Some Numerical Studies for a Complicated Hub Location Problem

J. Fabian Meier and Uwe Clausen

Abstract We consider a complicated hub location problem which includes multi-allocation, different hub sizes and different transport volumes on different week days. Furthermore, we consider transport costs per vehicle and not per volume which transforms the cost function into a step function and makes the problem numerically very hard. In our previous work we developed a heuristic approach which we now want to compare to CPLEX results for general and simplified models.

1 Introduction

Hub location problems have become classic challenges in the area of discrete optimization. The original problems, as they are very well described in [1], use a graph of depots which are connected by transport arcs. The task is to transport a given set of shipments from their sources to their sinks in a cost-optimal way. For that, some depots are equipped as *hubs*; then one assigns to every shipment a path from its source to its sink using only hubs in between. The total cost is the sum of the transport costs (for every shipment on every arc it uses) and the costs for the hubs (some problems require a fixed number of hubs p which is equivalent to assigning zero cost to the first p hubs and infinite costs to the following ones).

The main idea of hub location problems is *economies of scale*: Bundling shipments usually decreases the unit transport costs and may hence be beneficial even if it requires detours and costly facilities. Classic hub location problems usually assume

Supported by Deutsche Forschungsgemeinschaft, Project *Lenkung des Gueterflusses in durch Gateways gekoppelten Logistik-Service-Netzwerken mittels quadratischer Optimierung.*

J.F. Meier (✉)
Institut Fuer Transportlogistik, TU Dortmund, Leonhard-Euler-Str. 2, 44227 Dortmund, Germany
e-mail: meier@itl.tu-dortmund.de

U. Clausen
Institut Fuer Transportlogistik, Fraunhofer-Institut Fuer Materialfluss Und Logistik, Joseph-von-Fraunhofer-Str. 2-4, 44227 Dortmund, Germany
e-mail: clausen@itl.tu-dortmund.de

H.-J. Sebastian et al. (eds.), *Quantitative Approaches in Logistics and Supply Chain Management*, Lecture Notes in Logistics,
DOI 10.1007/978-3-319-12856-6_2

fixed unit costs for each transport arc; to simulate economies of scale they reduce the unit costs on hub-hub-connections by a factor α because "usually" these arcs carry more overall weight. This simplification increases the solvability but also limits the applicability of the model.

We investigate the problem from the point of view of a less-than-truckload network planner. A large number of small shipments has to be transported from their source depots to their sink depots. A vehicle can transport many of these shipments, so that it is advantageous to bundle shipments for transport: Instead of direct transport we establish hubs as transhipment points. We ask the strategic questions:

- Where should hubs be established?
- What transhipment capacities should be assigned to them?

To give a cost-efficient answer to such questions, we have to balance the strategic costs of establishing transhipment capacities with the prospective tactical/operational costs of the transport. Our model incorporates the following challenges:

- *Truck-based transport costs*. If we send a truck from A to B, the resulting cost depends on general vehicle costs, driver salary and fuel consumption. The filling quota of the truck has little influence on the fuel consumption and nearly no influence on the other terms. Thus we obtain a good approximation of the real costs if we measure transport costs "by vehicle" instead of "by volume" [2]. On each connection $A \rightarrow B$, the cost per volume becomes a step function. Such vehicle based costs were already considered in the mathematical models of [8, 9].
- *Multi-allocation*. Every shipment can be independently routed, so that each depot can be connected to many others. If it turns out to be cheaper to have some direct transports, this is also possible.
- *A weekday-based schedule*. We consider a European network where the travel time between two depots/hubs is one to four days. As our shipping volumes depend on the day of the week, we use a cyclic model with five time slices to represent the working days.
- *Variable hub sizes*. We assign a transhipment to every possible hub. It can be chosen on a continuous scale. Hubs of different capacities were considered by [7], but our weekly schedule adds an additional flavour: Strategic decisions have to be equal for every day of the week, i.e. the transhipment capacity of a hub is the same on every week day.
- *Buffering*. We want to analyse the effect of buffering, i.e. the possibility of storing a shipment in a hub for a day to get a cheaper transport on the next day. Therefore, we consider a scenario with a separate buffering capacity for every hub, which can also be chosen on a continuous scale at the strategic level. Buffering is considered as a "transport in time" lasting one day. In principle, buffering actions can be chained, but as our real world instances have strict transport time limits this is usually not possible.

Section 2 will define and discuss mathematical models for the strategic planning problem that was just outlined. These have a huge number of binary variables, so that we define restricted models in Sect. 3 that are easier to solve. Section 4 briefly describes our heuristic approach which is detailed in [3]. In Sect. 5, we state and discuss the numerical results of both the MIP approaches and the heuristic approach. A short conclusion completes the paper.

2 The Mathematical Models

We start with a given set of D of *depots*. The depots are all interconnected by *transport connections*. A transport connection can be used by shipments to get from one depot to another; it has a *travel time* (in days) and *cost factor* which is the cost per vehicle on this connection. We consider a homogeneous fleet as one usually uses the maximal allowed truck size on European connections. The situation can be depicted by a directed graph: The nodes are formed by the (depot, weekday) pairs, and a transport connection from depot A to depot B which needs n days connects (depotA, d) with (depotB, $d + n \pmod 5$)) for each weekday d (shown in Fig. 1).

Furthermore we have a large number of shipments. These shipments all have a source depot, a sink depot, a volume and a maximal travel time. The routes can use every depot for transhipment or buffering which is equipped with the appropriate hub capacity. The transhipment capacities of the hubs have to be chosen large enough to work for *every* weekday, i.e. they need to handle the maximal transhipment that happens throughout the week. A route can consist of arbitrary many steps as long as the maximal travel time is not exceeded.

There are two general approaches to model the routing of the shipments: A *route-based* or a *flow-based* view.

In the route-based view, each possible route for a shipment is represented by a binary variable from which exactly one has the value 1. Without any restrictions on

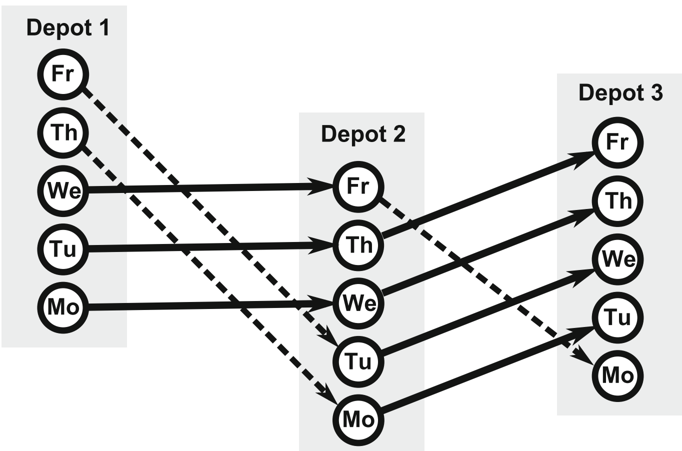

Fig. 1 The time expanded network: each arc represents a movement in space and (cyclic) time

the possible paths, the number of variables is exponential in the number of edges of the graph. Hence this kind of model can only be sensibly applied when we have strong restrictions on the number of possible paths (like in our paper [5] or in classical hub location problems [1], where at most three edges per path were allowed) or apply an approach like column generation. As column generation failed to give good results in a simpler model with truck-based costs [6], we will not use this approach in the general case, but we will come back to it in the next section to build restricted models.

The flow-based view is inspired by the multi-commodity flow problem, but with the major difference that shipments cannot be split, so that we need a binary variable for each (shipment, edge) pair, stating if the edge is used by the shipment. As the number of commodities and the number of edges are each of order $(\#\text{depots})^2 \cdot (\#\text{days per week})$, we get approximately $(\#\text{depots})^4 \cdot (\#\text{days per week})^2$ binary variables. Due to the maximal travel time of each shipment, some of these variables can be set to zero in preprocessing.

Firstly, we will construct a mathematical model without shipment buffering, then we will add this feature later. Let D be the set of depots and $W = \{0, 1, 2, 3, 4\}$ be the set of week days. Furthermore we have a set Q of shipments. For every $q \in Q$ we denote by $q_{\text{so}} \in D$, $q_{\text{si}} \in D$, $q_{\text{day}} \in W$, am_q and time_q the source, sink, starting day, amount and maximal travel time respectively.

The most important variable is the binary flow variable f_{qabw}. It states whether the shipment $q \in Q$ uses the arc $a \to b$, $a \neq b$ with starting day w. To form a flow it has to fulfill the following three conditions (tr_{ab} is the number of days to travel from a to b):

$$\sum_{d \in D, d \neq q_{\text{so}}} f_{q q_{\text{so}} d q_{\text{day}}} = 1 \qquad q \in Q \tag{1}$$

$$\sum_{d \in D, d \neq q_{\text{si}}, w \in W} f_{q d q_{\text{si}} w} = 1 \qquad q \in Q \tag{2}$$

$$\sum_{a \in D, a \neq d} f_{qad(w - \text{tr}_{ad} \bmod 5)} = \sum_{b \in D, b \neq d} f_{qdbw}$$
$$q \in Q, d \in D, w \in W, \quad d \neq q_{\text{si}}, \ d = q_{\text{so}} \Rightarrow w \neq q_{\text{day}} \tag{3}$$

Equation (1) states that each shipment leaves its source depot on the respective day, while Eq. (2) indicates that each shipment has to reach the sink depot (on an arbitrary week day). Equation (3) matches the flows for any other depot and time. As the last parameter of f stands for the starting day, we have to reduce w on the left hand side. To model the maximal travel time of each shipment we add the times of all used edges (4):

$$\sum_{a \neq b \in D, w \in W} f_{qabw} \cdot \text{tr}_{ab} \leq \text{time}_q \qquad q \in Q \tag{4}$$

Let us introduce some auxiliary variables: We define t_{abw} to be the transport volume on the edge $a \to b$ starting on day w, v_{abw} the number of vehicles on that edge, u_{dw} the total transhipment at depot d on day w and $u_d^{\max}$ the maximum over w of u_{dw}. These variables are defined by the Eqs. (5–8) (size is the size of a vehicle in units of volume):

$$t_{abw} = \sum_{q \in Q} \mathrm{am}_q \cdot f_{qabw} \qquad a \neq b \in D, w \in W \tag{5}$$

$$v_{abw} \cdot \mathrm{size} \geq t_{abw} \qquad a \neq b \in D, w \in W \tag{6}$$

$$u_{dw} = \sum_{b \in D} t_{dbw} - \sum_{q \in Q: q_{\mathrm{so}=d}, q_{\mathrm{day}}=w} \mathrm{am}_q \qquad d \in D, w \in W \tag{7}$$

$$u_d^{\max} \geq u_{dw} \qquad d \in D, w \in W \tag{8}$$

Using the parameters $\mathrm{truckcost}_{ab}$ for the costs of using a truck on connection $a \to b$ and $\mathrm{transcost}_d$ for the strategic costs of having transhipment capacity, we can state the objective function as:

$$\sum_{a,b \in D, w \in W} v_{abw} \cdot \mathrm{truckcost}_{ab} + \sum_{d \in D} u_d^{\max} \cdot \mathrm{transcost}_d \tag{9}$$

Let us call this problem *Multi-Allocation Weekday Scheduled Strategic Planning Problem* MAWSSPP. To get the buffering version BMAWSSPP, we need to introduce the possibility to store a commodity in a hub for a day. For this, we use the flow variables f_{qddw} which describe a "transport" from d to itself lasting one day. The buffering is also a strategic cost which is charged similarly to the transshipment cost (but with the factor buffcost). For that b_{dw} and $b_d^{\max}$ are analogously defined to u_{dw} and $u_d^{\max}$. We write:

$$b_{dw} = \sum_{q \in Q} f_{qddw} \cdot \mathrm{am}_q \qquad d \in D, w \in W \tag{10}$$

$$b_d^{\max} \geq b_{dw} \qquad d \in D, w \in W \tag{11}$$

The constraints (1–5) are transformed to:

$$\sum_{d \in D} f_{q q_{\mathrm{so}} d q_{\mathrm{day}}} = 1 \qquad q \in Q \tag{12}$$

$$\sum_{d \in D, w \in W} f_{q d q_{\mathrm{si}} w} = 1 \qquad q \in Q \tag{13}$$

$$\sum_{a \in D} f_{qad(w - \mathrm{tr}_{ad} \bmod 5)} = \sum_{b \in D} f_{qdbw}$$

$$q \in Q, d \in D, w \in W, \quad d \neq q_{\text{si}}, \; d = q_{\text{so}} \Rightarrow w \neq q_{\text{day}} \tag{14}$$

$$\sum_{a \neq b \in D, w \in W} f_{qabw} \cdot \text{tr}_{ab} + \sum_{d \in D, w \in W} f_{qddw} \leq \text{time}_q \qquad q \in Q \tag{15}$$

$$t_{abw} = \sum_{q \in Q} \text{am}_q \cdot f_{qabw} \qquad a, b \in D, w \in W \tag{16}$$

These five equations only differ from their counterparts by usage of buffering flows f_{qaaw}. For (15) we added a term adding one day for every buffering. We note that the (following) Eqs. (17) and (19) are unchanged, while (18) gets an additional term: Without it, buffered shipments would be charged twice for transhipment, but we chose only to charge incoming shipments.

$$v_{abw} \cdot \text{size} \geq t_{abw} \qquad a \neq b \in D, w \in W \tag{17}$$

$$u_{dw} = \sum_{b \in D} t_{dbw} - \sum_{q \in Q: q_{\text{so}=d}, q_{\text{day}}=w} \text{am}_q - b_{d(w-1 \bmod 5)} \qquad d \in D, w \in W \tag{18}$$

$$u_d^{\max} \geq u_{dw} \qquad d \in D, w \in W \tag{19}$$

The cost function is extended by an extra term:

$$\sum_{a,b \in D, w \in W} v_{abw} \cdot \text{truckcost}_{ab} + \sum_{d \in D} u_d^{\max} \cdot \text{transcost}_d + \sum_{d \in D} b_d^{\max} \cdot \text{buffcost}_d \tag{20}$$

Let us note that our modelling of the buffering feature includes the possibility of chaining buffering edges which means buffering a shipment for more than one day. We see no theoretic reasons for stronger constraints, but in practice the maximal transport time restrictions often exclude long buffering.

One can improve the solvability of the problem by discarding some variables in preprocessing. The largest number of eliminated variables can normally be achieved by the following argument:

$$f_{qabw} = 1 \quad \Rightarrow \quad \text{tr}_{q_{\text{so}}a} + \text{tr}_{ab} + \text{tr}_{bq_{\text{si}}} \leq \text{time}_q, \tag{21}$$

if we make the reasonable assumption that transport times fulfill the triangle inequality. Variables not fulfilling condition (21) can hence be eliminated from the equations.

3 Restricted Mathematical Models

A way to improve solvability of the MAWSSPP is to drastically reduce the number of involved binary variables. We want to define restricted models whose solutions are still valid solutions for MAWSSPP (and hence also for BMAWSSPP). We consider two approaches:

1. We consider the same transport plan for every day (SAMEDAY). By this, the number of routes that have to be assigned is reduced by a factor of five. Furthermore, it reflects the reality in many non-automatized settings.
2. We allow at most one hub on each route (ONEHUB). This leads to a massive reduction in the number of binary variables (detailed below).

Let us first discuss the elimination of weekdays. Until now, we assumed one shipment for every (source depot, sink depot, week day) triple, possibly of size zero. Now we define a "super shipment" for every pair (source depot, sink depot) which has the maximal size of all five attached shipments. We call the set of super shipments Q^s and furthermore reuse the variables f, v, t and u, which are now time independent (we can thus dispense with $u^{\max}$). Solving the routing problem for these super shipments (with quintupled costs) automatically gives a solution for the original problem. The model now looks like this:

$$\sum_{d \in D, d \neq q_{\text{so}}} f_{q q_{\text{so}} d} = 1 \quad q \in Q^s \tag{22}$$

$$\sum_{d \in D, d \neq q_{\text{si}}} f_{q d q_{\text{si}}} = 1 \quad q \in Q^s \tag{23}$$

$$\sum_{a \in D, a \neq d} f_{qad} = \sum_{b \in D, b \neq d} f_{qdb} \quad d \in D, d \neq q_{\text{si}}, d \neq q_{\text{so}} \tag{24}$$

$$\sum_{a \neq b \in D} f_{qab} \cdot \text{tr}_{ab} \leq \text{time}_q \quad q \in Q^s \tag{25}$$

$$t_{ab} = \sum_{q \in Q^s} \text{am}_q f_{qab} \quad a \neq b \in D \tag{26}$$

$$v_{ab} \cdot \text{size} \geq t_{ab} \quad a \neq b \in D \tag{27}$$

$$u_d = \sum_{b \in D} t_{db} - \sum_{q \in Q^s : q_{\text{so}} = d} \text{am}_d \quad d \in D \tag{28}$$

$$\min \quad 5 \cdot \left(\sum_{a,b \in D} v_{ab} \cdot \text{truckcost}_{ab} + \sum_{d \in D} u_d \cdot \text{transcost}_d \right) \tag{29}$$

For the ONEHUB model, we opted for a route-based-approach, because we can now easily describe the routing of a shipment by giving the intermediate hub d as

r_{qd}. A direct routing can be represented by $d = q_{\text{si}}$ or $d = q_{\text{so}}$. In this way we reduce the number of binary variables from about (#days per week)^2 · (#depots)^4 to approximately (#days per week) · (#depots)^3.

We keep all the other variables except for f. Hence, we can leave the objective function unchanged. Essentially, we have to make four changes:

- Delete the flow conditions (1–4).
- The transport volume t_{abw} is now calculated as:

$$t_{abw} = \sum_{\substack{q \in Q,\\ q_{\text{so}}=a,\\ q_{\text{day}}=w}} r_{qb} \cdot \text{am}_q + \sum_{\substack{q \in Q,\\ q_{\text{si}}=b,\\ q_{\text{day}}=w-\text{tr}_{da}}} r_{qa} \cdot \text{am}_q \quad a \neq b \in D, w \in W \tag{30}$$

- We have to ensure that for every commodity exactly one route is chosen:

$$\sum_{d \in D} r_{qd} = 1 \quad q \in Q \tag{31}$$

- The maximal travel time constraint has to be rewritten:

$$\sum_{d \in D} r_{qd} \cdot \left(\text{tr}_{q_{\text{so}}d} + \text{tr}_{dq_{\text{si}}}\right) \leq \text{time}_q \quad q \in Q \tag{32}$$

In our paper [4, Sect. 4] we considered additional inequalities to strengthen the formulation, which were not very successful in the two-hub-case. Following Martin Baumung's good unpublished results for the one-hub-case, we reconsider them. The idea is that we can calculate the minimal flow from a subset $K \subset D$ to $D \backslash K$ as

$$\text{minflow}\left(K, D \backslash K\right) = \sum_{q \in Q, q_{\text{so}} \in K, q_{\text{si}} \in D \backslash K} \text{am}_q \tag{33}$$

From that we know that the number of vehicles going from K to $D \backslash K$ is at least $\lceil \text{minflow}(K,\ D \backslash K)/\text{size} \rceil$. Due to the rounding up procedure, this bound is stronger than the original LP bound. To avoid adding huge numbers of inequalities, we consider this only for $|K| = 1$ and $|K| = |D| - 1$. In the first case, we can even consider the outgoing flow of every day separately. In the end, we get:

$$\sum_{b \in D} v_{abw} \geq \left\lceil \Bigg(\sum_{q \in Q: q_{\text{so}}=a, q_{\text{day}}=w} \text{am}_q \Bigg) / \text{size} \right\rceil \quad a \in D, w \in W \tag{34}$$

$$\sum_{a \in D, w \in W} v_{abw} \geq \left\lceil \Bigg(\sum_{q \in Q: q_{\text{si}}=b} \text{am}_q \Bigg) / \text{size} \right\rceil \quad a \in D \tag{35}$$

4 A Short Description of the Heuristic Approach from [3]

In [3] we developed a general modelling language for shipment based transport problems, i.e. problems that consist of a large number of (unsplittable) shipments which have to be transported through a graph. It is based upon the following paradigms:

1. The routes of the shipments are considered as independent variables, while all other variables (like number of trucks, transhipment capacities, buffering capacities, etc.) are calculated from the chosen routes.
2. Each shipment has a set of admissible routes. Constraints that depend on more than one route are modelled in the cost function.
3. A neighbour of a given solution is created by taking a small subset of the shipments and replacing their routes by others. To avoid extremely large neighbourhoods we discard neighbours which fail to fulfill a special local optimality criterion detailed in [3].

The model and the neighbourhood creation scheme allow us to implement a Simulated Annealing algorithm. The numerical results are given in the next section.

5 Numerical Results

We want to compare three different approaches:

1. The full model solved by CPLEX.
2. Heuristic results based upon Sect. 4.
3. The results of CPLEX for the restricted models.

We will use the seven benchmark instances I5, I10, I20, I30, I40, I50 and I60 with the respective number of depots which are based upon data from a large European road freight company. We solved each of it twice: with buffering and without buffering. The results are summarized in Table 1. We used a computer with 3.4 GHz and 16GB RAM for six hours, both for the heuristic and for CPLEX 12.6.0.

Contrary to our expectation, the buffering advantage does not show up in our heuristic results. Especially for the larger instances, we tend to get better results in the non-buffering case. There are practical and numerical reasons for this: Firstly, the option “buffering” increases the size of the search space and so slows down the heuristic. Secondly, the larger instances offer more other possibilities for consolidation so that buffering is not so important.

Comparing heuristic and CPLEX we see that although we drastically reduced the number of variables by preprocessing CPLEX fails for each of the instances over 20 depots. We see that the heuristic works well for the small instances; for the larger ones, we have no comparison.

Hence we solved the restricted models SAMEDAY and ONEHUB with the same solver and computer. The results are shown in Tables 2 and 3. We see that our heuristic outperforms all of them.

Table 1 Results of the heuristic approach compared to the results of CPLEX with preprocessing

Inst	Heur	Heur buf	CPLEX	CPLEX buf
I5	29,206,387	19,221,750	29,206,387	19,221,750
LB			29,206,387	19,221,750
I10	101,029,802	90,112,421	101,029,802	91,972,141
LB			101,029,802	82,633,347
I20	156,016,297	149,190,951	201,129,327	No solution
LB			135,228,841	No solution
I30	324,679,889	347,634,809	No solution	No solution
I40	468,567,791	470,264,568	No solution	No solution
I50	668,569,617	683,945,737	No solution	No solution
I60	978,018,115	977,338,269	No solution	No solution

Every problem is considered with and without the possibility of buffering

Table 2 Results from SAMEDAY compared to the best known results

Inst	SAMEDAY	Lower bound	Best known	Best known with buff
I5	30,171,383	30,171,383	29,206,387	19,221,750
I10	133,023,575	133,023,575	101,029,802	90,112,421
I20	222,652,292	163,991,639	156,016,297	149,190,951
I30	597,737,936	363,056,830	324,679,889	324,679,889
I40	No solution		468,567,791	468,567,791
I50	2,169,759,985	670,424,018	668,569,617	668,569,617
I60	No solution		978,018,115	977,338,269

Table 3 Results from ONEHUB (minimum of the results with and without strengthening inequalities) compared to the best known results

Inst	ONEHUB	Lower bound	Best known	Best known with buff
I5	29,206,387	29,206,387	29,206,387	19,221,750
I10	119,128,292	119,128,292	101,029,802	90,112,421
I20	200,685,015	188,607,303	156,016,297	149,190,951
I30	642,446,816	295,834,506	324,679,889	324,679,889
I40	7,881,411,010	381,556,712	468,567,791	468,567,791
I50	3,225,667,487	526,796,494	668,569,617	668,569,617
I60	4,813,948,039	739,362,497	978,018,115	977,338,269

6 Conclusion

The results show that the realistic hub location problem that we stated is very difficult for standard MIP solvers. This difficulty persists not only if we do preprocessing but also when we drastically reduce the complexity of the model. On the other hand, a heuristic approach based on [3] performs well. Hence we will follow two paths for the further development:

On the one hand, we will improve our heuristic by proper gauging. A heuristic procedure involves a huge number of search parameters which have to be calibrated by statistical methods. On the other hand, we will aim for better lower bounds by solving relaxed hub location problems, preferably with a Benders' decomposition approach.

References

1. Alumur S, Kara BY (2008) Network hub location problems: the state of the art. Eur J Oper Res 190(1):1–21
2. Fleischmann B (2008) Transport- und Tourenplanung. In: Arnold D, Isermann H, Kuhn A, Tempelmeier H, Furmans K (eds) Handbuch logistik, vol 3. Springer, pp 137–152
3. Meier JF (2014) A versatile heuristic approach for generalized hub location problems. Preprint, Provided upon personal request
4. Meier JF, Clausen U (2013) Heuristic strategies for a multi-allocation problem in LTL logistics. In: Operations research proceedings 2012: selected papers of the international annual conference of the German operations research society (GOR). Springer, Germany
5. Meier JF, Clausen U (2013) Strategic planning in LTL logistics—increasing the capacity utilization of trucks. Electron Notes Discret Math 41:37–44
6. Meier JF, Clausen U, Baumann F (2013) A column generation approach for strategic planning in LTL logistics. In: Proceedings of VeRoLog 2013
7. Sender J, Clausen U (2011) Hub location problems with choice of different hub capacities and vehicle types. In: Pahl J, Reiners T, Voss S (eds) Network optimization. Springer, Berlin, pp 535–546
8. Sender J, Clausen U (2013) Heuristics for solving a capacitated multiple allocation hub location problem with application in German wagonload traffic. Electron Notes Discret Math 41:13–20
9. Voll R, Clausen U (2013) Branch-and-price for a European variant of the railroad blocking problem. Electron Notes Discret Math 41:45–52