ELSEVIER European Journal of Operational Research 110 (1998) 474–489

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Theory and Methodology

Cost optimal allocation of rail passenger lines

M.T. Claessens [a], N.M. van Dijk [b], P.J. Zwaneveld [c,*]

[a] Netherlands Railways Travelers, P.O. Box 2025, NL-3500 HA Utrecht, The Netherlands
[b] University of Amsterdam, Roeterstraat 11, NL-1018 WB Amsterdam, The Netherlands
[c] Erasmus University Rotterdam and TNO INRO, P.O. Box 6041, NL-2600 JA Delft, The Netherlands

Received 15 October 1996; accepted date 25 April 1997

Abstract

We consider the problem of cost optimal railway line allocation for passenger trains for the Dutch railway system. At present, the allocation of passenger lines by Dutch Railways is based on maximizing the number of direct travelers. This paper develops an alternative approach that takes operating costs into account. A mathematical programming model is developed which minimizes the operating costs subject to service constraints and capacity requirements. The model optimizes on lines, line types, routes, frequencies and train lengths. First, the line allocation model is formulated as an integer nonlinear programming model. This model is transformed into an integer linear programming model with binary decision variables. An algorithm is presented which solves the problem to optimality. The algorithm is based upon constraint satisfaction and a Branch and Bound procedure. The algorithm is applied to a subnetwork of the Dutch railway system for which it shows a substantial cost reduction. Further application and extension seem promising.

Keywords: Branch and Bound; Integer programming; Optimization; Timetabling

1. Introduction

The Dutch railway network is one of the most intensively used railway networks in the world. The railway network has 2800 km of track and 372 stations. Almost 1,000,000 travelers and 60,000 tonnes of cargo are transported each day. The most important characteristic of the Dutch railway system is the use of a cyclical timetable. This means that the timetable is more or less identical during all hours of the day over a year.

In this paper we consider the problem of cost optimal railway line allocation for passenger trains. A line is a direct railway connection between two stations. A line is characterized by its origin and destination station, its frequency per hour, the route between these two stations and the intermediate stops at passing railway stations. The line system is the collection of all lines. The developed method for obtaining a cost optimal line system is designed to be used by Railned and Netherlands Railway Travelers within the process of generating timetables. For more details

* Corresponding author. TNO INRO, P.O. Box 6041, 2600 JA Delft, The Netherlands. Fax: +31-15 269 7782; e-mail: pzw@inro.tno.nl

PII S0377-2217(97)00271-3

about the overall process of generating timetables, we refer to Odijk et al. [1].

1.1. Motivation

Currently, Dutch Railways use a line allocation model to determine a line system which aims to maximize the total number of travelers on direct connections. This objective is also used in other countries (e.g. [2]). The underlying motivation of this approach is to minimize the inconvenience for a passenger of changing from one train to another. Hence, the objective is to transport as many travelers as possible directly from their origin to their destination station. This approach often results in *long routes* for the lines, where the notion of long refers to the (geographical) distance between its origin and its destination station. The direct travelers approach not only results in long routes, but also in *long trains*. The length of a train, determined by the number of cars, has to be adjusted to the number of passengers that the train has to transport at each segment (track) of its route. The train length and train capacity of a passenger line are roughly determined by the maximum number of passengers along the route of the train. This may result in long trains. Long lines often have considerable differences between the number of passengers on each track. Therefore, long lines may lead to a substantial amount of unused train capacity at less busy tracks and can thus be rather expensive. Conversely, a cost approach may increase the number of passenger transfers, but it may decrease operating costs. This decrease in operating costs could translate into lower prices for customers.

1.2. Objective and results

The objective of this study is to investigate the effects on line allocation when using a *cost* approach, as opposed to the current *direct travelers* approach. We specify our cost approach and compare this approach to the classical direct travelers approach. We present two new model formulations for the line allocation problem. We present an algorithm to solve the problem to optimality. This algorithm is based upon constraint satisfaction and a Branch and Bound procedure. We compare the results of our approach with the results of the direct travelers approach for a subnetwork of the Dutch railway system. This comparison indicates a possible, substantial cost improvement.

1.3. Outline

In Section 2 we describe the problem in more detail and we introduce the new approach based on operating costs. This new approach is discussed in relation with the existing direct travelers approach. In addition, the complexity of the problem is discussed. The mathematical programming models are discussed in Section 3. In Section 4 the algorithm is described to obtain an optimal solution. To show the practical use of the model as well as the performance of the solution procedure, an empirical study is carried out. This is presented in Section 5. The paper concludes with a discussion of possible extensions of the model and improvement of the solution procedures.

2. Problem description, complexity and notation

The problem of determining a line system in this paper can be stated as follows:

> Given the railway infrastructure between stations, the traveler flows on each track, the operating costs associated with the exploitation of trains, and service and capacity constraints, determine a cost optimal allocation of lines to passenger flows. The allocation of lines involves the determination of the origin and destination stations of the lines with their frequencies per hour and the length of the trains on each line.

The allocation of lines results in a line system. A line type may be Intercity (IC), Inter Regional (IR), or Agglo Regional (AR). The line type determines the stations at which the line halts: the IC lines halt at just the IC stations, the IR lines halt at both IC and IR stations, and the AR lines halt at all stations. The length of a train is determined by its number of cars. The line systems we consider, cover 1 h.

We consider the allocation of passenger flows to tracks as given. This allocation is determined by a procedure called System Split (see [3]). This procedure is widely used by many authors (see [4,5,2]).

System Split assigns the passengers to the different train types (IC, IR, and AR). The assumption of this passenger-split procedure is that the passengers would travel via the shortest route. Consider, for example, the passengers at AR station x who want to travel to IC station z. Between these stations there is an IC station y. The passengers at x have two possibilities to travel from x to z. They can take an AR line directly from x to z, or they can take an AR line from x to y, and change at y to an IC line from y to z. Depending on the assumptions on the behavior of the passengers, the passengers from x to z are split between these two possibilities. This is done for each pair of stations. After this assignment, the passenger flows on each track can easily be derived.

In the remainder of this section we make the problem description more specific. In Section 2.1 we formally introduce our cost approach. In Section 2.2 we discuss the computational complexity of the problem.

2.1. The cost approach

Our cost approach differs from the existing approach as described in [4,5,2] with respect to the following aspects:

1. We focus on the minimization of operating costs of a line system instead of the maximization of the number of direct travelers.
2. We do not consider the length of the trains belonging to a certain line as fixed. The length of the trains is therefore determined by optimization. For example, a choice will be made between a long train once per hour or a short train twice per hour. Since the number of cars can be balanced with the number of passengers, the number of 'empty' cars during a ride can be reduced.
3. We do not fix the number of trains per hour on each railway track. However, the number of trains per hour on each railway track has to be between certain upper and lower bounds. The upper bound is included for robustness at the operational level, since the railway track must not be used too intensively. The lower bound is included for service considerations, namely to provide a regular connection between the stations at both ends of the railway track.
4. The subdivision in IC, IR and AR passenger flows generated by System Split is considered to be not absolutely fixed. When the travelers on a track can be transported by more line types, a cross optimization is performed over all these line types simultaneously.
5. We take the circulation of rolling stock into account. The circulation of stock is typically considered after the timetable is determined. Since the line system determines to a large extent the best possible circulation of rolling stock, we want to take the circulation of stock into account in the determination of the line system. Based on some realistic assumptions, our model determines an optimal circulation plan, and thus, the costs for the circulation of rolling stock are taken into account.

Our cost optimal approach is not only determined by the above-mentioned aspects. We also had to adopt the following basic assumptions which are commonly accepted by Dutch Railways and other European railway companies.

Firstly, all passenger flows between pairs of stations are symmetric and each line is always operated in both directions, due to the policy of Dutch Railways. The former assumption is a consequence of the data provided by the Dutch Government [6]. The latter is set to provide maximum service to the passengers. As a result of both assumptions, we can restrict our attention to lines in one direction only.

Secondly, we only consider a maximum of three connecting routes of each line between a given pair of origin and destination stations. This maximum is an assumption of Dutch Railways and of System Split. It should be noted that more connected routes can easily be incorporated into our model without changing the mathematical structure. The selected routes have the smallest traveling times, and, in addition, the longest traveling time of these routes has to be less than two times the shortest traveling time. This criterion is tested (see [7]) and found appropriate for the Dutch railway network.

Finally, we only consider the line types as used by Dutch Railways (i.e. IC, IR, and AR). However, other lines types for passenger trains can be implemented easily in our approach. Freight lines and international passenger lines are not considered in this paper, since they are determined by other organizations than Dutch Railways.

All operating costs are taken into account in the cost optimization model. These costs are divided into the following three categories after ample discussions with financial experts of Dutch Railways. The categories presented are identical to the categories used by Dutch Railways. The categories are:

- Fixed costs per car of type t per hour in one direction ($cfix_t$). These costs include depreciation of stock, cost of capital, fixed maintenance costs, and the costs of overnight parking.
- Variable costs per car of type t per kilometer (ckm_t). These costs include ticket collectors costs, cleaning costs, variable maintenance costs, energy costs, infrastructure costs, and mechanics costs.
- Variable costs per train of type t per kilometer (trkm_t). These costs include the driver, the fixed part for ticket collectors, and energy costs.

2.2. Complexity

Bussieck et al. [2] prove that the problem of finding a feasible line system is NP-complete, if the number of trains on each track is fixed. We will prove that the problem of finding a cost optimal allocation of trains without these fixed amount of trains is NP-hard. For that reason, we consider a simplified, abstract version of the line allocation problem.

The simplified cost optimal line problem: Consider a railway network consisting of stations $s \in S$ and tracks $k \in K$ between some pair of stations. The number of passengers on track k is given by n_k. A line $l \in L$ is characterized by its origin and destination station and by the sequence of tracks it passes. The maximum frequency of line l is denoted by $f_l^{\max}$ and the capacity of line l per frequency is ccap_l. The cost of line l per frequency is denoted by c_l. A feasible line system is a set of lines $M \subset L$ such that all passengers can be transported. The problem of finding a cost optimal line system is to find a feasible line system with minimal cost.

Theorem 2.1 *The simplified line problem is NP-hard.*

Proof. We will prove the theorem by a reduction from the vertex cover problem [8], which is known to be NP-hard. The vertex cover problem can be described as follows: Consider a graph $G = (V, E)$. A vertex cover is a subset $W \subset V$ such that $e \cap W \neq \emptyset$ for each edge e of G. The vertex cover problem is to find a vertex cover of minimal cardinality.

Let $G = (V, E)$ be an instance of the vertex cover problem. Number the edges of G consecutively. For each edge e_i introduce two stations s_{2i-1} and s_{2i}. Introduce tracks between all pairs of stations. The number of passengers on a track between stations s_{2i-1} and s_{2i} is equal to 1. The number of passengers on all other tracks is equal to 0. For each vertex v_j introduce a line l_j. Line l_j passes the track between railway stations s_{2i-1} and s_{2i} if and only if $e_i \cap v_j \neq \emptyset$. The capacity and the maximum frequency of all lines are equal to 1. The cost of operating a line is equal to 1. This instance of the simplified cost optimal line problem can be constructed in polynomial time. It is easy to see that a feasible vertex cover of minimal cardinality is equivalent to a cost optimal line system. □

The reduction is illustrated in Fig. 1.

3. Model formulation

In this section we present two mathematical programming models for the cost efficient allocation of passenger trains. In Section 3.1 the notation is introduced that is used throughout this paper. An integer nonlinear programming model

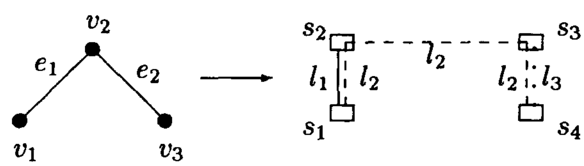

Fig. 1. A graph and its reduced lines and railway stations.

is presented in Section 3.2 as a natural model to capture the various cost components involved. In Section 3.3 this model is transformed into an integer linear programming model.

3.1. Notation

In this paper the following notation for the input data is used:

Indices

- c number of cars
- f frequency of a line ($f = 1,\ldots,f^{\max}$);
- i,j station ($i,j = 1,\ldots,s^{\max}$)
- k track ($k = 1,\ldots,k^{\max}$)
- r connecting route ($r = 1,\ldots,r^{\max}$)
- t line type ($t \in T = \{\text{IC, IR, AR}\}$)

Parameters

- $c_t^{\max}$ the maximum number of cars per train of line type t
- $c_t^{\min}$ the minimum number of cars per train of line type t
- ccap_t the car capacity (number of passengers) for line type t
- $cfix_t$ the fixed costs per hour of one car in one direction of line type t
- ckm_t the costs of one car kilometer with line type t
- cp_{ij}^{rt} factor that multiplied with the frequency and rounded up gives the number of compositions used on a line
- d_{ij}^{rt} the distance between stations i and j over connecting route r with line type t
- $f_k^{\max}$ the maximum train frequency on track k
- $f_k^{\min}$ the minimum train frequency on track k
- i_{ijk}^{rt} takes the value 1 if the line from station i to station j with line type t passes track k on connecting route r, 0 otherwise
- n_k the number of travelers on track k of the right type
- $r^{\max}$ the maximum number of connecting routes between two stations. This number equals three by an assumption of System Split
- trkm_t the costs of one train kilometer with line type t;

The above described input data are determined by three aspects: *the railway infrastructure*, *the costs of operating trains*, and *the process of generating timetables*.

The *railway infrastructure* determines the index sets of i, j, r, t, and k. These index sets express among others the fact that not every station can be used as origin and/or destination station of a line. An actual track between two railway stations may be part of several tracks k. Consider for example two railway stations of type IC with one IR station in between. A line of type IC runs directly between the two IC stations, it does not halt at the IR station in between. An artificial track k_1 is defined for IC lines between these two IC stations and this track consists of two physical tracks, namely the track between the first IC station and the IR station and the track between the IR station and the final IC station. Tracks, k_2 and k_3, are also defined for these two physical tracks separately, since both the IR and AR trains may use the IR station as origin and destination station. Track k_1 can only be used by lines of type IC, not by lines of the types IR and AR. Thus, $i_{ijk_1}^{rt} = 0$ for $t \in \{\text{IR, AR}\}$ for all i, j, r. Each track k in the model has a corresponding line type. The number n_k of travelers on track k represents the number of travelers of the same type.

The infrastructure further determines the possible routes r between each pair of origin and destination stations, the distance between two stations d_{ij}^{rt}, the maximum frequency on each track $f_k^{\max}$ due to limited track capacity, and the maximum number of cars of a line $c_t^{\max}$. The latter is determined by the platform length and the locomotive power.

Costs and other financial aspects do not only determine the variable and fixed costs ($cfix_t$, ckm_t and trkm_t), but also the minimum length $c_t^{\min}$ of a train.

The *hierarchical approach of generating timetables* also has a large impact on several parameters. Since the actual number of passengers is determined by the combination of a line system and the timetable, and the latter is not yet known, an estimation is required. The estimated number n_k of travelers on each track is determined by the procedure System Split, which was explained earlier in this paper. Based on n_k, the minimum frequency of a track $f_k^{\min}$ is determined by criteria of Dutch Railways.

The circulation of rolling stock affects the number of compositions (i.e. identical trains) needed to operate the lines. A composition is the name for a train as a whole. If a train consists of three cars, then three cars are called one composition. We assume that compositions are only used on one line and that the timetable, yet to be determined, will put no extra restrictions on the stock circulation. The number of compositions needed to operate a particular line can be determined based on these assumptions. The assumptions guarantee that the proposed circulation plan is feasible and easy to implement in practice. The proposed circulation plan fits very well with the present circulation plan of Dutch Railways. At present, many lines are operated in this way. For a line the number of compositions depends on the travel time between its origin and its destination station, its frequency and its minimum turn-around time. The number of compositions can be determined from the circulation time. The circulation time is the time needed for a train to travel from its origin to its destination station and back, including the turning time at both its destination station and its origin station. This turning time is needed for the unloading and loading of passengers, the cleaning of the train, maintenance and changing the crew. The turn-around time can be increased by a penalty for the possible extra turn-around time forced by the timetable. The minimum circulation time is divided by 60 (min) to obtain a lower bound on the number of compositions needed for operating a line once per hour, which is denoted by cp_{ij}^{rt}. The total number of compositions needed per day for operating a line is given by cp_{ij}^{rt} multiplied by the frequency of the line and rounded up.

3.2. The integer nonlinear programming formulation

The planners at Dutch Railways apply the following strategy when determining a line system manually. They first decide which *lines* are included in the line system. Then they decide about the *frequency* of the lines. Finally, they determine the *capacity or length* of the trains on each line. To preserve this strategy in our cost approach, the problem can be formulated with the following decision variables: F_{ij}^{rt}, the frequency of a line from station i to station j via route r and of type t. These decision variables represent the first two stages of the described strategy. If a line is not included in the line system, its frequency must be set to zero. C_{ij}^{rt}, the number of cars per train on the line from station i to j via route r and of type t. These decision variables represent the third stage of the strategy. When taking costs into account, these decision variables lead to the following mathematical programming model:

$$\text{minimize} \quad \sum_{r=1}^{r^{\max}} \sum_{t\in T} \sum_{i=1}^{s^{\max}} \sum_{j=i+1}^{s^{\max}} cfix_t \cdot \lceil cp_{ij}^{rt} F_{ij}^{rt} \rceil C_{ij}^{rt} + d_{ij}^{rt}\, \text{ckm}_t\, F_{ij}^{rt}\, C_{ij}^{rt} + d_{ij}^{rt}\, \text{trkm}_t\, F_{ij}^{rt} \tag{1}$$

$$\text{subject to} \quad \sum_{r=1}^{r^{\max}} \sum_{t\in T} \sum_{i=1}^{s^{\max}} \sum_{j=i+1}^{s^{\max}} i_{ijk}^{rt}\, \text{ccap}_t\, F_{ij}^{rt}\, C_{ij}^{rt} \geqslant n_k \quad \forall k, \tag{2}$$

$$f_k^{\min} \leqslant \sum_{r=1}^{r^{\max}} \sum_{t\in T} \sum_{i=1}^{s^{\max}} \sum_{j=i+1}^{s^{\max}} i_{ijk}^{rt}\, F_{ij}^{rt} \leqslant f_k^{\max} \quad \forall k, \tag{3}$$

$$c_t^{\min}\, F_{ij}^{rt} \leqslant F_{ij}^{rt}\, C_{ij}^{rt} \leqslant c_t^{\max}\, F_{ij}^{rt} \quad \forall i, j > i, r, t, \tag{4}$$

$$0 \leqslant F_{ij}^{rt} \leqslant f^{\max} \text{ and integer} \quad \forall i, j > i, r, t, \tag{5}$$

$$C_{ijk}^{rt} \geqslant 0 \text{ and integer} \quad \forall i, j > i, r, t. \tag{6}$$

The objective function (1) represents the costs of a line system. Hence, the sum is taken over all cost components. The objective function considers: the costs of the number of cars being used, the costs of car kilometers and the costs of train kilometers. Constraints (2) ensure that on every track all travelers can be transported. Constraints (3) guarantee that on every track the frequency is between a certain lower (due to service constraints) and upper bound (due to capacity restrictions). Constraints (4) limit the minimum and maximum number of cars per train if the frequency of the line is positive. Constraints (5) ensure a limited positive integer line frequency and constraints (6) state that the length of a train is a positive integer. If the frequency of a line is zero, its train length will be set to zero during optimization. The actual number of operating trains on a line per day is determined in this model by $\lceil cp_{ij}^{rt} F_{ij}^{rt} \rceil$. This discontinuous term is a result of the assumed circulation plan for railway stock.

The mathematical model has *discontinuous* terms in the objective function, *quadratic* terms in both the objective function and the constraints and *integer* decision variables. For this kind of models, no general solution procedures are available [9].

programming model. Tests showed that the results are too unstable for practical use. For more details about this heuristic and the mathematical programming model we refer to Claessens [10].

3.3. The integer linear programming formulation

In order to overcome the computational difficulties mentioned above, we transform the integer nonlinear programming model into an integer linear programming model. This transformation is done by introducing the following binary decision variable for each unique combination of line frequency and train length of a line.

$$X_{ij}^{rtfc} = \begin{cases} 1 & \text{if the line from station } i \text{ to station } j \text{ via route } r \text{ and of type } t \text{ is included in the line system with frequency } f \text{ and } c \text{ cars,} \\ 0 & \text{otherwise.} \end{cases}$$

This decision variable leads to the following mathematical programming model:

$$\text{minimize} \quad \sum_{r=1}^{r^{\max}} \sum_{t \in T} \sum_{f=1}^{f^{\max}} \sum_{c=c_t^{\min}}^{c_t^{\max}} \sum_{i=1}^{s^{\max}} \sum_{j=i+1}^{s^{\max}} \left(cfix_t \lceil cp_{ij}^{rt} f \rceil c X_{ij}^{rtfc} + d_{ij}^{rt} \, \text{ckm}_t \, fc X_{ij}^{rtfc} + d_{ij}^{rt} \, \text{trkm}_t \, f X_{ij}^{rtfc} \right) \tag{7}$$

$$\text{subject to} \quad \sum_{r=1}^{r^{\max}} \sum_{t \in T} \sum_{f=1}^{f^{\max}} \sum_{c=c_t^{\min}}^{c_t^{\max}} \sum_{i=1}^{s^{\max}} \sum_{j=i+1}^{s^{\max}} i_{ijk}^{rt} \, \text{ccap}_t \, f \, c X_{ij}^{rtfc} \geqslant n_k \quad \forall \, k, \tag{8}$$

$$f_k^{\min} \leqslant \sum_{r=1}^{r^{\max}} \sum_{t \in T} \sum_{f=1}^{f^{\max}} \sum_{c=c_t^{\min}}^{c_t^{\max}} \sum_{i=1}^{s^{\max}} \sum_{j=i+1}^{s^{\max}} i_{ijk}^{rt} f X_{ij}^{rtfc} \leqslant f_k^{\max} \quad \forall \, k, \tag{9}$$

$$\sum_{f=1}^{f^{\max}} \sum_{c=c_t^{\min}}^{c_t^{\max}} X_{ij}^{rtfc} \leqslant 1 \qquad \forall \, i, j > i, r, t, \tag{10}$$

$$X_{ij}^{rtfc} \in \{0, 1\} \qquad \forall \, i, j > i, r, t, f, c. \tag{11}$$

Classical relaxation methods for obtaining a lower bound, such as the Lagrange relaxation, are difficult because of the nonlinear terms in the model. We have studied the possibility of using a heuristic based on a relaxation, and using GAMS/MINOS to solve the relaxed mathematical

The objective function (7) represents the costs of a line system. Constraints (8) ensure that on every track all travelers can be transported and constraints (9) guarantee that on every track the frequency is between certain upper and lower bounds. Constraints (10) state that each line must obtain a unique frequency and train length. Finally, constraints (11) state the binary restriction for the decision variable X_{ij}^{rtfc}.

Clearly, Eqs. (7)–(9) resemble Eqs. (1)–(3) of the integer nonlinear programming model, respectively. A multiplication with $f\,c$ is included if in the nonlinear model $F_{ij}^{rt}\,C_{ij}^{rt}$ is used. For terms in the nonlinear model with only F_{ij}^{rt}, a multiplication with f is included in the integer linear model.

4. The algorithm

Our first attempt was to solve the integer linear programming model of Section 3.3 by the CPLEX 3.0 MIP solver. Unfortunately, CPLEX was unable to find a good, feasible solution. Therefore, we have developed a specially designed algorithm. The algorithm has the following major components:

- Model reformulation to reduce the problem size. The derived model reformulation techniques, which are described in Section 4.1, are highly effective for the initial problem formulation of Section 3.3. We were unable to solve the problem to optimality without these techniques.
- Lower bounding. A good lower bound reduces the size of the Branch and Bound tree. Without the techniques of Section 4.2, we were unable to prove optimality for the obtained solutions.
- Selecting a subproblem. This is an important aspect of the search process. The selection procedure, see Section 4.3, makes it possible to obtain a good (or optimal) solution quickly, thereby accelerate the search process.

4.1. Model reformulation: Reducing the problem size

By using model reformulation techniques based on the special structure of our problem, the problem size can be reduced. This is a necessity, since, even for small railway networks, the size of the model stated in Section 3.3 is large. Four model reformulation techniques are presented. Technique 1 aims at reducing the number of constraints. Techniques 2 and 3 aim at reducing the number of variables. Technique 4 aims at the overall improvement of the problem formulation by adjusting the coefficients of the integer linear programming model and by identifying superfluous constraints and variables. The first three techniques are especially developed for this problem and are based on constraint satisfaction: if the stated constraints are satisfied, then the problem can be reduced. Technique 4 is a well known technique applicable to all integer linear programming models.

4.1.1. Technique 1: Redundant tracks

This technique aims at reducing the number of tracks, represented by $k^{\max}$. If the number of tracks is reduced, the number of constraints will also be reduced, since for each track a constraint like (8) and a constraint like (9) must be included in the model. The number of tracks can be reduced by combining connected tracks. No further reductions can be obtained, if each end-point of each resulting track has at least one of the following properties:

1. The end-point represents a railway station that can be used as origin and/or destination station of a line.
2. The track is adjacent to at least two other tracks at the end-point.

Originally, each railway station is represented by one or more end-points of tracks. Since many railway stations do not satisfy one of the above requirements, many tracks can be combined. The minimum required frequency of a composed track is the maximum of the minimum required frequencies of the tracks of which it consists. The maximum required frequency of a composed track is the minimum of the maximum required frequencies of the tracks of which it consists. The number of travelers on a composed track is the maximum of the number of travelers of the tracks of which it consists.

Example: Consider the network in Fig. 2. All tracks can only be used by lines of type AR. Point y represents a railway station that cannot be used as origin and/or destination station. Therefore, all lines traveling over the track between the points x and y must also travel over the track between the points y and z and vice versa. Thus, a composed track xyz can be introduced with $n_{xyz} = \max\{n_{xy}, n_{yz}\}$, $f_{xyz}^{\min} = \max\{f_{xy}^{\min}, f_{yz}^{\min}\}$ and

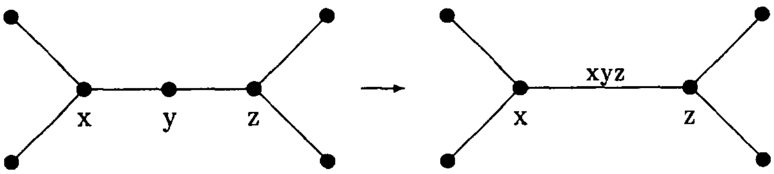

Fig. 2. The original network and the contracted network.

$f_{xyz}^{\max} = \min\{f_{xy}^{\max}, f_{yz}^{\max}\}$. The resulting network is also shown in Fig. 2.

4.1.2. Technique 2: Superfluous variables

The idea behind this technique is that there is a certain upper bound for the train capacity required. Some variables represent a train capacity which is sufficient to transport all passengers at the passed tracks. As a consequence, more train capacity is superfluous and should not be considered, since it will only lead to higher costs. Based upon any of the following conditions, certain decision variables can be removed from a problem. The first requirement gives a sufficient condition to exclude variables with a certain fixed frequency. The second requirement gives a sufficient condition to exclude variables with certain train lengths. The third requirement excludes infeasible frequencies for lines. The conditions are:

1. If $\exists i,j,r,t,f,c$ such that $\forall k$ with $i_{ijk}^{rt} \equiv 1$: $(n_k - (f\, c\, \mathrm{ccap}_t + \max(0, (f_k^{\min} - f)) \min\ \mathrm{ccap}_{t'}\, c_{t'}^{\min} \mid i_{i'j'k}^{r't'} \equiv 1,\ i \neq i' \vee j \neq j' \vee r \neq r' \vee t \neq t'\})) \leqslant 0$, then all variables $X_{ij}^{rtfc^*}$ with $c^* > c$ are superfluous and can be removed from the problem.

   Explanation: if the line corresponding to X_{ij}^{rtfc} is selected, then at least $(f_k^{\min} - f)$ other trains have to pass track k, since $f_k^{\min}$ is the required minimal number of trains on track k. If these other trains are in all cases able to transport the remaining passengers on all relevant tracks, then the line corresponding to X_{ij}^{rtfc} should not have more than c cars. Thus variables corresponding to this line with a number of cars exceeding c can be removed from the problem. $(\max(0,(f_k^{\min} - f)) \min\{\mathrm{ccap}_{t'} \cdot c_{t'}^{\min} | \exists i_{i'j'k}^{r't'} \equiv 1, i \neq i' \vee j \neq j' \vee r \neq r' \vee t \neq t'\})$ represents the minimum capacity offered by other trains that have to pass track k because of the frequency requirements.

2. If $\exists i,j,r,t,f,c$ such that $\forall k$ with $i_{ijk}^{rt} \equiv 1$: $f\, c\, \mathrm{ccap}_t \geqslant n_k$ and $f \geqslant f_k^{\min}$, then all variables $X_{ij}^{rtf^*c^*}$, with $f^* \geqslant f$, $c^* \geqslant c$, and ($f^* \neq f$ or $c^* \neq c$) are superfluous and can be removed from the problem.

   Explanation: if the line corresponding to X_{ij}^{rtfc} is selected, then all passengers can be transported by this line on all the visited tracks. This line also satisfies the minimal frequency requirements on all visited tracks. Thus, variables for this line corresponding to higher frequencies and/or more cars can be removed from the problem.

3. If $\exists i,j,r,t,f,c$ such that $f > \min\{f_k^{\max} \mid i_{ijk}^{rt} \equiv 1\}$, then all variables $X_{ij}^{rtf^*c^*}$ with $f^* \geqslant f \wedge c^* \geqslant 0$ are superfluous and can be removed from the problem.

   Explanation: if the line corresponding to X_{ij}^{rtfc} is selected, then at least one of the maximal frequency requirements on the visited tracks is violated. Thus, variables corresponding to the line with the observed and higher frequencies can be removed from the problem.

4.1.3. Technique 3: Dominated variables

The third technique relies on the idea that some variables can always be replaced by another variable in any feasible line system and this replacement will result in a line system which is feasible and at least as cheap as the original. A variable can be replaced by another variable if, for example, the other variable is able to transport more passengers against lower costs and if the frequency requirements are also satisfied by the resulting line system. This dominance rule is formally defined as follows; Variable $X_{ij}^{rtf^1c^1}$ can be replaced, or is dominated by, variable $X_{ij}^{rtf^2c^2}, f^1 \neq f^2 \vee c^1 \neq c^2$, if the following three requirements are met:

1. $cfix_t \lceil cp_{ij}^{rt} f^1 \rceil c^1 + d_{ij}^{rt}\ \mathrm{ckm}_t\, f^1 c^1 + d_{ij}^{rt}\ \mathrm{trkm}_t\, f^1 \geqslant cfix_t \lceil cp_{ij}^{rt}\, f^2 \rceil c^2 + d_{ij}^{rt}\ \mathrm{ckm}_t f^2 c^2 + d_{ij}^{rt}\ \mathrm{trkm}_t\, f^2$.

Explanation: this condition states that the costs associated with $X_{ij}^{rtf^1c^1}$ have to be greater than or equal to the costs associated with $X_{ij}^{rtf^2c^2}$.

2. $f^1c^1\mathrm{ccap}_t \leqslant f^2c^2\,\mathrm{ccap}_t$
   or
   $\forall k$ with $i_{ijk}^{rt} \equiv 1$: $f^2c^2\mathrm{ccap}_t + \max(0, (f_k^{\min} - f^1)\min\{\mathrm{ccap}_{t'}c_{t'}^{\min} \mid i_{i'j'k}^{r't'} \equiv 1, i \neq i' \vee j \neq j' \vee r \neq r' \vee t \neq t'\}) \geqslant n_k$.
   Explanation: this condition states that the capacity of $X_{ij}^{rtf^1c^1}$ is less than the capacity of $X_{ij}^{rtf^2c^2}$ or that the extra capacity of $X_{ij}^{rtf^1c^1}$ is of no use.
3. $f^1 \geqslant f^2$
   and
   $f^2 + \max\left(0, \left\lceil \frac{(n_k - \min\{f^1c^1\mathrm{ccap}_t, f^2c^2\mathrm{ccap}_t\})}{(\max\{\mathrm{ccap}_{t'}c_{t'}^{\max} \mid \exists i_{i'j'k}^{r't'} \equiv 1, i \neq i' \vee j \neq j' \vee r \neq r' \vee t \neq t'\})} \right\rceil\right) \geqslant f_k^{\min} \forall k : i_{ijk}^{rt} \equiv 1.$
   Explanation: this condition states that the frequency requirements are always satisfied whenever $X_{ij}^{rtf^1c^1}$ is replaced by $X_{ij}^{rtf^2c^2}$.

Thus, if all three requirements are fulfilled, then $X_{ij}^{rtf^1c^1}$ is dominated by $X_{ij}^{rtf^2c^2}$ and the former can be removed from the problem.

4.1.4. Technique 4: CPLEX preprocessor

The technique tries to improve the overall problem formulation by sharpening the coefficients of the integer linear programming model, by identifying superfluous constraints and variables and by substituting variables. We used the MIP preprocessing of CPLEX 3.0. More details can be found in [11–13]. This technique is well known [11–13] and implemented in many optimization packages.

4.2. Lower bounding

A lower bound for the integer linear programming model is obtained by removing the integrality restrictions (11) for variables X_{ij}^{rtfc}. For calculating the corresponding LP relaxation we used the primal simplex method at the root of the Branch and Bound tree and the dual simplex method in all other nodes.

We applied two techniques for improving the value of the LP relaxation. Technique 5 results from the observation that only an integer number of cars can be used in any feasible solution, and must be executed before the Branch and Bound procedure is started. Technique 6 involves the addition of valid inequalities, namely cover inequalities, during the Branch and Bound procedure.

4.2.1. Technique 5: Right-hand side

This technique is based upon the fact that only an integer number of cars can be used on each track. Hence, we can increase the right-hand sides of constraints (8). If the passengers on track k can only be transported by a single line type t, n_k can be increased to $\lceil n_k/\mathrm{ccap}_t \rceil \mathrm{ccap}_t$. If several line types, suppose $T' \subseteq T$, can be used, n^k can only be increased to the lowest reachable value greater than or equal to n^k for an integer combination of the available car capacities $\mathrm{ccap}_t, t \in T'$. This technique is applied after Techniques 1–3, but before Technique 4. This is a consequence of the fact that the effect of Technique 4 depends on the effect of Technique 5.

4.2.2. Technique 6: Cover inequalities

The value of the LP relaxation can be further improved by the use of cover inequalities. A cover is a set S of variables that has the property that if all variables of the set are set to one, then the remaining problem becomes infeasible. A cover inequality states that the sum of the variables in an associated cover must be less than or equal to the cardinality of the cover minus one, e.g. $\sum_{s \in S} X_s \leqslant |S| - 1$. The cover inequalities are derived from minimal covers. A minimal cover is a cover (i.e. it satisfies the above-mentioned property), but the resulting set would no longer be a cover if any of the variables would be removed from the cover. For a further discussion of cover inequalities, we refer to Hoffman and Padberg [14].

We search for violated cover inequalities after having solved the LP relaxation at a node in the Branch and Bound tree. Violated cover inequalities may exist because of the maximum frequency constraints. If violated cover inequalities are found, they are added to the problem description and the LP relaxation is solved again. This process is repeated until no more violated cover inequalities exist.

4.3. Selecting a subproblem

The subdivision of a problem into subproblems and deciding which subproblem is considered next from the available list of subproblems are major aspects of a Branch and Bound procedure. Our branching rule is to force a variable with a fractional value to integrality. This is done in the regular fashion of setting the variable to one in one subproblem and to zero in the other. The selection of the fractional variable is based upon the automatically selection of CPLEX 3.0, see [11]. The selection of the next subproblem is based upon an estimate of the best obtainable integer feasible solution for the subproblem. This estimate of the best obtainable integer objective value of the considered subproblem is obtained by removing all variables with a fractional value from the objective function value of the LP relaxation. Heuristically speaking, this selection procedure selects the best partial solution. This selection rule is able to find a good solution quickly, and thus accelerates the search process in the Branch and Bound tree. Selecting the subproblem based upon the best objective value of the associated LP relaxation turned out to be unsuccessful since the gap between the value of the LP relaxation and the value of the resulting integer solution was too large.

Finally, all variables can be subdivided into several mutually exclusive sets, which are useful in the Branch and Bound procedure. These sets belong to the so called Special Ordered Sets (SOSs) of type 1, see [15,3,16]. An SOS of type 1 has the property that at most one of the variables in the set has a nonzero value in any feasible integer solution. These sets are defined by the variables occurring in the same constraint (10). Thus, each SOS of type 1 has the interpretation that each line must obtain a single frequency and train length.

5. Application to a Dutch railway subnetwork

We carried out an empirical study on real world data to illustrate the application of our model as well as the performance of our algorithm. The basic variant used for our computations is described in Section 5.1. In Section 5.2 the performance of the model reformulation techniques and the Branch and Bound procedure are presented. In Section 5.3 we compare the line system obtained by our cost approach with the line system obtained by the direct travelers approach. We end this section with a sensitivity analysis.

5.1. Introduction

We use the north west part of The Netherlands as the basic variant for our computational experiments. The infrastructural lay-out of this subnetwork of the Dutch railway network is shown in Fig. 3. This subnetwork contains 28 railway stations, of which one is an IC station, nine are IR stations and 18 are AR stations. The subnetwork has 10 railway stations that can be used as origin and/or destination station.

The software system PROLOP is used to assign the passenger flows to the railway tracks by the described procedure System Split and to obtain a line system determined by the direct travelers approach.

We have used CPLEX 3.0 [11] as the basis for our Branch and Bound procedure. GAMS [9] is used for modeling the problem as an integer linear programming model and for implementing our model reformulation (Techniques 1–3) and lower bounding (Technique 5) techniques.

5.2. Performance of the algorithm

The obtained initial integer linear programming model has 5629 variables, 194 constraints, about 110,000 nonzero coefficients (nonzeros) and an LP lower bound of about 6920. Before actually starting the Branch and Bound part of our algorithm, the problem is reduced to 1547 variables, 139 constraints, and about 18,000 nonzeros. The LP lower bound is increased to about 7577. Thus, a substantial improvement of the initial problem formulation is obtained. The cumulative effect of the model reformulation and lower bounding techniques is described in Table 1. The percentages indicate the total effect in comparison with the initial problem size.

Techniques 2 and 3 were able to identify many redundant decision variables and thus to decrease the problem size significantly. The use of the lower bounding Technique 5 was crucial, since the stronger lower bound prevented that the size of the Branch and Bound tree exceeded the available

Fig. 3. Dutch railway network in the north west of The Netherlands.

Table 1
Problem size and LP lower bound after each of the model reformulation and lower bounding techniques.

	Initial	Tech 1	Tech 2	Tech 3	Tech 5	Tech 4	Total%
# Variables	5629	5629	1708	1548	1548	1547	−73
# Constraints	194	143	143	143	143	139	−28
# Nonzeros	111,733	64,591	21,313	19,913	19,913	18,192	−84
Lower bound	6920	6920	6975	6975	7255	7577	+10

computer memory space. The MIP preprocessor of CPLEX 3.0, that we used as Technique 4, eliminated four constraints and one variable and modified 3860 coefficients. Especially the improvement of the objective value of the LP relaxation is noticeable for this technique. For this problem no violated cover inequalities existed during the search of the Branch and Bound tree. However, many cover inequalities were found for only slightly changed problems. Thus, cover inequalities can be useful for other problems.

The computing times (in CPU seconds on an SUN LX workstation 50 MHz) are described in Table 2. The CPU times represent the Branch and Bound part of the algorithm, including Technique 4. Techniques 1–3 and 5 were performed on an MS-DOS operated 486DX2-66 computer. The total computation times of these techniques accumulated to about 10 s. The conclusion can be drawn from the table that reasonable solutions are found rather quickly. Most time was spent on proving optimality for the found solution, since the optimal solution itself was found already after 850 CPU seconds. The total computing time of 3989 CPU seconds is not significant in practice, since the determina-

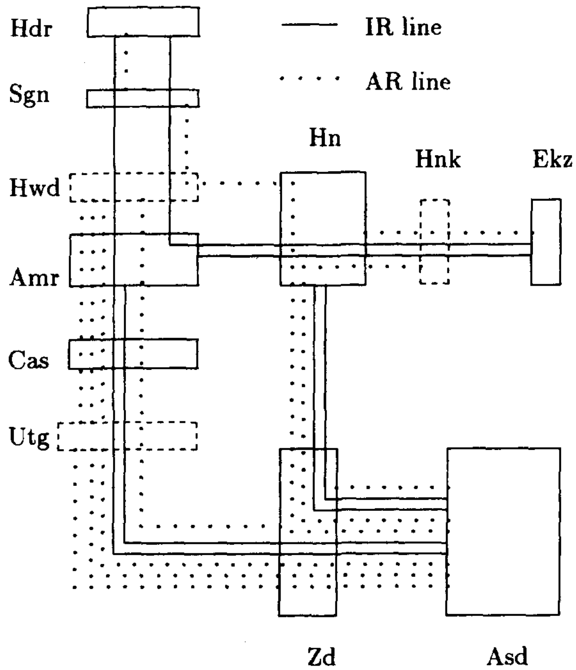

Fig. 4. Line system obtained by the cost approach. Each line in the figure represents one train per hour in both directions.

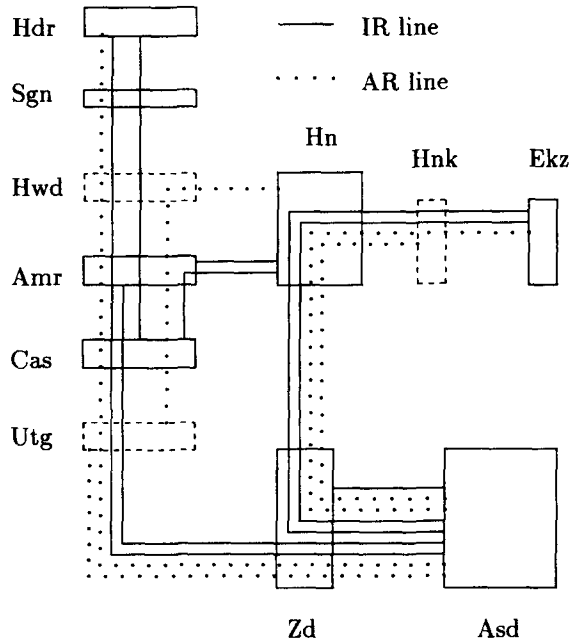

Fig. 5. Line system obtained by the direct travelers approach. Each line in the figure represents one train per hour in both directions.

tion of a line system occurs only a few times per year.

5.3. Comparison of cost versus direct travelers approach

The line system obtained by the cost approach is displayed graphically in Fig. 4. A line system was also determined for the described problem based on the direct travelers approach of PROLOP. The obtained line system for this approach is displayed graphically in Fig. 5. The line system was determined using the same allocation of the travelers to the tracks and the same maximum and minimum frequencies on each track. A line in these figures corresponds to one line in the line system, operated once per hour in both directions. The IC and IR stations, which are displayed as a solid box, are Alkmaar (Amr), Amsterdam CS (Asd), Castricum (Cas), Den Helder (Hdr), Enkhuizen (Ekz), Hoorn (Hn), Schagen (Sgn), and Zaandam (Zd). The AR stations, displayed as a dashed box, are Heerhugowaard (Hwd), Hoorn Kersenboogerd (Hnk), and Uitgeest (Utg).

Table 2
CPU time of the Branch and Bound procedure

	CPU time (s)
Solving LP	0.82
10% MIP gap	29
5% MIP gap	77
0% MIP gap	3989

Both line systems are compared on their operating costs (Table 3) and on their numbers of direct travelers (Table 4). Table 3 clearly indicates a substantial cost reduction. The total operating costs are reduced by 1627 Dutch guilders per hour (17.2%), which amounts to nearly 18 million Dutch guilders per year for this part of the railway network only. Dutch Railways assume to operate their trains 17 h per day for 320 days per year in both directions. The cost reduction is obtained by reducing the unused capacity as well as the necessary railway stock. The total number of unused car seats is decreased by 4435 (43%). The decrease

Table 3
One-direction operating costs and other solution characteristics

Line system	Direct travelers	Cost optimal	Difference	
Costs/hour	9473.3	7845.9	−1627.4	(−17.2%)
Unused seats	10,335	5900	−4435	(−42%)
Empty seat kilometers	66,391	33,606	−32,785	(−49%)
Car kilometers	1970	1573	−397	(−20%)
Train kilometers	619	670	+51	(+8%)
Cars needed per day	100	77	−23	(−23%)
Av. train length (in cars)	3.2	2.7	−0.5	(−15%)
Av. route length (in km)	51.6	37.2	−14.4	(−27%)

Table 4
Number of direct travelers and other solution characteristics

Line system	Direct travelers	Cost optimal	Difference	
Direct travelers (max = 68200)	65,996	62,051	−3945	(−6%)
Direct links (max = 351)	210	154	−56	(−26.6%)
Average number of travelers on a direct link	314	398	+84	(+26.8%)

in the total number of kilometers covered by these unused car seats is even slightly higher, namely 49%. The reduction of required railway stock is measured by the number of cars needed per day to operate the designed line system. This number is based upon our assumptions on the circulation plan for railway stock. Clearly, the lines are selected for which a much more efficient circulation plan can be constructed.

The difference in the total number of travelers that can travel directly from their origin station to their destination station is given in Table 4. [1] All figures correspond to travelers in only one direction per day. A conclusion from the table is that the less busy connections are no longer offered a direct link, because of cost implications. Since the decrease in operating costs is much higher than the decrease in the total number of direct travelers, the overall conclusion is that the line system of the cost approach is highly competitive with the line system obtained by the direct travelers approach.

[1] Based on data offered by NS Travelers, Marketing Research and Advice.

5.4. Sensitivity analysis

In this section, the influence of several constraints and assumptions of our approach are investigated. We investigate the influence of the removal of the minimum frequency constraints and the influence of determining a cost optimal line system for the IR and AR network separately.

First, the influence of the minimum frequency requirements is analyzed by removing them. These minimum frequency requirements are included in our model to provide a regular connection between the railway stations adjacent to the railway track. The costs of the resulting line system are 7274 Dutch guilders per hour, which is a reduction of 7.3% in comparison with the cost approach of Section 5.3 and a reduction of 23% in comparison with the direct travelers approach. The total number of direct travelers for this problem is 62022. This is a 0.05% reduction with respect to the original cost approach and a 6% reduction with respect to the direct travelers approach.

Secondly, we determined a line system independently for each subnetwork, i.e. the subdivision in

IR and AR passenger flows generated by System Split is considered absolutely fixed. The costs of the resulting line systems are 3033.5 Dutch guilders per hour for the IR subnetwork and 5220 Dutch guilders for the AR subnetwork. Therefore, the costs per hour for the combined line system are 8253.5 Dutch guilders which is an increase of 5.2% in comparison with the cost approach of Section 5.3 and a reduction of 12.9% in comparison with the direct travelers approach. The total number of direct travelers for this problem is 62,262. This is a 0.3% increase with respect to the original cost approach and a 5.7% reduction with respect to the direct travelers approach. The computing time for the IR line system is 4 CPU seconds and 205 CPU seconds for the AR line system. So the total computing time is reduced with almost 95% in comparision with the computing time of the cost approach of Section 5.3.

6. Summary and conclusion

In this paper we considered the problem of the allocation of lines to passenger flows. We presented a new approach that takes operating costs into account. This approach introduces cost elements early in the overall timetable generation process. This approach was translated into a mathematical programming model that optimizes on lines, line types, routes, frequencies and train lengths. A Branch and Bound procedure was outlined and implemented to solve the problem to optimality. We applied the solution procedure to a part of the Dutch railway network. The test results show that for this particular study the decrease in costs is significant.

The usefulness of our model lies not only in the fact that it can be used to obtain a cost optimal line system but also be used to determine the costs of an existing line system. The model can therefore serve as an aid when choosing the best line system from several alternatives.

Future research can be directed towards combining our cost approach with the direct travelers approach. One possibility is to implement in our model the manual planning method of Railned [17] for optimizing the number of direct travelers. This method requires an estimation of the number of direct travelers on each possible line in advance. This can be done by adjusting the coefficients of our objective function.

Acknowledgements

The work described in this article is supported by Railned and Nederlandse Spoorwegen (Netherlands Railways), The Netherlands. The detailed and thoughtful comments of the referees have been highly appreciated.

References

[1] M.A. Odijk, P.J. Zwaneveld, J.S. Hooghiemstra, L.G. Kroon, M. Salomon, Decision support systems help railned to search for 'win-win' solutions in railway network design ERASM management Report series no. 286, Rotterdam School of Management, Erasmus University Rotterdam, Rotterdam, The Netherlands, 1996.

[2] M.R. Bussieck, P. Kreuzer, U.T. Zimmermann, Optimal lines for railway systems, European Journal of Operational Research 96 (1997) 54–63.

[3] C. Oltrogge, Linienplanung für mehrstufige Bedienungssysteme im öffentlichen Personenverkehr, Ph.D. Thesis, TU Braunschweig, 1994 (In German).

[4] H. Dienst, Linienplanung im spurgeführten Personenfernverkehr mit Hilfe eines heuristischen Verfahrens, Ph.D. Thesis, TU Braunschweig, 1978 (In German).

[5] A. Bouma, C. Oltrogge, Linienplanung und Simulation für öffentliche Verkehrswege in Praxis und Theorie, ETR – Eisenbahntechnische Rundschau, 1994, pp. 369–378 (In German).

[6] Hague Consulting Group, Het Landelijk model systeem voor verkeer en vervoer, Rapport C: Methoden en Modellen, Rijkswaterstaat Dienst Verkeerskunde, Afdeling Landelijke Verkeersprognose en Beleidsanalyse, Governmental Report Rotterdam, 8, 1990 (In Dutch).

[7] A. Dam, C. Oprel, PROLOP toedelings- en vergelijkingsmodel, Technical Report, Dutch Railways, 1995 (In Dutch).

[8] M.R. Garey, D.S. Johnson, Computers and Intractability: A guide to the Theory of NP-Completeness, Freeman, San Francisco, 1979.

[9] A. Brooke, D. Kendrick, A. Meeraus, GAMS: A User's Guide Release 2.25, The Scientific Press, San Francisco 1992.

[10] M.T. Claessens, De kosten-lijnvoering, Master's Thesis, University of Amsterdam, 1994 (In Dutch).

[11] CPLEX Optimization Inc. Using the CPLEX Callable Library, Version 3.0, 1994, Manual.

[12] H.P. Crowder, E.L. Johnson, M.W. Padberg, Solving large-scale zero-one linear programming problems, Operations Research 31 (1983) 803–934.
[13] M.W.P. Savelsbergh, Preprocessing and probing techniques for mixed integer programming, ORSA Journal on Computing 6 (1994) 445–454.
[14] K.L. Hoffman, M. Padberg, LP-based combinatorial problem solving, Annals of Operations Research 6 (1985) 145–194.
[15] E.M.L. Beale, J.A. Tomlin, Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables, in: J. Lawrence (Ed.), Proceedings of the Fifth International Conference on Operational Research, Tavistock Publications, London, pp. 447–454.
[16] E.Y.H. Lin, Multiple choice programming: a state-of-the-art review, International Transactions in Operational research 14 (1994) 409–421.
[17] F.L. Härte, Het verbeteren van dienstregelingen en het opstellen van lijnvoeringen, Master's thesis, Free University Amsterdam, 1995 (In Dutch).