

Optimal location of intermodal freight hubs

Illia Racunica

Institut Eurecom

2229 Route des Crêtes, B.P. 193

06904 Sophia Antipolis, France

and

Laura Wynter

PRISM, Université de Versailles

and INRIA

Domaine de Voluceau, B.P. 105

78153 Le Chesnay, France

Abstract

Attempts at reducing the externalities of freight transport in Europe are generally focused on the incorporation of a more significant use of rail into freight itineraries. One new scenario for increasing the share of rail in intermodal transport involves the development of a dedicated subnetwork of freight rail lines. Within this European Union project, the use of hub-and-spoke type networks, in combination with fixed-length shuttle services for freight, are under discussion. We present this innovative project and the proposed optimization model. The problem is cast as a nonlinear concave-cost hub location problem. A linearization procedure along with two efficient variable-reduction heuristics are proposed for its resolution, making use of recent results on polyhedral properties of this class of problems. Computational experience is provided on the European network of potential hub terminals.

The most recent scenarios under study for integrating freight transportation in Europe efficiently and with minimum social and environmental cost involve extensive use of intermodal transport. The intent of these new transport scenarios is to

make maximum use of rail transport, not only for long haul and low value freight distribution, as has been the case, but over medium-length distances as well.

To permit rail to be competitive with road haulers in Europe, one of the few policies that remains viable is the use of a subnetwork of dedicated or semi-dedicated (that is, in which freight is given priority) freight rail lines. Indeed, it has been observed that complete integration with passenger rail services has rendered it difficult to increase the market share of rail in the freight transport sector, since freight slots are generally given only at night, or between passenger trains during the day. In all cases, on mixed-use lines, passenger services are given priority, to the detriment of the quality of service offered by rail freight providers.

The primary reasons cited for the predominance of road haulers are the low travel times they offer, and their high flexibility in terms of departure times. Currently, in Europe, freight rail schedules are planned weeks in advance, and last-minute requests can be accommodated only through insertions between passenger trains and planned freight service. Consequently, flexibility is severely limited. The development and integration, therefore, of a dedicated or semi-dedicated freight subnetwork is the focus of several large-scale European Union research projects (see INRETS final reports: EUFRANET, IQ, and SCENES [7]). With respect to reducing freight travel times on rail services, high-speed freight rail services are being considered as a potential component of the proposed freight subnetwork.

Within this context, investigations have examined the use of network configurations prevalent in airline transportation, such as shuttle services operating on regular, frequent schedules. Of particular interest is the fact that shuttle services would permit the use of a viable hub-and-spoke network configuration for rail freight transport, in that rapid and reliable hub-to-hub transfers could be included in freight itineraries. These itineraries could then include high-speed freight train services on a few corridors. Fixed-length shuttle services mean reduced terminal times, since trains need not be reconfigured, and would allow far greater flexibility, since reservations would not be required, much as is the case with airline shuttles.

An essential feature of shuttle services, and hub-and-spoke networks in general, is the economies of scale that can be gained by consolidation as well as by reducing shunting costs. This is crucial for rail freight transport, since current operating costs are high compared to those of road haulers, and freight consolidation would permit substantial cost savings to the rail operators.

This paper is concerned with devising a model for the conception of such an innovative hub-and-spoke network for multimodal freight transport on dedicated freight

rail lines. Of particular interest is the incorporation of the scale economies resulting from freight consolidation at hub terminals. Desired results of the model include information on the optimal hub locations: How much freight can each (potential) hub capture? What percentage of the market can the shuttle services take from both the current direct complete block train service, and from the share of the road haulers? Which corridors can become competitive enough to warrant the construction costs associated with high-speed freight train services? How would an evolution of freight flows effect these conclusions?

Since it is necessary to evaluate both the market share of the hubs themselves and of the lines (shuttle services versus direct train service and road service), coupled with the need to explicitly include economies of scale, the choice of a model representing each path explicitly appears judicious as opposed to a more compact arc-based formulation of the network flow problem. Furthermore, the importance of the construction and development costs in converting terminals into “mega hub” nodes, capable of handling shuttle services, leads us to adopt a network design approach.

We therefore model this rail freight network design problem as an uncapacitated hub location problem with concave increasing costs on certain links. The concave costs represent economies of scale that can be obtained between two hub nodes, and from hubs to non-hub destination terminals. Explicit capacity constraints were not included, as one desired result of the model is to evaluate the maximum average frequency attainable by rail on any freight corridor.

The objective of the model is then to minimize a linear combination of hub construction costs and travel costs, where the latter takes into account the effect of freight consolidation where applicable. The rate of increase of direct (origin-to-destination) complete block train service is taken to be a linear function of distance, since the number of wagons is directly proportional to the tonnage carried. Conversely, given the fixed composition of shuttle services, the marginal cost to the user on these lines decreases considerably with increasing flow, as the additional cost for each ton hauled is lower than the cost of running the train. Similarly, for non-shuttle services between hub terminals and destination terminals, sufficient consolidation is believed to allow for some economies of scale, though less than can be attained on the fixed-length shuttle lines.

The resulting model, described in the following section, is cast as a nonlinear mixed-integer program. Constraints on the model concern the maximum number of hubs that can be opened, as well as the maximum number of hubs traversed in any path. The latter constraint is included implicitly within the definition of the flow variables. We present further some recently developed polyhedral properties of this

model. In order to correctly model the influence of train frequency on choice decisions and on travel costs, we have chosen to implicitly include frequency effects through a calibration of the cost function, rather than through the use of a dynamic model with time as a parameter. This model is therefore similar to that of *frequency network design* with frequencies as derived output, within the context of the classification proposed by Crainic (2000).

In Section 2, we discuss solution procedures for the model. To solve the program efficiently, a linearization scheme is used for the nonlinear flow term. The linearized problem is then coupled with a very efficient heuristic that permits approximating the effect of the concave costs with a linear optimization solver. We demonstrate as well the effect of incorporating the polyhedral information into the formulation.

Section 3 provides numerical results on the European rail freight network and interprets the benefits acquired through the use of this model. Finally, in Section 4, we discuss a number of interesting avenues to be explored within the scope of this problem.

1 Formulation as a nonlinear mixed-integer hub location model

The problem of the optimal location of hubs in a network has received attention over the past decade due to its importance in air transportation, and also in telecommunications. See, for example, Bryan and O’Kelly (1999), Campbell (1994), or Campbell (1996). The objective is to determine *a posteriori* the number of hubs to be opened and the paths used in the network, where a hub is opened only if it is profitable to do so. The definition of “profitable” is given in terms of hub opening costs and travel time savings, where the latter are, in principle, due both to sufficient consolidation and as well as trip time reduction. The basic model, presented below, has been studied and some algorithms proposed in the references O’Kelly et al. (1995), Klincewicz (1996), Skorin-Kapov et al. (1996), Ernst and Krishnamoorthy (1998), and Hamacher et al. (2000).

1.1 The basic hub location model

Let us define first the graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ where \mathcal{N} is the set of all terminal nodes, and the set of potential hub nodes is $\mathcal{H} \subset \mathcal{N}$. The set of arcs is defined as \mathcal{A} , where

each arc is weighted by its travel time. Paths in the graph are identified as a sequence of the nodes traversed, where this number is limited to at most two hub nodes per path by the definition of the path variables. The standard uncapacitated hub location model is then given as follows:

$$\min_{x,z} F(x, z) := \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{H}} \sum_{k \in \mathcal{H}} \sum_{l \in \mathcal{N}} c_{ijkl} x_{ijkl} + \sum_{j \in \mathcal{N}} f_j z_j \quad (1)$$

subject to

$$\sum_{j \in \mathcal{H}} \sum_{k \in \mathcal{H}} x_{ijkl} = d_{il}, \quad \forall (i, l) \in \mathcal{W}, \quad (2)$$

$$\sum_{j \in \mathcal{N}} x_{ijkl} \leq Q_{il} z_k, \quad \forall k \in \mathcal{H}, (i, l) \in \mathcal{W}, \quad (3)$$

$$\sum_{k \in \mathcal{N}} x_{ijkl} \leq Q_{il} z_j, \quad \forall j \in \mathcal{H}, (i, l) \in \mathcal{W}, \quad (4)$$

$$x_{ijkl} \geq 0, \quad \forall j \in \mathcal{H}, \forall k \in \mathcal{H}, (i, l) \in \mathcal{W}, \quad (5)$$

$$0 \leq z_j \leq 1, \quad \forall j \in \mathcal{N}, \quad (6)$$

$$z_j \in \{0, 1\}, \quad \forall j \in \mathcal{N}, \quad (7)$$

where $d \in \mathfrak{R}_+^{|\mathcal{W}|}$ is the vector of demands over the set $\mathcal{W} \subset \mathcal{N}^2$ of origin-destination (o-d) pairs. The flow variables are given by $x \in \mathfrak{R}_+^{|\mathcal{N}|^2 \times |\mathcal{H}|^2}$, and $z \in \{0, 1\}^{|\mathcal{H}|}$ is the vector of discrete decision variables indicating whether a hub is to be opened or not.

The constant Q_{il} is defined such that $Q_{il} \geq d_{il}$, $(i, l) \in \mathcal{W}$, in which case, equations 3 and 4 ensure that hub i (resp. j) is open for the flow through that hub to be non-zero. Costs on the path (i, j, k, l) are given by c_{ijkl} , while f_j is the cost associated with converting the terminal j into a hub node.

1.2 Taking into account economies of scale

In order to take into account the cost reductions that are obtained by consolidation at hub nodes, the technique used in the standard hub location model is to apply a so-called *discount factor* on the interhub links of the network, so that the per-unit price on interhub links is lower than that on extremal links of the network. See, for example, Ernst and Krishnamoorthy (1998), Hamacher et al (2000), O’Kelly et al. (1995), or Skorin-Kapov et al. (1996) where the authors use the following objective function, which makes use of the discount factor, α , where $0 < \alpha < 1$:

$$c_{ijkl} := c_{ij} + \alpha c_{jk} + c_{kl}. \quad (8)$$

That is, the unit cost per distance is lower between pairs of hubs than otherwise, but the marginal costs on all links are constant with flow.

It is clear, however, that objective function (1) with costs given by (8) does not model scale economies, which require that the the marginal price *decreases* with increasing flow, in which case, the cost function must be strictly concave increasing, rather than linear. Clearly, this simplification is costly in terms of accuracy of the solution since large and small flow values all receive the same discount.

To deal with this deficiency, we generalize the definition of some of the terms in (8). While it is technically feasible to generalize the definition of all three terms in (8), our motivation is to accurately model the economies of scale arising from the use of hub-and-spoke shuttle service for freight transport. Therefore, direct (origin-to-destination) complete block train service is taken to be a linear function of distance, since the number of wagons is directly proportional to the tonnage carried. Shuttle services between two (mega-) hub nodes are designed to operate with a fixed composition, so as to reduce shunting costs; consequently, the marginal costs decrease considerably with increasing flow on hub-to-hub shuttle lines. Similarly, for non-shuttle services between mega-hub terminals and destination terminals, sufficient consolidation, along with somewhat reduced shunting costs at the mega-hub node is believed to allow for some economies of scale, though less than can be attained between two mega-hubs. To summarize, then, we have included nonlinear concave cost terms between pairs of mega-hub terminals, and from mega-hub to destination terminals, where the nonlinear coefficient on the latter routes will be higher than on the former.

Furthermore, since a primary objective of the model is to evaluate the market share of the hubs with respect to currently existing direct, complete block train itineraries, it is necessary to include paths not passing through the hub nodes as well. Flow on a direct path from i to l is therefore represented by the variable \hat{x}_{iil} , where nodes i and l are not hub nodes. Note that itineraries passing through exactly one hub are implicitly defined in the original model, where the flow is given as x_{ijkl} with $j = k$. Adding these characteristics to objective function (1), we obtain:

$$\Psi(x) := \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{H}} c_{ij} \sum_{k \in \mathcal{H}} \sum_{l \in \mathcal{N}} x_{ijkl} + \sum_{j \in \mathcal{H}} \sum_{k \in \mathcal{H}} c_{jk}^1(x) + \sum_{k \in \mathcal{H}} \sum_{l \in \mathcal{N}} c_{kl}^2(x), \quad (9)$$

where $c^1(x)$ and $c^2(x)$ are nonlinear *discount functions*. Then, the objective function

can be expressed as:

$$\min_{x, \hat{x}, z} F(x, \hat{x}, z) := \Psi(x) + \sum_{(i,l) \in \mathcal{W} \setminus \mathcal{H}} \hat{c}_{il} \hat{x}_{iil} + \sum_{j \in \mathcal{H}} f_j z_j \quad (10)$$

Note that the constraint (2) must be replaced by:

$$\sum_{j \in \mathcal{H}} \sum_{k \in \mathcal{H}} x_{ijkl} + \hat{x}_{iil} = d_{il}, \quad \forall (i, l) \in \mathcal{W} \quad (11)$$

so as to include flows on direct routes.

We first make the following assumption.

Assumption 1 *The functions $c_{jk}^1 : \mathfrak{R}_+^{|\mathcal{W}|} \mapsto \mathfrak{R}_+$ (resp. $c_{kl}^2 : \mathfrak{R}_+^{|\mathcal{V}| \times |\mathcal{H}|} \mapsto \mathfrak{R}_+$) are continuous, concave increasing functions for all nonnegative x , for all pairs of nodes $(j, k) \in \mathcal{H}^2$ (resp. $(k, l) \in \mathcal{H} \times \mathcal{N}$).*

Then, we have the following property of the objective function.

Proposition 1 *The objective function $F(x, \hat{x}, z)$ is a continuous, nonnegative, and concave function of (\cdot, \cdot, z) , affine in (x, \hat{x}, \cdot) , for all nonnegative x, \hat{x} .*

The resulting concave-cost hub location model is given then by equations (9), (10) and constraint set (11), (3)–(7).

1.3 Estimating the effects of weekly frequency

In order to take into account the effect of the weekly frequency on the travel costs, one could formulate a fully dynamic model in which time is explicitly included as a parameter, and, for example, over a weekly horizon. However, the resulting complexity of the model would be significant. Instead, we have chosen to implicitly take into account the effect of time within a static model through an appropriate calibration of the nonlinear term in the cost function.

Since both direct itineraries and itineraries passing through hub nodes are maintained in the model, we can define a generic concave cost function, and calibrate its parameters so that the interhub shuttle trip becomes more economic than the direct service at the minimum weekly frequency needed to make the shuttle service

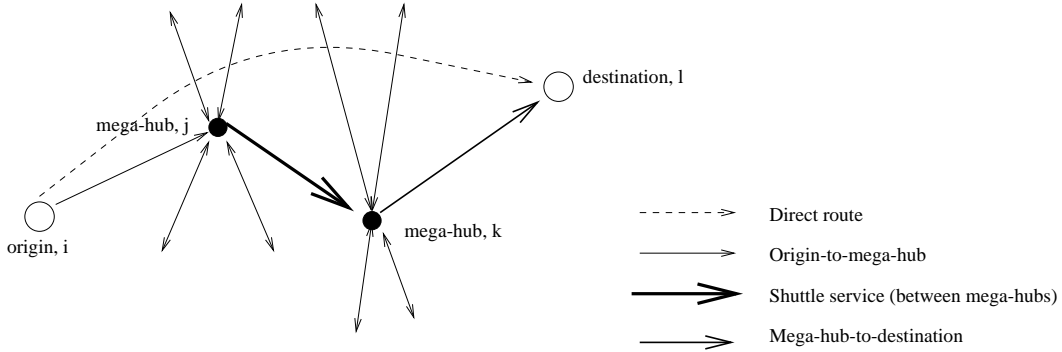


Figure 1: Direct versus hub-based itineraries for a typical origin-destination pair

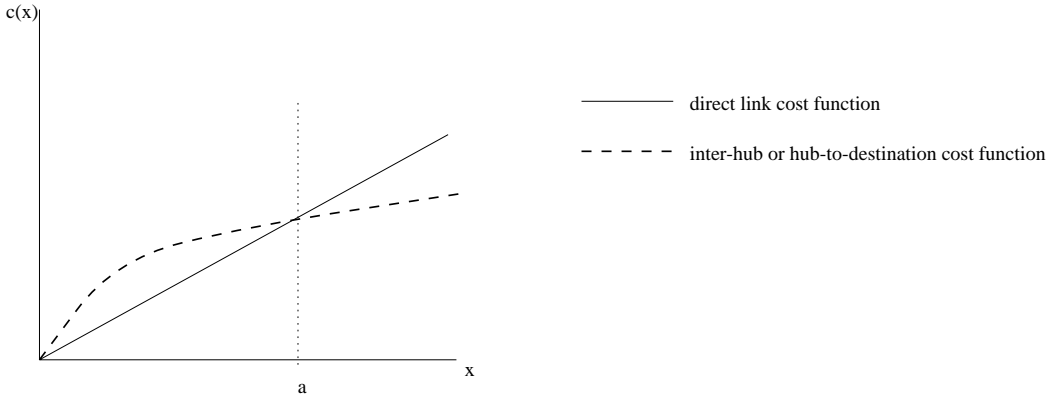


Figure 2: Cost function calibration between a direct and hub-based trip leg

viable. Consider the Figure 1 which represents the itinerary choices for a given origin-destination pair, (i, l) . The Figure 2 illustrates the calibration of the concave cost curves with respect to the linear, direct (complete block train) cost curves on a typical example. The two curves should intersect precisely where the shuttle services becomes advantageous. Quantitative data on this point of intersection permits calibration of the cost function parameters.

Letting the concave interhub discount function c_{jk}^1 be given by $c_{jk}^1 = a_{jk}^1 (\sum_{(i,l) \in \mathcal{W}} x_{ijkl})^{0.5}$, we set

$$a_{jk}^1 \left(\sum_{(i,l) \in \mathcal{W}} \bar{x}_{ijkl} \right)^{0.5} = \bar{c}_{jk} \sum_{(i,l) \in \mathcal{W}} \bar{x}_{ijkl} \quad (12)$$

where \bar{c}_{jk} is the cost associated with direct, complete block train service between each pair of terminal nodes $j \in \mathcal{H}$ and $k \in \mathcal{H}$. The term $\bar{x}_{jk} := \sum_{(i,l) \in \mathcal{W}} \bar{x}_{ijkl}$ is the amount of flow in tonnes corresponding to the minimum number of complete block trains needed to make shuttle service viable, where an average conversion factor (e.g., 480 T per complete block train) is used. The resulting calibration gives the following function $c_{jk}^1(\cdot)$ with constant terms a_{jk} :

$$c_{kl}^1 := a_{jk}^1 \left[\sum_{(i,l) \in \mathcal{W}} x_{ijkl} \right]^{0.5} \quad (13)$$

for each hub-hub pair $(j, k) \in \mathcal{H}^2$. An explicit form is obtained for the hub-to-destination discount function, c^2 , in an analogous manner, given as:

$$c_{kl}^2 := a_{kl}^2 \left[\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{H}} x_{ijkl} \right]^{0.6} \quad (14)$$

for each hub-destination node pair $(k, l) \in \mathcal{H} \times \mathcal{N}$. Note that the exponent of the discount function c^2 is higher than that of c^1 , corresponding to a lower potential for scale economies on these paths.

1.4 A linearized model

The concave increasing cost terms on the interhub and hub-to-destination portions of each itinerary are approximated by a piecewise-linear function so as to permit the use of linear programming solvers for its resolution. To this end, the concave functions $c_{jk}^1(\cdot)$ and $c_{kl}^2(\cdot)$ are divided each into a number M of equal unit pieces and the flows are normalized to this scale. Additional binary variables are introduced to represent the flow between each kink in the piecewise-linear approximation, defined as:

$$\begin{aligned} hh_{jkm} &= 1, \text{ if the } m^{\text{th}} \text{ portion of flow is captured by the hub shuttle service} \\ &\quad \text{between hubs } j \text{ and } k, \\ hd_{klm} &= 1, \text{ if the } m^{\text{th}} \text{ portion of flow is attained between hub } k \text{ and destination} \\ &\quad \text{terminal } l, \end{aligned}$$

and 0 otherwise, for all $(j, k) \in \mathcal{H}^2$ and for all $(k, l) \in \mathcal{H} \times \mathcal{N}$. Then, the slopes of each linear unit piece of the functions $c_{jk}^1(\cdot)$ and $c_{kl}^2(\cdot)$ can be expressed as $\phi_m^{hh} := m^{0.5} - (m-1)^{0.5}$ and $\phi_m^{hd} := m^{0.6} - (m-1)^{0.6}$, respectively.

The costs in the linearized objective function are then expressed as:

$$\begin{aligned} \Psi(x, hh, hd) &:= \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{H}} c_{ij} \sum_{k \in \mathcal{H}} \sum_{l \in \mathcal{N}} x_{ijkl} + \sum_{j \in \mathcal{H}} \sum_{k \in \mathcal{H}} a_{jk}^1 \sum_{m \in [1, M]} \phi_m^{hh} hh_{jkm} \\ &+ \sum_{k \in \mathcal{H}} \sum_{l \in \mathcal{N}} a_{kl}^2 \sum_{m \in [1, M]} \phi_m^{hd} hd_{klm} \end{aligned} \quad (15)$$

and then linearized objective function is then given by:

$$\min_{x, \hat{x}, hh, hd, z} \Psi(x, hh, hd) + \sum_{(i, l) \in \mathcal{W} \setminus \mathcal{H}} \hat{c}_{il} \hat{x}_{iil} + \sum_{j \in \mathcal{H}} f_j z_j. \quad (16)$$

The pieces of the piecewise linear approximations must be filled in order from the most costly marginal cost in the first piece, to the least costly in the last piece; this restriction is given by the following constraints for the interhub links:

$$hh_{jkm} \in \{0, 1\}, \quad j \in \mathcal{H}, \quad k \in \mathcal{H}, \quad m \in [1, M], \quad (17)$$

$$hh_{jkm} \geq hh_{jk, m+1}, \quad m \in [1, M-1], \quad j \in \mathcal{H}, \quad k \in \mathcal{H}, \quad (18)$$

so that $hh_{jk2} = 1$ only if $hh_{jk1} = 1$ and so on, and similarly for the hub-to-destination links:

$$hd_{klm} \in \{0, 1\}, \quad k \in \mathcal{H}, \quad l \in \mathcal{N}, \quad m \in [1, M], \quad (19)$$

$$hd_{klm} \geq hd_{kl, m+1}, \quad m \in [1, M-1], \quad k \in \mathcal{H}, \quad l \in \mathcal{N}. \quad (20)$$

The following definitional constraints link the flow variables with the binary approximation variables:

$$\sum_{(i, l) \in \mathcal{W}} x_{ijkl} - \gamma_{hh} \sum_{m \in [1, M]} hh_{jkm} = 0, \quad j \in \mathcal{H}, \quad k \in \mathcal{H}, \quad (21)$$

$$\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{H}} x_{ijkl} - \gamma_{hd} \sum_{m \in [1, M]} hd_{klm} = 0, \quad k \in \mathcal{H}, \quad l \in \mathcal{N}, \quad (22)$$

where the constants γ_{hh} and γ_{hd} are used to normalize flow values to the unit scale used in the piecewise-linear approximation.

The resulting linearized concave-cost hub location model is then given by objective function (16) and (15), subject to constraints (11),(3)–(7) and (17)–(22).

1.5 Polyhedral properties of the model

In this section, we provide some polyhedral properties of the above model. In Hamacher et al. (2000), the hub location polytope was studied by lifting facets from the polytope associated with the facility location problem. It was shown that a set of inequalities define facets for the hub location problem when the definition of the flow variable represents percentages of the OD demand. (That is, the formulation is such that the sum of the flows between each OD pair over all intermediate hub nodes is equal to one: $\sum_{(j,k) \in \mathcal{H}^2} x_{ijkl} = 1$ for each $(i,l) \in \mathcal{W}$.) This requires a normalization of the constraints (2) to (6) with respect to the OD demand d_{il} . To this end, let the constraints (2) be given instead by

$$\sum_{j \in \mathcal{H}} \sum_{k \in \mathcal{H}} x_{ijkl} = 1 \quad (23)$$

for each OD pair $(i,l) \in \mathcal{W}$ and remove the constant term Q_{il} from (3)–(4). Then the following result holds:

Proposition 2 *Let constraint (23) replace constraint (2) in the hub location model (1)–(7), where the constant Q is then no longer needed in the constraints (3)–(4). Let the number of OD pairs, $|\mathcal{W}| \geq 3$. Then the following inequalities are valid and define facets of the abovedefined polytope:*

$$x_{ijkl} \geq 0 \quad j, k \in \mathcal{H}, (i,l) \in \mathcal{W} \quad (24)$$

$$z_j \leq 1 \quad k \in \mathcal{H} \quad (25)$$

$$\sum_{j \in \mathcal{H}} x_{ijkl} + \sum_{j \in \mathcal{H} \setminus \{k\}} x_{ikjl} \leq z_k \quad k \in \mathcal{H}, (i,l) \in \mathcal{W}. \quad (26)$$

Proof: See Corollary 3.5 of Hamacher et al. (2000).

Note that this set of inequalities differs from the basic model, in that rather than dividing the hub opening constraints into two inequalities, that is, (3) and (4), a sum of the two terms is preferable, since it is tighter.

In the case of the our model, in which demands are explicitly included, we modify the constraints (26) as follows:

$$\sum_{j \in \mathcal{H}} x_{ijkl} + \sum_{j \in \mathcal{H} \setminus \{k\}} x_{ikjl} \leq d_{il} z_k \quad k \in \mathcal{H}, (i,l) \in \mathcal{W}. \quad (27)$$

Next, we show that (27) is both valid and facet-defining for the model (1), (11), (3)–(7).

Proposition 3 *The inequality (27) is valid and is a facet for (1), (11), (3)–(7).*

Proof: The first part of the proposition is straightforward, since the sum of the OD flow can never exceed the sum of the OD flow through the hub k and the OD flow sent on a direct route. The second follows as a direct extension of Proposition (2) above.

2 Solution method

In O’Kelly and Bryan (1998), the authors present a hub location model with piecewise-linear costs, and test the model on a 20-node network using application-specific linear programming software. Each piecewise-linear approximation to their concave cost function, however, has only three pieces. This affects, in particular, the threshold of efficiency of the interhub trip as compared to (in our case) the direct trip. In addition, fixed costs for hub opening are not included in the above reference.

In our case, the model makes use of from 25 to 100 linear pieces to approximate the nonlinear cost functions for each hub-to-hub and hub-to-destination pair, meaning in practice that the linear approximation is very close to the original nonlinear function. Although we can eliminate some of the possible hub-hub and hub-destination pairs by simple pre-processing techniques, the number of binary variables is on the order of 10,000. To solve the resulting model, we therefore develop an efficient heuristic method to solve the linearized problem.

In the case of linear hub location models, a classic resolution method consists in performing a Benders decomposition and solving independently the 0-1 programming problem over the binary variables z , which indicate whether or not a hub is to be opened, and the linear network flow problem for a fixed set z of open hubs. However, the first phase integer programming master problem is generally very difficult to solve if there is no particular structure to exploit.

Rather than decompose the problem in this fashion, we have chosen to make use of recent polyhedral properties of the linear hub location problem and solve the problem as a large mixed-integer program. Indeed, preliminary results have shown the cuts to

be quite efficient at removing non-integer solutions of the hub-location variables, z_i . For more information on this work, the reader is referred to [6] (or to a related work on the single allocation problem by Sohn and Park (2000) for the case in which the number of hub nodes is exactly three).

However, as opposed to the abovementioned references, our problem in the flow variables is still a very difficult problem to solve. This is due, of course, to the piecewise linear relaxation of the concave cost curves. Indeed, the number of binary variables associated with the segments of the piecewise linear curve is on the order of several thousands. In order to approximate exactly the concave cost curve, it is necessary to impose the constraint that the second piece of each piecewise-linear function is used (that is, $hh_{jk2} = 1$) only if the first piece is used (that is, $hh_{jk1} = 1$), and similarly for the third piece ($hh_{jk3} = 1$ only if $hh_{jk2} = 1$) and so on for all remaining pieces, and each of these variables (hh and hd) is binary.

To solve the model in practice, then, we have devised a variable-reduction heuristic which solves a sequence of relaxed subproblems, in which constraints (17) and (19) are replaced by

$$hh_{jkm} \in [0, 1], \quad j \in \mathcal{H}, \quad k \in \mathcal{H}, \quad m \in [1, M], \quad (28)$$

and

$$hd_{klm} \in [0, 1], \quad k \in \mathcal{H}, \quad l \in \mathcal{N}, \quad m \in [1, M]. \quad (29)$$

The following variable-reduction heuristic technique successively reduces the number of free variables, thereby forcing the initial pieces to 1, allowing only the final used piece to be fractional, as should be the case.

2.1 A variable-reduction heuristic: heuristic 1

1. Set $\hat{M}_{jk} := M$ and $\hat{M}_{kl} := M$ for each $(j, k) \in \mathcal{H}^2$ and each $(k, l) \in \mathcal{H} \times \mathcal{N}$.
2. Solve the (relaxed) subproblem with constraints (28) and (29) replacing (17) and (19).
3. For each arc $(j, k) \in \mathcal{H}^2$ with fractional values of hh_{jkm} ,
 - (a) Calculate the total amount of flow assigned to the arc. That is, obtain

$$y_{jk} := \sum_{m \in [1, \hat{M}_{jk}]} \{hh_{jkm} \mid hh_{jkm} > \epsilon\}$$

for $0 < \epsilon \ll 1$.

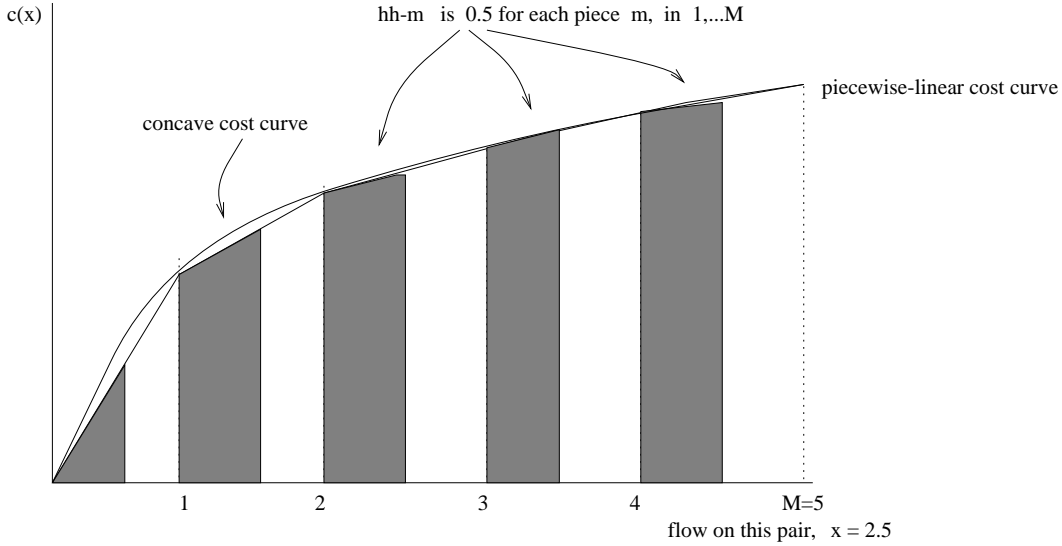


Figure 3: Before variables are fixed, the (fractional) flow on each piece is equal

- (b) Set $hh_{jk(\bar{y}_{jk}+i)} = 0$ for $i = 2, \dots, \hat{M}_{jk} - \bar{y}_{jk}$, where $\bar{y}_{jk} := y_{jk} - (y_{jk} \bmod 1)$ is the integer part of the assigned flow on arc (j, k) . Set $\hat{M}_{jk} = \bar{y}_{jk} + 1$.
4. Goto step 2 and perform the same procedure for hd_{klm} , for each $(k, l) \in \mathcal{H} \times \mathcal{N}$.
5. Check whether either all $hh_{jk1} \in \{0, 1\}$ and $hd_{kl1} \in \{0, 1\}$ or $hh_{jk1} \in (0, 1)$ and $hh_{jk2} = 0$ [resp. $hd_{kl1} \in (0, 1)$ and $hd_{kl2} = 0$]. Otherwise, return to Step 2.

A few remarks are in order.

Remark 1 *In Step 3(a), the portions of the piecewise-linear curve that should have been unused in the case of binary hh (resp. hd) are forced to zero. The process is then repeated, thereby eliminating at each iteration the number of free variables in the relaxed subproblem, that is, \hat{M} decreases at each iteration. That is, at each iteration of the heuristic, there is a reduction in the number of variables in the problem.*

One iteration of the heuristic is illustrated on the example given in Figure 3 in the Figure 4. Since the total flow on the pair is equal to 2.5, and should therefore use the first three (unit-length) pieces only, the remaining pieces, 4 and 5 in this case, are fixed and removed from the problem in the next iteration.

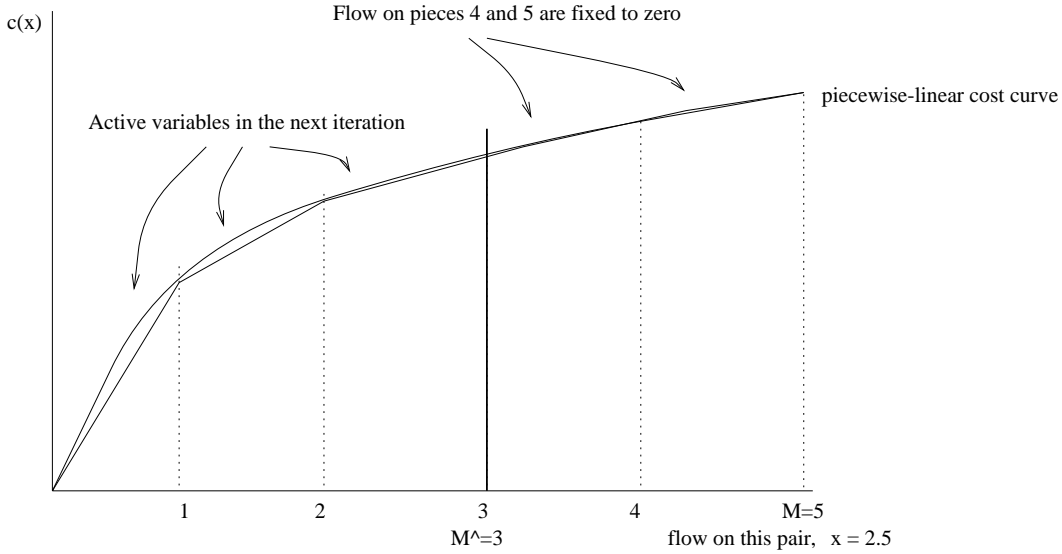


Figure 4: Illustration of heuristic 1 on the example of Figure 3

Remark 2 Note that, in the first iteration, if any $hh_{jk_m} > \epsilon$ and fractional, then all pieces satisfy precisely $hh_{jk_1} = hh_{jk_2} = \dots = hh_{jk_{\hat{M}}}$ due to the constraints (18) and (20). (See Figure 3 for an illustration.) Therefore, setting the last piece to zero has the effect of “forcing” the higher (marginal) cost pieces to be used, or shifting the flow to a different path. This characteristic holds for the remaining iterations, with the index \hat{M} reduced at each iteration for which there are fractional flows.

Remark 3 As a stopping criterion, it is thus sufficient to check the value of pieces $m = 1$ and $m = 2$ only, since either each piece $i \in \hat{M}^t$ (where \hat{M}^t is the number of free variables remaining at iteration t) has an equal fractional value, or else only the last piece is fractional.

For example, if one considers Figure 3 to represent the output of the relaxed subproblem resolution at some iteration, then Step 3 of the variable-reduction heuristic would add the following constraints (letting the decision variable for each piece be expressed as hh_m): $hh_m = 0$ for $m \in [4, 5]$, since the total flow on the path is 2.5, and the piecewise curve is defined by 5 pieces. In other words, variables hh_4 and hh_5 are removed from the problem in subsequent iterations.

2.2 A faster variable-reduction heuristic: heuristic 2

The following heuristic speeds up the variable reduction process of Step 3 in heuristic 1. In particular, in Step 3(c), the additional fractional pieces are fixed to at most the value of the fractional part of the total flow present on the route. (In Figure 3, then, the piece $m = 3$ would have the following constraint added to those defined by heuristic 1: $hh_3 \leq 0.5$.)

1. Set $\hat{M}_{jk} := M$ and $\hat{M}_{kl} := M$ for each $(j, k) \in \mathcal{H}^2$ and each $(k, l) \in \mathcal{H} \times \mathcal{N}$.
2. Solve the (relaxed) subproblem with constraints (28) and (29) replacing (17) and (19).
3. For each arc $(j, k) \in \mathcal{H}^2$ with fractional values of hh_{jkm} ,

(a) Calculate the total amount of flow assigned to the arc. That is, obtain

$$y_{jk} := \sum_{m \in [1, M]} \{hh_{jkm} \mid hh_{jkm} > \epsilon\}$$

for $0 < \epsilon \ll 1$.

(b) Set $hh_{jk(\bar{y}_{jk}+i)} = 0$ for $i = 2, \dots, \hat{M}_{jk} - \bar{y}_{jk}$, where $\bar{y}_{jk} := y_{jk} - (y_{jk} \bmod 1)$ is the integer part of the assigned flow on arc (j, k) . Set $\hat{M}_{jk} = \bar{y}_{jk} + 1$.

(c) Set $hh_{jk(\bar{y}_{jk}+1)} \leq y_{jk} \bmod 1$.

4. Goto step 2 and perform the same procedure for hd_{klm} , for each $(k, l) \in \mathcal{H} \times \mathcal{N}$.
5. Check whether either all $hh_{jk1} \in \{0, 1\}$ and $hd_{kl1} \in \{0, 1\}$ or $hh_{jk1} \in (0, 1)$ and $hh_{jk2} = 0$ [resp. $hd_{kl1} \in (0, 1)$ and $hd_{kl2} = 0$]. Otherwise, return to Step 2.

Heuristic 2 should therefore converge faster, since the problem is more constrained at each iteration, the price to pay being naturally that the local solution obtained may be of inferior quality.

3 Computational experience and analysis

The numerical experiments of the hub location model were carried out on the data set described in the Appendix. The set of terminals used was obtained from the European Freight Railway Network (EUFNET) study [7] by selecting the 32 terminals

# hubs, # nodes, # pieces	Heuristic 1	Heuristic 2	Exact solution
3-4-25	8277 (0.3%)	9744 (18.1%)	8252
3-5-25	25,273 (3.9%)	28,223 (16.1%)	24,315
3-6-25	37,899 (4.9%)	39,610 (9.6%)	36,138
4-5-15	23,352 (4.8%)	26,024 (16.8%)	22,274
4-6-15	38,130	42,337	–
5-6-15	36,249	39,649	–

Table 1: Validation of the Variable-reduction heuristics

in the Alpine region with the largest international intermodal freight flows. Potential mega-hub nodes were those with the highest *gravity* potential to attract freight consolidation. Flows between all terminals were obtained from the NEAC data set for combined transport traffic.

First, however, the two variable-reduction heuristics are validated on a subset of the data.

3.1 Validation of the variable-reduction heuristics

In this section, the two variable-reduction heuristics are tested on a subset of the data small enough to permit an exact resolution of the linearized subproblem by branch and bound. The Table 1 provides the results of these tests. The first column of the table provides the characteristics of the data set, in terms of number of possible hub nodes, total number of nodes (where each pair of nodes has a non-zero demand to every other pair), and the number of segments used in the approximation of the concave cost curve. The integer programming solver of CPLEX (version 6) was used to obtain the exact solution. Similarly, the CPLEX linear programming solver was used to solve the linear subproblems in the two heuristics.

Note that, for test sets of more than three hub nodes, the integer programming solver was unable to provide a solution, even with the number of segments reduced to 15 for each concave cost curve. As concerns the quality of the two heuristics, the first heuristic is clearly superior to the second, more constrained heuristic. The first heuristic provides solutions within five percent of the optimal value, for those test sets for which an optimal value was obtained. On the other hand, the second, more constrained heuristic provided, on average, solutions on the order of 15% more costly than the optimal solution. Since the computation time of the both heuristics is very reasonable (less than 0.5 seconds for each of the test sets above), the first

Costs, f_j	$\{i : z_i = 1\}$	$\{i : z_i = 0\}$	$I := \{i : z_i \in (0, 1)\}, \mu(z_{\bar{I}})$
random 1	0,1,3,4,5,6,7,8,9,10,11,13	–	$ I = 2, \mu = 0.04$
random 2	0,1,3,11,13	5,6	$7, \mu = 0.05$
high est.	0	12	$12, \mu = 0.16$
med. est.	0,6,11	13	$10, \mu = 0.08$
low est.	0, 5, 6, 11	10, 13	$8, \mu = 0.07$

Table 2: Results on the full-size data set

heuristic is clearly preferable. It is this heuristic that was therefore used in the code for solving the overall hub location problem. Numerical results on the overall problem are presented below.

3.2 Numerical results on the full data set

The purpose of the model developed is to permit studying the potential of a hub-and-spoke type network of intermodal freight terminals in Europe. One key corridor for freight transport improvement in Western Europe is the Alpine crossing. The numerical results presented in this section are taken from the application of the model to this important corridor.

The number of nodes in the graph is 32, representing the major sources of emission and reception of intermodal freight traffic in Europe. Of these, 14 terminals were selected as potential *mega-hub* nodes; that is nodes which are candidates for construction of efficient container swapping equipment and fixed-composition freight shuttle service. All tests in this section were therefore run on this full set of 32 nodes, 14 potential hub nodes, and 100 linear pieces for each of the concave cost curves.

The CPLEX linear programming solver was used to solve the overall problem along with variable-reduction heuristic 1. The addition of the constraint (27) was sufficient in eliminating non-integer solutions on very small data sets, on the order of those presented in Table 1.

To test the model and solution method on the full-size data set, initially, two sets of random hub opening costs were generated. The results are provided in Table 2. The first column provides the type of hub-opening costs used in the test set. The second column gives the number of hubs with values z_i identically 1, and then those with values equal to 0. The last column provides the number of hubs with fractional values of z_i and the average fractional value. The main thing to observe is that, when

fractional values were present in the results, the values were very small, particularly in the tests with randomly generated hub opening costs. In the last three tests, realistic hub-opening costs were used (with high, medium, and low estimates of the costs). The quality of the polyhedral information good, but could be stronger. This is seen from the fact that the average value of the non-integer results is significantly higher than in the first two results, particularly in the test set *high est.*. Nevertheless, if one takes two averages in that set, the first over the non-integer values of z_j such that $z_j \in [0.5, 1)$ (call it $\mu_{>}$) and the second over the z_j for which $z_j \in (0, 0.5)$ (call it $\mu_{<}$), we obtain $\mu_{>} = 0.85$ (with only one element in the set) and $\mu_{<} = 0.07$; that is, the non-integer values are in fact quite close to 0 and 1. This is important since for values of $z_j < 0.08$, a single branch and bound node often provides integer solutions.

Qualitative results with the model show that hubs 0, 6, and 11 (Munich, Verona, and Mannheim) are of great importance in reducing transportation costs through consolidation across the Transalpine region, and hub 5 of importance when the opening costs are sufficiently low. Indeed, the mega-hub at Munich is shown to consolidate freight flows on the heavily used North-South Alpine corridor, that is to Verona and Naples, in particular, as well as from North-east Europe to West Europe. Furthermore, the hub at Munich is shown to consolidate flow to other German cities such as Hamburg and Nuremburg. The shuttle service on the corridor Mannheim–Milan (hub 11 – hub 5) consolidates significant flow from north to south, as well flow with destinations in Spain and France. The pair Milan–Verona (hubs 5, 6) is important in consolidating flow across Italy, such as to Genoa and Bologna. Further analysis of the model results are available in the INRETS Final Report of the project IQ [7].

4 Conclusions

We have presented the novel application of locating the optimal configuration of intermodal freight transport hubs and obtaining their usage levels. The model we propose for this application is based on the uncapacitated hub location problem. We further add to this model an accurate representation of the economies of scale due to consolidation; this is accomplished through explicit use of concave cost functions for the interhub (and hub-to-destination) portions of each trip. In order to solve in practice this more complex model, we propose two heuristics for solving a piecewise approximation of the nonlinear, concave cost curves that permit handling even very large problems quickly. The efficiency of the heuristics is such that the piecewise linear approximation need not lose much; indeed the number of pieces we consider for each curve is on the order of 25 to 100, and problems having 30 nodes are still

easily solvable. Comparisons with exact solutions on smaller test sets show that the percentage deviation of the heuristic is within five percent of optimal. The use of recent results on polyhedral properties of the model enable us to obtain quasi-integer solutions with a mixed-integer formulation of the problem, with better results on small-scale problems. Interesting extensions of this work could involve the comparison of the proposed heuristic with other techniques for handling concave costs, both linear programming-based as well as nonlinear programming techniques. The results have shown as well that further work on the polyhedral properties of the problem would be of benefit on problems of medium and large size.

Acknowledgments

This work was supported in part by the European Union DG-7 projects IQ, and SCENES. Data was supplied by the Department of Transport Economics (DEST), INRETS, Arcueil, France. The second author thanks Tim Sonneborn of ITWM, Kaiserslautern, for valuable comments.

References

- [1] Bryan, D.L. and O’Kelly, M.E. (1999) Hub-and-spoke networks in air transportation: an analytical review, *Journal of Regional Science* **39**(2) 275–295.
- [2] Campbell, J.F. (1994) Integer programming formulations of discrete hub location problems, *European Journal of Operations Research* **72**, 387–405.
- [3] Campbell, J.F. (1996) Hub location and the p-hub median problem, *Operations Research* **72**, 387–405.
- [4] Crainic, T.G. (2000) Service network design in freight transportation , *European Journal of Operations Research* **122**, 272–288.
- [5] Ernst, A.T. and Krishnamoorthy, M. (1998) Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem, *European Journal of Operations Research* **104**, 100–112.
- [6] Hamacher., H. W., Labbe, M., Nickel, S., and Sonneborn, T., (2000) Polyhedral properties of the uncapacitated multiple allocation hub location problem, Institut

fuer Techno- und Wirtschaftsmathematik (ITWM), Technical Report 20, 2000, available at <http://www.itwm.fhg.de/zentral/berichte/bericht20.html>

- [7] INRETS, final reports on the EU DG-7 projects EUFRANET, IQ, and SCENES. (2000) Reports of the Département d’Economie et Sociologie des Transports, available at
<http://www.inrets.fr:80/ur/dest/eufranet.htm>,
<http://www.inrets.fr:80/ur/dest/iq.htm>,
<http://www.inrets.fr:80/ur/dest/scenes.htm>, respectively.
- [8] Klincewicz, J.G. (1996) A dual algorithm for the uncapacitated hub location problem. *Location Science* **4**, 173–184.
- [9] Nickel S., Schoebel, A., and Sonneborn, T. Hub location problems in urban traffic networks, in Niittymaki and Pursula, editors, *Mathematical Methods and Optimization in Transportation Systems*, Kluwer Academic Publishers, to appear.
- [10] O’Kelly, M.E. and Bryan, D. (1998) Hub location with flow economies of scale, *Transportation Research-B* **32**(8), 605–616.
- [11] O’Kelly, M. Skorin-Kapov, D. and Skorin-Kapov, J. (1995) Lower bounds for the hub location problem. *Management Science* **41**(4), 713–721.
- [12] Skorin-Kapov, D., Skorin-Kapov, J., and O’Kelly, M. (1996) Tight linear programming relaxations of uncapacitated p-hub median problems, *European Journal of Operations Research* **94**, 582–593.
- [13] Sohn, J. and Park, S. (2000) The single allocation problem in the interacting three-hub network, *Networks* **35**(1), 17–25.

Appendix

The set of terminals used within this study was obtained from the European Freight Railway Network (EUFRANET) study [7] by selecting the 32 terminals in the Alpine region with the largest international intermodal freight flows. Potential mega-hub nodes were those with the highest *gravity* potential to attract freight consolidation. The resulting terminals, their status as a potential hub or not, and their code names are listed below. The demand data and the distance matrix were obtained from the EUFRANET study. The demand data is based upon the combined transport estimates from the annual NEAC-2 1992 origin-destination matrix. The flows are

expressed in tonnes of combined freight transport per year between each pair of terminals. The figures were divided by 50 to obtain weekly estimates. Cost data was compiled by INRETS.

Number	City	Status
0	MUNCHEN	Potential mega-hub
1	LYON	Potential mega-hub
2	MARSELLE	Potential mega-hub
3	TORINO	Potential mega-hub
4	GENOVA	Potential mega-hub
5	MILANO	Potential mega-hub
6	VERONA	Potential mega-hub
7	BOLOGNA	Potential mega-hub
8	KÖLN	Potential mega-hub
9	METZ	Potential mega-hub
10	BASEL	Potential mega-hub
11	MANNHEIM	Potential mega-hub
12	STRASBOURG	Potential mega-hub
13	STUTT GART	Potential mega-hub
14	FREDERICIA	Non hub terminal
15	NAPOLI	Non hub terminal
16	NURNBERG	Non hub terminal
17	BERLIN	Non hub terminal
18	HAMBURG	Non hub terminal
19	BRAUNSCHWEIG	Non hub terminal
20	ROMA	Non hub terminal
21	BARCELONA	Non hub terminal
22	NOISY LE SEC (PARIS)	Non hub terminal
23	DIJON	Non hub terminal
24	LAUSANNE	Non hub terminal
25	MECHELEN	Non hub terminal
26	SOPRON	Non hub terminal
27	WIEN	Non hub terminal
28	DONCASTER	Non hub terminal
29	CHIASSO	Non hub terminal
30	WARSAWA CENTRALNA	Non hub terminal
31	PRAHA LIBEN	Non hub terminal