ELSEVIER

European Journal of Operational Research 96 (1996) 54–63

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Theory and Methodology

# Optimal lines for railway systems [1]

Michael R. Bussieck [a], Peter Kreuzer [b], Uwe T. Zimmermann [a,*]

[a] *TU Braunschweig, Abteilung für Mathematische Optimierung, Pockelsstraße 14, D-38106 Braunschweig, Germany*

[b] *IVV Ingenieursgeseelschaft für Verkehrsplanung und Verkehrssicherung GmbH, Braunschweig, Germany*

Received 1 January 1995; revised 1 November 1995

**Abstract**

We discuss the optimal choice of traffic lines with periodic timetables on a railway system. A chosen line system has to offer sufficient capacity in order to serve the known amount of traffic on the system. The *line optimization problem* aims at the construction of a feasible line system optimizing certain objectives. We introduce a mixed integer linear programming formulation. For real world data we succeed in solving the model by means of suitable relaxations and sufficiently strong cutting planes with the commercial LP solver CPLEX 3.0.

*Keywords:* Integer programming; Railway networks; Periodic timetable; Line optimization; Cutting planes

## 1. Introduction

Nowadays planning problems of railway systems become more manageable due to efficient algorithms and better implementations on faster computers. Especially solving huge linear programs, which is a substantial part of solving mixed integer problems, became much more efficient in the last ten years. Nevertheless a lot of mathematical work has to be done to solve "real-world" instances of a complex problem.

In this paper we describe a problem which occurs in a railway system with periodic timetables. Nearly every urban public transportation system (tramway, bus) and a growing number of railway companies (e.g. Nederlandse Spoorwegen) use periodic timetables. In a railway system with periodic timetable a junction or line connecting two stations runs several times, in a fixed time interval (e.g. one hour), across the network. This number is called the frequency of the line. The problem considered in this paper consists of choosing some lines with their frequencies to serve passenger demand and to optimize a given objective. Several different objective functions are proposed. On one hand you may try to minimize operational costs for a fixed service [4], on the other hand you may wish to maximize service quality for fixed operational costs.

One way to improve the service is to minimize the *total travel time* of all passengers. At this stage of planning there is no timetable, hence you cannot determine the exact waiting period while changing lines. Changing of lines itself is a major inconvenience, hence one possible way to optimize service is to minimize the total number of changes, or even simpler to maximize the total number of travellers on direct connections (or simply *direct travellers*).

Public transportation companies offer several services to meet the requirements of their customers. Typ-

* Corresponding author.

[1] This research was supported by the Federal Ministry for Education, Science, Research and Technology.



*SSDI* 0377-2217(95)00367-3

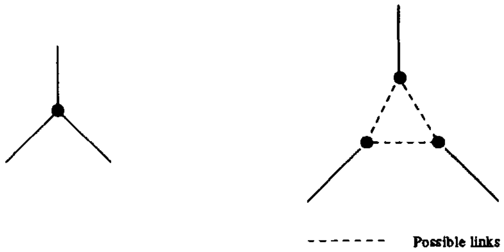


Fig. 1. Network transformation for applying usual network design techniques.

ically, the railway companies set up fast far-reaching InterCity(Express) trains (IC/ICE), InterRegio trains (IR) connecting district towns and commuter trains (CT). Travellers will be assigned to the different networks by a procedure called *system split* [13]. The idea of this split is very simple. Assume there are some passengers at a small station $a$ who want to travel to another far away small station $b$. No fast train (IC/ICE or IR) stops at these stations, hence there is only a slight hope for a direct connecting train, and if it exists, it will be very slow. Therefore we assume that some travellers take a CT train to the next station $c$, where an IC/ICE or IR train stops, use this fast train to reach a station $d$ near station $b$ and finally get on a CT train to station $b$. Hence we split journeys from $a$ to $b$ in the following way: In the network for CT trains we move passengers between $a$ and $c$, just as between $d$ and $b$. In the IC/ICE respectively IR network we move passengers between $c$ and $d$. The exact split depends on the assumption on the behaviour of the passengers and the topology of the network.

After this procedure we obtain mostly three different networks (IC, IR, and CT) with their specific data. The data for each network consists of a set of stations, the direct connections between two stations (links), the travel time and distance for these links, and a given amount of traffic between each pair of stations. The problem of finding optimal lines, in short *line optimization*, can now be performed independently on the different subnetworks like other phases of tactical railway planning [2].

In the context of network design [9, 10] the problem can be formulated as an *optimum network design for minimum cost multicommodity flows*. The set of *possible links* consists of the connections of tracks inside a station (Fig. 1). If some travellers find a suitable travel path with all tracks connected by these inner links, then these travellers have a direct connection between their origin and destination. Due to a small number of suitable travel paths we prefer another formulation of the line optimization problem. We derive a mixed integer linear program (MIP) related to models proposed in [14] (aircraft) and [1] (railroad freight transportation). Dienst et al. [6, 8] consider the problem for passenger transport and introduce basic terminology. They propose a branch-and-bound algorithm for solving the line optimization problem. In the next section we introduce our model and compare it with Dienst's approach. First experiments are reported in Section 3. The results of the experiments lead to some changes which are discussed in Section 4. In Section 5 we take advantage of the integrality of our problem to introduce some valid inequalities which help to solve the MIP and offer some concluding remarks in Section 6.

## 2. Modelling railway networks and lines

Let us first introduce the basic elements of our problem. For reasons of symmetry (we assume that passengers from $a$ to $b$ come back to $a$) we model the railway network using an undirected graph $G = (V, E)$, where $V$ denotes the set of vertices which describe the stations. $E$ is the set of edges which define direct connections or links between two stations. Furthermore we know some evaluation of the edges, like $T : E \to \mathbb{Z}_+$, the travel time on a single link, or $D : E \to \mathbb{Z}_+$, the travel distance. Possible lines in a railway network are modeled by (simple) paths in $G$. A station in which a line may start/end must have a special equipment (e.g. sidings to compose trains). Let $\mathcal{CY} \subseteq V$ describe these *classification yards*. Only paths in $G$ with start- and endpoint in $\mathcal{CY}$ are possible lines. Let $\mathcal{L}_0$ denote the set of all possible lines, then $f : \mathcal{L}_0 \to \mathbb{Z}_+$ denotes the frequencies of the possible lines in a fixed time interval (e.g. in one hour).

Next we model the behaviour of the travellers. Let $tr : \{\{a,b\} \mid a,b \in V, a \neq b\} \to \mathbb{Z}_+$ denote the volume of traffic between the stations. Let $\mathcal{T} := \{\{a,b\} \mid a,b \in V, a \neq b, tr(\{a,b\}) \neq 0\}$ denote the set of origin-destination-pairs with nonzero volume of traffic. Instead of $tr(\{a,b\})$ for $\{a,b\} = t \in \mathcal{T}$ we shortly write $tr(a,b)$ or $tr(t)$. Obviously this in-

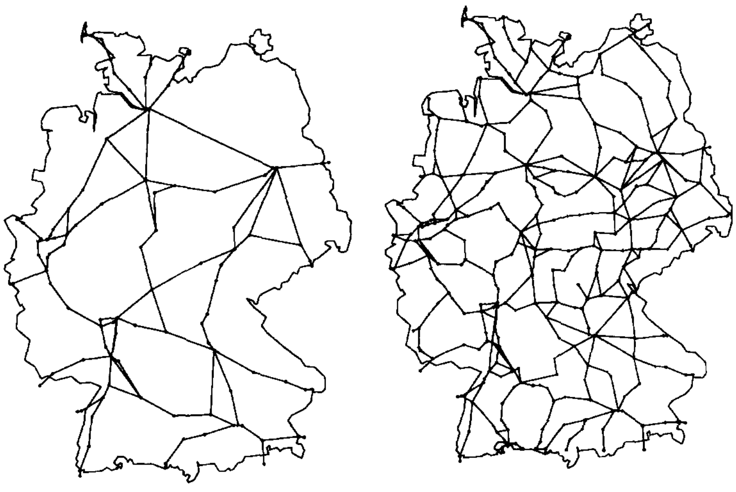

Fig. 2. The German IC/ICE and IR railway network.

formation is not enough to define the traffic flow on the network. Therefore, we have to make certain assumptions on the behaviour of the travellers.

**Assumption** Travellers between $a$ and $b$ $(a, b \in V)$ use a *shortest path* between $a$ and $b$ in $G$ with respect to some edge evaluation, i.e. w.r.t. travel time $T$ or w.r.t. travel distance $D$.

For most of the long-distance networks, this is a realistic assumption. For very dense local networks, like urban bus networks, this will not reflect reality. The assumption is sufficient to fix the traffic load through the links of the railway network when we assume that all shortest paths are uniquely determined. Let $P_t$ denote the shortest path in $G$ with respect to some edge evaluation between $a$ and $b$ $(t = \{a, b\} \in \mathcal{T})$. Then the traffic load $tl : E \longrightarrow \mathbb{Z}_+$ is given by

$$tl(e) := \sum_{\substack{\{a,b\}=t\in\mathcal{T} \\ e\in P_t}} tr(t).$$

If we assume a maximal fixed *train capacity* $C$, we may compute the minimum number of trains/lines, called *line-frequency-requirement*, which have to run along link $e$ to serve the demand for transportation. A reasonable calculation of the line-frequency-requirement $lfr : E \longrightarrow \mathbb{Z}_+$ would be

$$lfr(e) := \left\lceil \frac{tl(e)}{C} \right\rceil.$$

Due to political, economical and other non-mathematical considerations, this calculation is not always used, hence in our model we have to treat the line-frequency-requirement as a fixed input parameter.

Since every traveler between $t \in \mathcal{T}$ moves along his shortest path $P_t$, direct travel maximization suggests to choose shortest paths or combinations of shortest paths as possible lines. Hence we shrink $\mathcal{L}_0$ to $\mathcal{L} := \{l \in \mathcal{L}_0 \mid l \text{ is a shortest path between some } a, b \in \mathcal{CY}\}$. Although we can handle any combination of lines in the model as well, combinations are usually done "by hand" at the end of the optimization when further operational constraints have to be satisfied.

## 2.1. *The mixed integer linear programming formulation*

A feasible solution of the line optimization problem is a set of lines with their frequencies satisfying the line-frequency-requirement for every edge. An optimal solution maximizes the number of direct travellers. Let $d_{t,l} \in \mathbb{Z}_+$ denote the number of direct travellers between $t \in \mathcal{T}$ $(t = \{a, b\})$ using line $l$. We remind that $f_l$ denotes the frequency of some line $l \in \mathcal{L}_0$. Then, we have the following MIP formulation of the line optimization problem:

$$D^* = \max \sum_{l\in\mathcal{L}} \sum_{\substack{t\in\mathcal{T} \\ P_t\subseteq l}} d_{t,l},$$

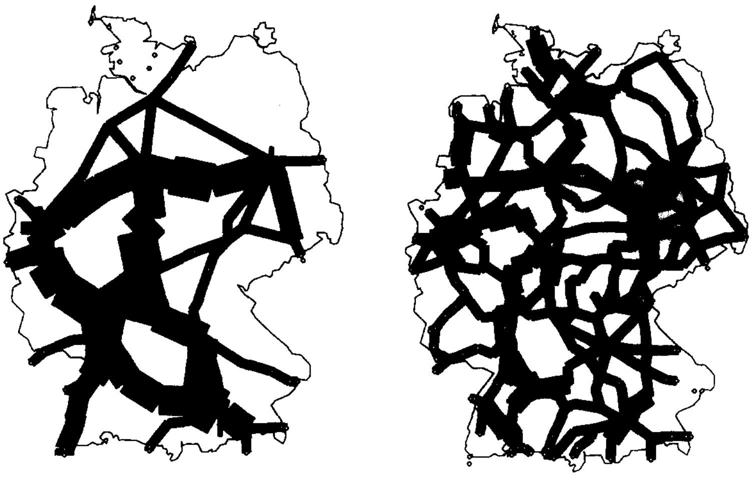

Fig. 3. The line-frequency-requirement for the German IC/ICE and IR railway network.

$$s.t. \quad \sum_{\substack{l\in\mathcal{L}\\ l\supseteq P_t}} d_{t,l} \le tr(t) \quad (\text{for all } t\in\mathcal{T}), \qquad (1)$$

$$\sum_{\substack{t\in\mathcal{T}\\ e\in P_t\subseteq l}} d_{t,l} \le C\cdot f_l \quad (\text{for all } e\in E, l\in\mathcal{L}), \qquad (2)$$

$$\sum_{\substack{l\in\mathcal{L}\\ e\in l}} f_l = lfr(e) \quad (\text{for all } e\in E), \qquad (3)$$

$$d_{t,l}, f_l \in \mathbb{Z}_+ \quad (\text{for all } t\in\mathcal{T}, l\in\mathcal{L}).$$

We will allow fractional travellers, i.e. we relax $d_{t,l} \in \mathbb{Z}_+$ to $d_{t,l} \geq 0$, for several non mathematical reasons. If we choose a feasible set of lines with certain frequencies then the remaining smaller maximization problem describes the quality of our choice. Now, the train capacity $C$ is only a vague estimation of the real situation. Moreover, the number of direct travellers is huge. Therefore, it seems not to be very important to find the exact integral optimum of this subproblem just for the comparison of the quality of our choice. It may be sufficient to base our evaluation on its linear programming relaxation. We refer to the above MIP formulation with this relaxation as **LOP**.

Inequality (1) restricts the number of direct travellers between $t \in \mathcal{T}$ by the total number of travellers between $t$. By inequality (2) no line can be overloaded and Eq. (3) ensures that the edges are covered with a sufficient number of lines/frequencies. $f$ defines a set $L_f := \{l \in \mathcal{L}_0 \mid f_l \neq 0\}$ with its frequencies. Therefore $f$ is called a *line partition* of $G$ if it fulfills (3). To improve the flexibility of the model you may attach some weights $w(t,l)$, e.g. travel distances or travel times, to $d_{t,l}$ in the objective function.

## 2.2. Heuristical approach

The main difference between our model **LOP** and the model described in [6,8] is that Dienst et al. assume an infinite train capacity. Whenever a direct connection exists for some travellers, they will be able to use this line neglecting the actual load. Setting $C := \sum_{t\in\mathcal{T}} tr(t)$, our model includes this approach, but you can take advantage of the infinite train capacity and find a more efficient model. In Section 4 we come back to this question. In this paragraph we give an outline of the method used by Dienst et al. [6,8].

His algorithm is based on a simple branch-and-bound (B&B) method which tries to build a line partition by adding lines one after another. Since $C = \infty$, the value of the line packing $L$, i.e. a set of lines with their frequencies which fulfills the partition equality (3) with "$\leq$", is $\sum_{l\in\mathcal{L}} \sum_{t\in\mathcal{T}, P_t\subseteq l} tr(t)$. After adding a line $l$ to the line packing we adjust the remaining data. In a node of the B&B-tree with a feasible line packing you branch on the choice of a line $l$ with maximal current direct travellers ($\sum_{t\in\mathcal{T}, P_t\subseteq l} tr(t)$) (Greedy). The remaining parts of the algorithm are standard B&B techniques. Due to the very slow performance of this method the algorithm is (usually) interrupted after a fixed time

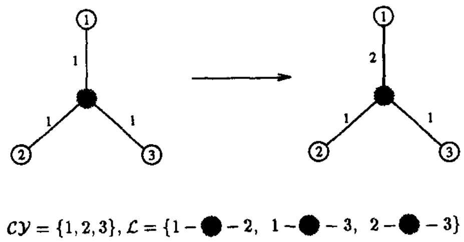

Fig. 4. An instance $(G,lfr,\mathcal{L})$ with no valid line partition and one of its adjustments.

(e.g. 100000 sec.) or a fixed number of operated nodes (e.g. 10000). Two further well-known problematic features of the algorithm are listed in the following.

- The current best (or the final optimal) line packing $\hat{L}$ may be infeasible. If we do not succeed in completing it by the remaining lines then there is no information at all whether the given data allow any valid line partition. Either there exists no line partition due to faulty problem data, or the algorithm missed to find one, or we missed to find some completion. All three cases are possible. Of course, if no line partition exists for the instance $(G, lfr, \mathcal{L})$, then we have to adjust the line-frequency-requirement (Fig. 4).
- In case of external interruption no information about the quality of the current best line packing $\hat{L}$ relative to the optimum line packing $L^*$ is known.

Nevertheless, in case of consistent data the algorithm seems to work quite well in practice, if we provide sufficient computer time (Table 1, problem instances described in the next section). For the instances tested, the gap between solutions generated by the B&B method and the optimal solution were always $\leq 4.1\%$. This and all other computational experiments were achieved on a HP 9000/715-50 workstation.

## 3. The problem instances and first results

Solving NP-hard problems like the line optimization problem (polynomial reducible to EXACT COVER BY 3-SETS [7]) has to be based on the actual structure of real-world data. Our current data pool consists of five "real-world" railway networks. Three of them (NS-IC, NS-IR and NS-CT) come from the Dutch railway company (Nederlandse Spoorwegen) and the remaining two (DB-IC and DB-IR) are from the German railway company (Deutsche Bahn AG). The parameters of the networks and the size of the concerning MIP formulation can be found in Table 2.

At the time being, a direct commercial solver of MIP's for networks of this size seems not to be available. Though using the fast CPLEX 3.0 LP solver [5], even the LP relaxation (replace $f_l \in \mathbb{Z}_+$ by $f_l \geq 0$) of the larger instances could **not** be solved on an HP 9000/715-50 with 212 MB core memory. Bixby [3] solved the LP relaxation for all instances with CPLEX on a SGI-Power-Challenge within 30 hours. Only for the smallest network (NS-IC), CPLEX 3.0 MIP solver managed to solve the MIP problem (Table 3).

Even if the solution of the LP-relaxation is found in reasonable time, the solution of the MIP remains difficult. Due to the large number of fractional variables $f_l$ in the optimal solution of the LP-relaxation, the successive B&B procedure of the MIP solver fails to find the optimal MIP solution. Therefore, we tried to ease the computational task in two ways:

- using a simpler model decreasing the size of the resulting MIP formulation,
- using integrality of variables to develop constraints which eliminate the generated fractional values of the variables $f_l$.

## 4. Reducing the size of the model

In the original model **LOP** travellers between $t \in \mathcal{T}$ using different lines $l_1$ and $l_2$ are counted in different variables $d_{t,l_1}$ and $d_{t,l_2}$. If the line-frequency-requirement is designed to carry the complete flow of travellers then we may try to aggregate the direct travellers in different lines. Let $D_t := \sum_{l \in \mathcal{L}, P_t \subseteq l} d_{t,l}$ denote the sum of all direct travellers between $t \in \mathcal{T}$ on all usable lines. In this smaller model, we do not take care of the exact distribution of the travels on the lines. The number of direct travellers $D_t$ is bounded by the total number of travellers (4) and capacity of lines connecting $t$ (5). In the context of multicommodity flows, this is a relaxation of the bundle constraints to individual capacities. However, we may run into difficulties if the relaxed model carries all travellers whereas in the original model the line-frequency-requirement is too small for the traffic volume.

Table 1
Results of the B&B method (interrupted after 10000 nodes)

| network | NS-IC | NS-IR | NS-CT |
|---|---|---|---|
| Optimality gap | 0.0% | 1.9% | 4.1% |
| Running time | 30 sec. | 1.5 h | 12 h |

Table 2
Net parameters and MIP formulation size

| network | $\|V\|$ | $\|\mathcal{CY}\|$ | $\|E\|$ | $\|\mathcal{L}\|$ | $\|\mathcal{T}\|$ | LOP | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Variables | Constraints | Nonzeros |
| NS-IC | 23 | 23 | 31 | 253 | 210 | 2120 | 1017 | 6278 |
| NS-IR | 86 | 86 | 114 | 3655 | 2147 | 116974 | 28487 | 556131 |
| NS-CT | 385 | 91 | 428 | 4095 | 11240 | 583598 | 96620 | 3589478 |
| DB-IC | 100 | 100 | 118 | 4950 | 3136 | 183235 | 43200 | 1038761 |
| DB-IR | 307 | 199 | 398 | 19701 | 9215 | 900173 | 261308 | 4853878 |

Table 3
Computational results with **LOP**.

| network | running time | | objective function value | | $\|\{f_l \mid f_l \notin \mathbb{Z}_+\}\|$ |
|---|---|---|---|---|---|
| | LP | MIP | LP | MIP | |
| NS-IC | 10.55 | 4912.32 | 9.168.554 | 8.203.412 | 141 |
| NS-IR | 5284.34 | - | 21.315.607 | - | 636 |
| NS-CT | - | - | *25.492.888 | - | - |
| DB-IC | 12777.66 | - | 9.768.973 | - | 1384 |
| DB-IR | - | - | *8.095.734 | - | - |

$^-$) no solution after 5h running time. Time in seconds. * computed in [3]

$$\hat{D} = \max \sum_{t \in \mathcal{T}} D_t,$$

$$\text{s.t.} \quad D_t \leq tr(t) \quad (\text{for all } t \in \mathcal{T}), \tag{4}$$

$$D_t \leq C \cdot \sum_{\substack{l \in \mathcal{L} \\ P_t \subseteq l}} f_l \quad (\text{for all } t \in \mathcal{T}), \tag{5}$$

$$\sum_{\substack{l \in \mathcal{L} \\ e \in l}} f_l = lfr(e) \quad (\text{for all } e \in E), \tag{6}$$

$$D_t, f_l \in \mathbb{Z}_+ \quad (\text{for all } t \in \mathcal{T}, l \in \mathcal{L}).$$

As in Section 3 we relax $D_t \in \mathbb{Z}_+$ to $D_t \geq 0$. The substantial smaller size of this model (referred as **lop**) is described in Table 4. The solution time for its LP relaxation for all five networks together decreases below 190 sec (Table 4).

Solving the MIP is still time consuming (3950 sec-for all instances). All MIP solutions were found using the CPLEX MIP solver based on a branch-and-bound algorithm. Before starting the B&B method CPLEX did a lot of preprocessing work called CPLEX MIP PRESOLVE. Without this preprocessing none of the instances could be solved (break after 5h running time). Any feasible solution of the original model yields a feasible solution of the smaller model. Therefore, the smaller model is a relaxation of the original problem. Solving the smaller model does not provide enough information to solve **LOP**. Due to the neglected capacities of single lines, it may be impossible to distribute all direct travellers $D_t$ in an optimal solution $D^*$ of the smaller model to the lines in the optimal line partition. Inequality (5) only assures that the travellers between one $t \in \mathcal{T}$ fit into the connecting lines. Still, the inequality is tighter than the one used in the model of Dienst (Section 3, [6,8]). Therefore, the model of Dienst is a relaxation of the **lop**.

Table 4
Computational results with **lop**

| network | size | | | running time | | objective function value | | $\lvert\{f_l \mid f_l \notin \mathbb{Z}_+\}\rvert$ |
|---|---|---|---|---|---|---|---|---|
| | Variables | Constraints | Nonzeros | LP | MIP | LP | MIP | |
| NS-IC | 463 | 234 | 3060 | 0.18 | 0.66 | 9.168.554 | 8.203.412 | 122 |
| NS-IR | 5802 | 2305 | 153752 | 11.07 | 187.50 | 28.909.033 | 21.172.411 | 502 |
| NS-CT | 15335 | 11763 | 696630 | 44.25 | 2128.47 | 38.922.436 | 37.118.270 | 390 |
| DB-IC | 8086 | 3276 | 236923 | 39.54 | 103.23 | 10.071.448 | 7.625.326 | 512 |
| DB-IR | 28915 | 9692 | 1239870 | 94.02 | 1527.92 | 8.106.707 | 6.114.448 | 1347 |

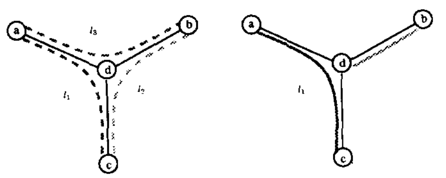


Fig. 5. Optimal fractional and integer solution. $tr(a,b) = tr(a,c) = tr(b,c) = 50$, train capacity $C = 100$. The optimal fractional solution is $f_{l_1} = f_{l_2} = f_{l_3} = 1/2$ with value 150. The optimal integer solution is e.g. $f_{l_1} = 1$, $f_{b-d} = 1$ with value 50.

## 5. Cutting planes

In the MIP formulation of **LOP** and **lop** we relaxed the integrality of $d_{t,l}$ respectively $D_t$. However, the integrality of the frequency variables $f_l$ is essential for a railway system with periodic timetable. The results reported in Section 4 point out that we should try to save computation time in the B&B part. Here, we propose some valid inequalities (*cutting planes*) which take advantage of the integrality of the problem. The method of using general cutting planes has been proposed by Chvátal and Gomory. More powerful for the MIP formulation of combinatorial problems (e.g. TSP, set partitioning, network design) is the use of problem specific cutting planes. Such cutting planes were successfully used to solve NP-hard problems [12].

### 5.1. *Cutting planes induced by (5)*

A close look at the solution of the LP relaxation of **lop** shows that the value of most frequency variables $f_l$ is

$$f_l = \frac{\sum_{t \in \tilde{\mathcal{T}}} tr(t)}{C},$$

for some $\tilde{\mathcal{T}} \subseteq \mathcal{T}$ (Fig. 5). Striving for integrality we add or change some corresponding inequalities.

**Lemma 1.** *Let $f$ be an integral line partition. For all $t \in \mathcal{T}$,*

$$D_t \le \left\lfloor \frac{tr(t)}{C} \right\rfloor (C - \Delta) + \Delta \sum_{\substack{l \in \mathcal{L} \\ P_t \subseteq l}} f_l, \qquad (7)$$

*with $\Delta := tr(t) - \lfloor \frac{tr(t)}{C} \rfloor C$, is valid for* **lop**.

**Proof.** We show that a solution $(D, f)$ of **lop** with integral line partition $f$ fulfills (7). Assume first that $\lfloor \frac{tr(t)}{C} \rfloor = \lceil \frac{tr(t)}{C} \rceil$, hence $\Delta = 0$ and (7) boils down to $D_t \le tr(t)$ which is equal to (4) of **lop**. Now assume $\lfloor \frac{tr(t)}{C} \rfloor \ne \lceil \frac{tr(t)}{C} \rceil$. Let $\sum_{\substack{l \in \mathcal{L} \\ P_t \subseteq l}} f_l =: i \in \mathbb{Z}_+$.

(1) $i \le \lfloor \frac{tr(t)}{C} \rfloor$: With (5) in **lop** we find

$$\begin{aligned} D_t &\le C \cdot i = C \cdot i - \Delta \cdot i + \Delta \cdot i \\ &\le \left\lfloor \frac{tr(t)}{C} \right\rfloor (C - \Delta) + \Delta \cdot i. \end{aligned}$$

(2) $i \ge \lceil \frac{tr(t)}{C} \rceil = \lfloor \frac{tr(t)}{C} \rfloor + 1$: With (4) in **lop** we find

$$\begin{aligned} &\left\lfloor \frac{tr(t)}{C} \right\rfloor (C - \Delta) + \Delta \cdot i \\ &= \left\lfloor \frac{tr(t)}{C} \right\rfloor C + \Delta \underbrace{\left(i - \left\lfloor \frac{tr(t)}{C} \right\rfloor\right)}_{\ge 1} \\ &\ge \left\lfloor \frac{tr(t)}{C} \right\rfloor C + \Delta \overset{\text{Def. of } \Delta}{=} tr(t) \ge D_t. \qquad \square \end{aligned}$$

For $t \in \mathcal{T}$ with $tr(t) \le C$ we substitute $D_t \le C \sum_{\substack{l \in \mathcal{L} \\ P_t \subseteq l}} f_l$ by

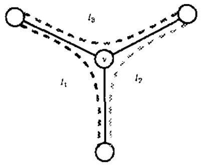

Fig. 6. A valid fractional line partition. $v \in \mathcal{CY}$, $lfr = 1$, $f(l_1) = f(l_2) = f(l_3) = \frac{1}{2}$.

$$D_t \leq tr(t) \sum_{\substack{l \in \mathcal{L} \\ r_t \subseteq l}} f_l,$$

which is obviously tighter than (5). For $t \in \mathcal{T}$ with $tr(t) > C$ we add (7) to **lop**.

### 5.2. Cutting planes induced by (6)

Another class of cutting planes is implied by the equation for a valid line partition (6). In Fig. 6 we give an example of a valid fractional line partition. Obviously, at least one line has to end in station $v$. The following lemma generalizes this observation.

**Lemma 2.** *Let $V' \subset V$, $E' \subseteq \{\{u,v\} \in E \mid |\{u,v\} \cap V'| = 1\}$ and $\sum_{e \in E'} lfr(e)$ be odd. Furthermore let $\alpha_l := |\{e \in l\} \cap E'|$. Then the following inequality holds for every valid (integer) line partition f.*

$$\sum_{\substack{l \in \mathcal{L} \\ \alpha_l \text{ even}}} \alpha_l f_l \leq \sum_{e \in E'} lfr(e) - 1. \qquad (8)$$

**Proof.** From (6) and the validity of $f$ we have

$$\sum_{l \in \mathcal{L}} \alpha_l f_l = \sum_{e \in E'} lfr(e),$$

$$\underbrace{\sum_{\substack{l \in \mathcal{L} \\ \alpha_l \text{ even}}} \alpha_l f_l}_{\in 2 \cdot \mathbb{Z}} \leq \sum_{e \in E'} lfr(e). \qquad (9)$$

The left hand side of (9) is even and the right hand side is odd, hence we may subtract 1 of the right hand side and keep validity.

$$\sum_{\substack{l \in \mathcal{L} \\ \alpha_l \text{ even}}} \alpha_l f_l \leq \sum_{e \in E'} lfr(e) - 1 \qquad \square$$

For the small example in Fig. 6 we may add

$$2f_{l_1} + 2f_{l_2} + 2f_{l_3} \leq 3 - 1 = 2$$

to prohibit the fractional line partition $f_{l_1} = f_{l_2} = f_{l_3} = 1/2$.

### 5.3. Computational results

A major difference between the classes of inequalities derived in Section 5.1 and 5.2 is their number. For $t \in \mathcal{T}$ we get at most one new inequality from the first class. However, the second class grows exponentially with the size of the network. Therefore, we add only the following subset of the inequalities to **lop**:

$$\sum_{\substack{l \in \mathcal{L} \\ l \text{ runs through } v}} 2f_l \leq \sum_{e=\{u,v\} \in E} lfr(e) - 1,$$

for $v \in V$ with $\sum_{e=\{u,v\} \in E} lfr(e)$ odd.

After adding both types of inequalities to **lop** the number of fractional frequency variables $f_l$ in the solution of the LP relaxation decreases substantially (Table 5). For NS-IC, NS-CT and DB-IC, all frequency variables are integral. For all problems, the difference between the solution time of the LP relaxation and the MIP becomes insignificant. All instances may be solved without using the CPLEX MIP PRESOLVER.

## 6. Generating upper and lower bounds

One drawback of the simple B&B algorithm (Section 3) is that there is no information on the quality of the "solution". During the B&B algorithm we get only lower bound on the optimal solution value. If the algorithm is interrupted, due to its time consuming behaviour, in most cases we obtain a solution but no information "how far" it is away from the optimal solution. In this section we derive upper and lower bounds in our model using **lop** and a projected version of **LOP**.

Since **lop** is a relaxation of **LOP** the value $\hat{D}$ of an optimal solution of **lop** is an upper bound for the value $D^*$ of an optimal solution of **LOP**. Trivially the maximum number of travellers through the network is an upper bound too. This number can be computed by solving

$$T^* = \max \sum_{t \in \mathcal{T}} D_t,$$

$$s.t. \quad D_t \leq tr(t) \quad (\text{for all } t \in \mathcal{T}),$$

Table 5
Computational results for **lop** tightened by cutting planes

| network | running time | | | objective function value | | $\|\{f_l \mid f_l \notin \mathbb{Z}_+\}\|$ |
|---|---|---|---|---|---|---|
| | LP | MIP | MIP* | MIP | LP | |
| NS-IC | 0.38 | 0.48 | 0.38 | 8.203.412 | 8.203.412 | 0 |
| NS-IR | 32.73 | 37.93 | 30.62 | 27.175.593,5 | 27.172.441 | 8 |
| NS-CT | 143.82 | 161.87 | 358.94 | 37.118.270 | 37.118.270 | 0 |
| DB-IC | 50.47 | 56.28 | 55.74 | 7.625.326 | 7.625.326 | 0 |
| DB-IR | 312.47 | 408.92 | 420.86 | 6.116.916,5 | 6.114.448 | 3 |

*) no preprocessing with CPLEX MIP PRESOLVER. Time in seconds.

$$\sum_{\substack{t \in \mathcal{T} \\ P_t \ni e}} D_t \leq C \cdot lfr(e) \quad (\text{for all } e \in E),$$

$$0 \leq D_t \quad (\text{for all } t \in \mathcal{T}).$$

On the other hand, we may take the optimal line partition $\hat{f}$ from some optimal solution of **lop** and insert it into **LOP**. Remind that $L_{\hat{f}} = \{l \in \mathcal{L} \mid \hat{f}_l \neq 0\}$ denotes the corresponding set of lines. Then we get a lower bound $D(\hat{f})$.

$$D(\hat{f}) = \max \sum_{l \in L_{\hat{f}}} \sum_{\substack{t \in \mathcal{T} \\ P_t \subseteq l}} d_{t,l},$$

$$s.t. \quad \sum_{\substack{l \in L_{\hat{f}} \\ l \supseteq P_t}} d_{t,l} \leq tr(t) \quad (\text{for all } t \in \mathcal{T}),$$

$$\sum_{\substack{t \in \mathcal{T} \\ e \in P_t \subseteq l}} d_{t,l} \leq C \cdot \hat{f}_l \quad (\text{for all } e \in E, l \in L_{\hat{f}}),$$

$$d_{t,l} \geq 0 \quad (\text{for all } t \in \mathcal{T}, l \in L_{\hat{f}}).$$

The huge size of the complete model **LOP** is due to the number of possible lines and the resulting number of variables $d_{t,l}$. However, for a fixed line partition the size of $L_{\hat{f}}$ is small (here always < 100) and results in LP models of small size (e.g. DB-IR: $|L_{\hat{f}}| = 89$, 1422 variables, 1774 constraints, 7246 nonzeros) solvable in a few seconds. Hence, we may compute a confidence interval (Table 6) containing the optimal value of the original large model:

The confidence intervals are small and acceptable. For three networks (NS-IC, DB-IC, DB-IR), the best upper bound is achieved by $\hat{D}$, but for NS-IR and NS-CT the trivial bound $T^*$ applies. This behaviour becomes clear if we test our assumption when relaxing **LOP** to **lop**. The line-frequency-requirement is too small to carry the complete traveller flow. Fig. 7 shows overcrowded arcs in the two networks. In Section 4 the poor performance of the upper bound $\hat{D}$ is explained. Anyway, the *solution* generated by **lop** leads to a satisfactorily small confidence interval.

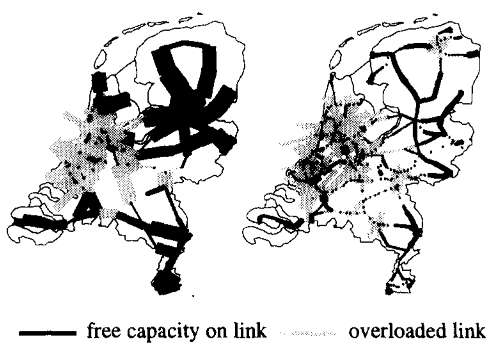


Fig. 7. Overcrowded links in NS-IR and NS-CT.

## 7. Conclusions

In this paper we derived a mixed integer linear programming formulation **LOP** for the line optimization problem. Due to its huge size we were forced to consider a smaller MIP model **lop** whose LP-relaxation could be solved using CPLEX 3.0. Adding suitable cutting planes we succeeded in solving the smaller MIP for all instances in less than **6 minutes**. A solution of **lop** leads to lower and upper bounds for **LOP**. For all instances this gap is less than **3.2%**. The use of more sophisticated MIP solvers like MINTO [11] is needless since the LP with the derived cutting planes is a good approximation for the MIP. Hence the B&B tree is very small and most of the computing time is spend by solving *one* LP. In our experiments we obtain the best results with CPLEX's **dual** simplex al-

Table 6
Lower and upper bounds on the optimal value $D^*$

| network | $D(\hat{f})$ | $\hat{D}$ | $T^*$ | gap between lower and best upper bound |
|---|---|---|---|---|
| NS-IC | 8.203.412 | 8.203.412 | 9.168.554 | 0.0% |
| NS-IR | 20.982.579 | 27.172.441 | 21.315.607 | 1.6% |
| NS-CT | 25.079.912 | 37.118.270 | 25.863.252 | 3.1% |
| DB-IC | 7.549.827 | 7.625.326 | 9.745.044 | 1.0% |
| DB-IR | 6.114.448 | 6.114.448 | 8.682.953 | 0.0% |

gorithm.

Future work is necessary to improve model and method. In **lop** as well as in **LOP** we admit only shortest path lines. The main reason for this choice in the past was the obvious inefficiency of former methods. The approach described is so efficient that we are quite sure that we may admit further possible, reasonable lines in the optimization. Similarly, one should reconsider the assumption that all travellers move along a shortest path. We may admit other paths, e.g. $k$-shortest paths for a small number $k$ or other reasonable paths. One can also consider the addition of some operational constraints, more flexible and complex linear objectives, and the parametric analysis of two different objectives.

In these strengthened models, it may be necessary to use all the inequalities derived in Section 5.2. Here we need some separation rules for the generation of violated inequalities "on demand" which can be used in a branch-and-cut framework. Another way to speed up the algorithm is to start the LP solver with a "good" initial solution, for which some efficient heuristic approach is missing by now.

Last but not least, we would like to thank the Ingenieurgesellschaft für Verkehrsplanung und Verkehrssicherheit (IVV) in Braunschweig for many helpful discussions and, in particular, for providing "real-world" data from ongoing consulting work.

## References

[1] Assad, Arjang A., "Modelling of rail networks: Toward a routing/makeup model", *Transportation Res. Part B* 14 (1980) 101-114.

[2] Assad, Arjang A., "Models for rail transportation", *Transportation Res. Part A* 14 (1980) 205-215.

[3] Bixby, R.E., private communication (1994).

[4] Claessens, M.T., and Van Dijk, N.M., "A mathematical programming model to determine a set of operation lines at minimal costs", Technical Report, University of Amsterdam (1994).

[5] CPLEX Optimization, Inc., "Using the CPLEX callable library", manual (1994).

[6] Dienst, Hartmut, "Linienplanung im spurgeführten Personenverkehr mit Hilfe eines heuristischen Verfahrens", Ph.D. Thesis, TU Braunschweig (1978).

[7] Garey, M.R., Johnson, D.S., *Computers and Intractability (A Guide to Theory of NP-completeness)*, Freeman, New York (1979).

[8] Pierick, K., and Weigand, K.D., "Methodical formulations for optimization of railway long-distance passenger transport", *Rail Int.* 7 (1976) 328-334.

[9] Magnanti, T.L., and Wong, R.T., "Network design and transportation planning: Models and algorithms", *Transportation Sci.* 18 (1984) 1-55.

[10] Minoux, M., "Network synthesis and optimum network design problems: Models, solution methods and applications", *Networks* 19 (1989) 313-360.

[11] Nemhauser, George L., Savelsbergh, Martin W.P., and Sigismondi, Gabriele C., "Minto, a mixed integer optimizer", *Operations Research Letters* 15 (1994) 47-58.

[12] Nemhauser, George L., Wolsey, Laurence A., *Integer and Combinatorial Optimization*, John Wiley & Sons, New York (1988).

[13] Oltrogge, Christine, "Linienplanung für mehrstufige Bedienungssysteme im öffentlichen Personenverkehr", Ph.D. Thesis, TU Braunschweig (1994).

[14] Richardson, R., "An optimization approach to routing aircrafts", *Transportation Sci.* 10 (1976) 52-71.