considered, focusing on the flow of goods along this direction. In our problem, we are given the network topology, the capacity of the existing line segments, and the position of the terminals or shunting yards on the rail network, which are points where goods enter and exit the network. Moreover, we are given the service demand, i.e. an assessment of the number of railcars which might be moved from origin to destination terminals in the network. Railcars are not all identical, being filled with goods that may have different values, requested transit times, etc.

Our objective is the design of the best paths, that will correspond to freight train services, on this network. A path on the network is a set of stations visited, along with a set of railcars collected and delivered along the path. Not all the railcars have the same requirements in terms of transit time on the network, and they do not yield the same profit if transported (profit can represent a monetary value, or a preference that the service provider gives to a particular good category).

Problems arising in railway transportation are surveyed by Cordeau, Toth and Vigo in [4], where the authors review the most widespread models for routing, fleet management, scheduling and assignment of rolling stock to services. Many authors dealt with specific problems of tactical level (which includes service design) in railway freight transportation. In [6] Kwon et al. use a network flow model to adapt an existing solution, consisting of a routing, train make-up and a target schedule, to meet load constraints. Holmberg and Hellstrand [5] use a Lagrangean heuristic and branch-and-bound to solve the Uncapacitated Network Design problem, which can model freight transportation. Recently, Campetella et al. [2] presented a model to design the set of origin and destination connections for a freight service, considering empty cars movement, handling cost and quality of service, but ignoring capacities, and proposed a tabu search solution approach. They report computational results on realistic instances from an Italian railway company. Ceselli et al. [3] present different models for planning the operations of a freight express service is Switzerland, operated as a hub and spoke system, and report computational results on realistic instances.

The paper is organized as follows. In Section 2 we formally define our problem and in Section 3 we present an *Integer Linear Programming* (ILP) model, having exponentially many variables. Section 4 discusses a heuristic solution approach based on this model, which avoids solving its continuous relaxation to optimality as this would be too time consuming on our instances. Computational experience on these large-size instances from the REORIENT project are discussed in Section 5, showing that we can find provably near-optimal solutions in all cases.

2 Formal Problem Definition

Formally, the problem input specifies the *railway network*, represented by an acyclic digraph $G = (V, A)$, where the node set V represents the set of stations and junctions in the network and the arc set A represents the set of railway line segments. As explained above, we are studying a main direction of the corridor in the network (North-South or South-North), which justifies the use of directed arcs (according to the corridor direction) and determines the acyclicity of the graph. Given that G is acyclic, we assume that the nodes in V are numbered according to a topological order, i.e. if $(i, j) \in A$ then $i < j$. Each arc $a = (i, j) \in A$ has a transit time $f_a = f_{(i,j)}$. Furthermore, there are a maximum weight W and a maximum length L for trains (seen as set of railcars, see below) travelling on the network.

In addition, the problem input specifies a set T of *railcar types* along with the number n_t of railcars of type $t \in T$ to be transported. Each railcar of type $t \in T$ has an origin $o_t \in V$, a destination $d_t \in V$ (with $o_t < d_t$), a weight w_t and a length l_t. Moreover, each railcar type $t \in T$ has value e_t and a maximum transit time D_t. The profit for transporting a railcar of type t from o_t to d_t, denoted by π_t, equals e_t when the railcar is transported within D_t. If this is not the case, the cost for the delay l_t with respect to D_t is modelled by a piecewise convex linear function (i.e. the marginal cost of the delay is increasing), namely:

$$\pi_t = e_t - \max\{(\mu_t^r l_t / D_t + \eta_t^r) e_t,\ r = 1, \ldots, R\}, \tag{1}$$

where R is the number of linear pieces, μ_t^r is the slope of the r-th piece and η_t^r the value of this piece for $l_t = 0$.

The problem aims at grouping railcars into *trains*, each represented by a pair (p, S), where p is a path in G and S is a vector of size $|T|$ specifying how many railcars of each type $t \in T$ are carried by the train, satisfying the following technical constraints. First of all, $S_t \leq n_t$ for $t \in T$, i.e. no more than the available railcars for each type can be carried. Moreover, the total weight and length of the railcars simultaneously carried along path p must not exceed W and L, respectively. Letting Q denote the set of the exponentially many trains, the objective of the problem is to select at most k trains so that the total profit of the railcars transported by these trains is maximized. The maximum number k of trains to be designed is given as an input parameter, since it represents a strategic decision which is taken a priori.

3 Modelling Approach

A natural ILP formulation of the problem is obtained by associating variables with trains, and then handling the exponentially-large number of these variables in the resulting *Master Problem* (MP) by generating columns via the solution of a *Slave Problem* (SP). This approach, which in most cases turns out to be the most successful in practice, here is strongly motivated by the complicated structure of the railcar profits.

The MP contains the integer variables $x_q = x_{(p,S)}$, representing the number of trains $q = (p, S) \in Q$ selected in the solution. Let $\pi_q = \pi_{(p,S)} := \sum_{t \in T} S_t \pi_t$ be the associated profit (recalling expression (1) for π_t). The MP is the following ILP:

$$\text{(MP)} \qquad \max \sum_{q \in Q} \pi_q x_q, \tag{2}$$

$$\sum_{q \in Q} x_q \leq k, \tag{3}$$

$$\sum_{(p,S) \in Q} S_t x_{(p,S)} \leq n_t, \qquad t \in T, \tag{4}$$

$$x_q \in \mathbb{Z}_+, \qquad q \in Q. \tag{5}$$

Constraint (3) imposes that no more than k trains are selected, while constraints (4) impose that, for every railcar type t, at most n_t railcars are transported.

The continuous relaxation C(MP) of the MP is solved by standard column generation techniques by considering the SP aimed at finding a train $q \in Q$ with positive reduced profit to be added to the MP. This can be done by looking for the feasible train (p, S) which maximizes the reduced profit $\sum_{t \in T} S_t(\pi_t - \tau_t^*)$ and checking if this value is greater than σ^*, where σ^* and τ_t^* are the current dual variables associated with the C(MP) constraints (3) and (4), respectively.

Accordingly, the SP can be formulated as the following (Mixed-)ILP. We introduce integer variables y_t, representing the number S_t of railcars of type t which are carried by the train; binary variables z_a, having value 1 if the arc $a \in A$ is included in the train path p and 0 otherwise; and binary variables u_i, having value 1 if node $i \in V$ is visited by path p and 0 otherwise. Moreover, as discussed later, we need additional variables to model the delay l_t of each railcar carried and the associated cost.

The first set of constraints that we consider are the classical path constraints, imposing that the path starts from an artificial origin node 0 and imposing that exactly one arc entering and exiting a visited node is part of the path: $\sum_{a \in \delta^+(0)} z_a = 1$, $\sum_{a \in \delta^-(i)} z_a = \sum_{a \in \delta^+(i)} z_a = u_i$, $i \in V$, where, as customary, $\delta^+(i)$ and $\delta^-(i)$ denote the set of arcs in A exiting and entering each node $i \in V$.

Visiting constraints are used to impose that, if the train is carrying railcars of type t, it must visit the origin node o_t and the destination node d_t of the railcars. Letting $m_t := \min\{n_t, W/w_t, L/l_t\}$ be an upper bound on the number of railcars of type t that can be transported, visiting constraints read: $y_t \leq m_t u_{o_t}$, $y_t \leq m_t u_{d_t}$, $t \in T$.

Capacity constraints impose that the total weight and length of the railcars simultaneously carried do not exceed the limits W and L, respectively. Rather than imposing one constraint for each node in V, it is sufficient to impose one constraint for each node which is the origin of a railcar type, given that the set of railcars carried simultaneously may change only in these nodes:

$$\sum_{s \in T : o_s \leq o_t, d_s > o_t} w_s y_s \leq W, \qquad t \in T, \tag{6}$$

$$\sum_{s \in T : o_s \leq o_t, d_s > o_t} l_s y_s \leq L, \qquad t \in T. \tag{7}$$

We complete the description of the SP model by discussing how to express the delay l_t of the railcars of type t loaded and the associated profit π_t in the objective function. Given that we are interested in a (Mixed-)ILP formulation, this is not possible if we do not introduce additional variables. Specifically, l_t could be set equal to the maximum between 0 and the difference between the travel time of the subpath of p from o_t to l_t and D_t:

$$l_t \geq 0, \qquad l_t \geq \sum_{(i,j) \in A : i \geq o_t, j \leq d_t} f_a z_a - D_t, \qquad t \in T,$$

and the associated cost in the piecewise linear function, say b_t, could be defined as the maximum over all costs of the linear pieces:

$$b_t \geq \mu_t^r l_t / D_t + \eta_t^r, \qquad t \in T,\ r = 1, \ldots, R.$$

On the other hand, the associated profit in the objective function would then by given by $(e_t - b_t)y_t$, which is quadratic in variables b_t and y_t.

The most natural solution to have a linear model would be to use, rather than the integer variables y, binary variables associated with each single railcar. This is what we did in our original approach, and led to very large and symmetric models whose solution was extremely time consuming (in our instances we have hundreds of railcars of the same type). An alternative that limits the number of variables and the symmetry is to define the binary expansion of the integer variables y_t through binary variables ϕ_{it}:

$$y_t = \sum_{i=0}^{\lfloor \log_2 m_t \rfloor} 2^i \phi_{it}, \qquad t \in T. \tag{8}$$

In practice, ϕ_{it} is a binary variable taking value 1 if 2^i railcars of type t are loaded on the train. By introducing continuous non negative variables l_{it}, the delay of these 2^i railcars can be expressed as:

$$l_{it} \geq \sum_{(i,j) \in A : i \geq o_t, j \leq d_t} f_a z_a - D_t - M_t(1 - \phi_{it}), \qquad t \in T,\ i = 0, \ldots, \lfloor \log_2 m_t . \rfloor. \tag{9}$$

When 2^i railcars of type t are carried ($\phi_{it} = 1$), their (positive) delay is constrained by (9) to be at least equal to the travel time of the subpath of p from o_t to d_t minus D_t. When the 2^i railcars of type t are not carried ($\phi_{it} = 0$), (9) is satisfied by $l_{it} = 0$ if the parameter M_t is large enough. Given that the nodes of the network are topologically ordered, we can set $M_t := f_{\max}(o_t, d_t) - D_t$, where $f_{\max}(o_t, d_t)$ is the maximum travel time of a path from a node v to a node w, with $v \geq o_t$ and $w \leq d_t$, according to the topological order of the vertices.

In order to represent the cost of a delay l_{it} affecting 2^i railcars of type t, we use an additional continuous variable b_{it} for each railcar type t and for each $i = 0, \ldots, \lfloor \log_2 m_t \rfloor$, and we impose:

$$b_{it} \geq \mu_t^r l_{it} / D_t + \eta_t^r, \qquad t \in T,\ i = 0, \ldots, \lfloor \log_2 m_t \rfloor,\ r = 1, \ldots, R \tag{10}$$

so that the SP objective function can be stated as:

$$\text{(SP)} \qquad \max \sum_{t \in T} (e_t - \tau_t^*) y_t - \sum_{t \in T} e_t \sum_{i=0}^{\lfloor \log_2 m_t \rfloor} 2^i \phi_{it} b_{it}. \tag{11}$$

4 Solution Approach

At each iteration of the column generation approach to solve C(MP), we first compute a few heuristic SP solutions by an algorithm called the *slave heuristic*. If no train with positive reduced profit is found, SP is then solved by a general-purpose Mixed-ILP solver, whose execution is stopped as soon as a train having positive reduced profit is found. Given the size of the instances we consider, it may happen that, when the column generation procedure is close to convergence, finding a train with positive reduced profit is fairly difficult, and the Mixed-ILP solver gets into memory and numerical troubles. To prevent this, it is essential to define a *tolerance* threshold, and stop the algorithm as soon as C(MP) is solved with this tolerance, as explained in Section 4. Heuristic solutions for MP, i.e. for our overall problem, are computed by a classical *fixing heuristic* based on the C(MP) solution, described in Section 4.

Approximate C(MP) solution

We next discuss how to compute, at each iteration of the column generation procedure, an upper and a lower bound on the optimal C(MP) value, denoted by $z(\text{C(MP)})$, so as to be able to stop the procedure when the associated relative gap is below the desired tolerance. A lower bound is simply given by the optimal value of the current LP, stated here in terms of the dual objective function:

$$z(\text{C(MP)}) \geq \sum_{t \in T} n_t \tau_t^* + k\sigma^*.$$

As to the upper bound, consider the Lagrangian relaxation of constraints (4) in MP, by using the Lagrangian multiplier vector τ^*:

$$(\text{L(MP},\tau^*)) \qquad \max \sum_{q \in Q} \pi_q x_q + \sum_{t \in T} \tau_t^* (n_t - \sum_{(p,S) \in Q} S(t) x_{(p,S)}), \tag{12}$$

$$\sum_{q \in Q} x_q \leq k, \tag{13}$$

$$x_q \in \mathbb{Z}^+, \qquad q \in Q. \tag{14}$$

Since objective function (12) maximizes the reduced profit of the selected trains with the addition of the constant term $\sum_{t \in T} \tau_t^* n_t$, the optimal solution of L(MP,τ^*) is obtained by selecting k times the train with most positive reduced profit (and by selecting no train if there is no train with positive reduced profit). In other words, if there are trains with positive reduced profit, the optimal Lagrangian value is $\sum_{t \in T} n_t \tau_t^* + kz(\text{SP})$, where $z(\text{SP})$ is the optimal value of the current SP. Noting that this Lagrangian relaxation has the integrality property, i.e. the removal of the integrality constraints does not affect its value, we have our upper bound:

$$z(\text{C(MP)}) \leq \sum_{t \in T} n_t \tau_t^* + kz(\text{SP}),$$

recalling that $z(\text{SP}) \geq \sigma^*$ if and only if there are trains with positive reduced profits.

Fixing

The algorithm described in the previous Section (approximately) solves the continuous relaxation C(MP) of the problem MP. However, we are interested in obtaining an integer solution to the problem, within an acceptable computing time. Thus, we use column fixing techniques in order to obtain a good feasible solution for the MP. The basic idea works as follows: we solve the C(MP) to near-optimality as illustrated in the previous section, and fix all variables having an integer value to this value. In addition, we also fix the least fractional variable (i.e. the fractional variable whose value is closest to an integer larger than 0) to the closest integer value. We remove all railcars that are transported by the fixed trains (variables), update the number of available trains, and iterate on the reduced problem (solving its continuous relaxation to near-optimality and fixing) until a feasible integer solution is obtained.

5 Computational Experiments

We report the results of our method on the railway network considered in the REORIENT project, which covers 11 European countries and counts 262 nodes and 523 arcs after having been preprocessed in order to remove the degree-2 nodes. We consider seven different service demands for a period of one week, which represent the potential demand for the corridor under alternative scenarios, and were assessed from the European Transport policy Information System (ETIS) trade data (see [1] for details on the network and demand data). For each demand scenario, we consider the two directions of the corridor, North-South (N-S) and South-North (S-N), and for each direction we set the number of available trains

Table 1: Computational results on the REORIENT corridor.

scen	*dir*	$\|T\|$	*cars*	*term*	*k*	%*tol*	*profit*	%*gap*	*cols*	%*heur*	*time*	%*MILP*
1	N-S	129	4054	32	7	3	576740	3.97	1772	99	440	11
1	S-N	238	13821	38	7	5	921990	5.20	3808	26	3808	96
2	N-S	137	3247	36	7	3	685498	3.11	1183	93	485	56
2	S-N	242	15135	40	7	3	777036	3.44	473	12	2801	98
3	N-S	127	3405	33	7	3	562963	3.25	1616	96	461	35
3	S-N	229	13335	41	7	3	824837	3.40	1700	66	5418	90
4	N-S	134	3520	35	7	3	556510	3.96	2944	95	1182	42
4	S-N	237	14488	40	7	5	860522	5.83	1138	50	6969	95
5	N-S	144	4484	35	7	3	580682	3.68	4004	97	1277	19
5	S-N	202	14117	38	7	3	837610	3.15	613	81	391	77
6	N-S	67	2897	21	7	3	226363	5.79	551	13	1028	99
6	S-N	112	11990	31	7	3	516360	3.13	609	76	364	90
7	N-S	133	3373	35	7	3	598010	3.87	1847	97	589	37
7	S-N	240	14978	41	7	3	783523	3.87	661	21	4523	97
1	N-S	129	4054	32	14	3	818558	3.22	1608	97	526	27
1	S-N	238	13821	38	14	5	1626810	5.69	1668	23	18801	97
2	N-S	137	3247	36	14	3	969312	2.87	3690	98	871	39
2	S-N	242	15135	40	14	3	1388610	3.31	902	42	10081	98
3	N-S	127	3405	33	14	3	767037	3.56	3114	99	562	14
3	S-N	229	13335	41	14	3	1419200	3.58	2730	71	8575	90
4	N-S	134	3520	35	14	3	785937	2.83	1085	96	300	24
4	S-N	237	14488	40	14	5	1457520	5.36	2454	81	7059	90
5	N-S	144	4484	35	14	3	812257	3.23	1011	99	273	23
5	S-N	202	14117	38	14	3	1437650	3.80	1275	65	3990	92
6	N-S	67	2897	21	14	3	320713	5.11	521	62	312	96
6	S-N	112	11990	31	14	3	883525	3.53	611	77	373	90
7	N-S	133	3373	35	14	3	812554	3.05	1409	99	255	21
7	S-N	240	14978	41	14	5	1453500	5.70	830	28	5729	97

to $k = 7$ (i.e., one train per day in each direction) and $k = 14$ (i.e., two trains per day in each direction), thus obtaining 28 instances. Table 1 reports, for each scenario (*scen*) and for each considered direction of the corridor (*dir*), the number of different railcars types, corresponding to the cardinality of set T, the total number of railcars (*cars*), and the number of nodes in the network being origin or destination terminals of some railcars (*term*).

We implemented our method in C and ran it on a PC Pentium 4, solving the LPs by CPLEX version 10. Within column generation, when the slave heuristic does not succeed in finding a train with positive reduced cost, the solution of the Mixed-ILPs is again by CPLEX. The percentage tolerance in the solution of C(MP) is 3%, namely we terminate the column generation process (and proceed with fixing) when the difference between the upper and lower bound on z(C(MP)) (of the previous section) relative to the upper bound is at most 3%. For five instances, the method could not converge, namely the CPLEX Mixed-ILP solver exceeded our time limit, and we had to increase the percentage tolerance to 5%.

In Table 1, we report: the number k of available trains; the percentage tolerance used (%*tol*); the total profit of the final integer solution (*profit*), corresponding to the k trains selected; the percentage relative gap between this profit and the upper bound on z(C(MP)) (%*gap*); the total number of columns generated (*cols*); the percentage of columns generated by the slave heuristic (%*heur*); the total computing time in seconds (*time*); and the percentage of this time spent by CPLEX for solving the Mixed-ILPs to generate columns (%*MILP*).

The table shows that we can find solutions which are close to optimal in all cases, within a running time ranging from a few minutes to five hours. In fact, the percentage gap tends to be close to the percentage tolerance, and actually in our experiments we observed that, by reducing the latter, the former is reduced as well too (when the instance can be solved). In other words, there seems to be evidence that the solutions that we compute are not only provably near-optimal, but also notably closer to the optimum than what we can prove. (The only noticeable exception is scenario 6 N-S, where the gap exceeds 5% whereas the tolerance is 3%.) Not surprisingly, the five most difficult instances are those having the largest number of different railcar types $|T|$.

References

[1] V.C. Arcot, A. Caprara, C. D'Ambrosio, J. Dong, A. Kozuki, A. Kuo, C.C. Lu, H.S. Mahmassani, E. Malaguti, S. Martello, E. Miller-Hooks, R. Nair, P. Toth, and K. Zhang. Modelling corridor freight transport for demonstration of seamless international rail freight services. Technical report, University of Maryland, MD, Usa, 2007.

[2] M. Campetella, G. Lulli, U. Pietropaoli, and N. Ricciardi. Freight service design for the italian railways company. *ATMOS 2006 - 6th Workshop on Algorithmic Methods and Models for Optimization of Railways.*

[3] A. Ceselli, M. Gatto, M. Lübbecke, M. Nunkesser, and H. Schilling. Optimizing the cargo express service of swiss federal railways. *Transportation Science (forthcoming)*, 2008.

[4] J. Cordeau, P. Toth, and D. Vigo. A survey of optimization models for train routing and scheduling. *Transportation Science*, 32:308–404, 1998.

[5] K. Holmberg and J. Hellstrand. Solving the uncapacitated network design problem by a lagrangean heuristic and branch-and-bound. *Operations Research*, 46:247–259, 1998.

[6] O. Kwon, C. Martland, and J. Sussman. Routing and scheduling temporal and heterogeneous freight car traffic on rail networks. *Transportations Research - E*, 34:101–115, 1998.