

The latest arrival hub location problem for cargo delivery systems with stopovers

Hande Yaman ^{*}, Bahar Y. Kara, Barbaros Ç. Tansel

Department of Industrial Engineering, Bilkent University, Bilkent 06800, Ankara, Turkey

Received 10 October 2005; received in revised form 28 March 2007; accepted 29 March 2007

Abstract

In this paper, we concentrate on the service structure of ground-transportation based cargo delivery companies. The transient times that arise from nonsimultaneous arrivals at hubs (typically spent for unloading, loading, and sorting operations) can constitute a significant portion of the total delivery time for cargo delivery systems. The latest arrival hub location problem is a new minimax model that focuses on the minimization of the arrival time of the last item to arrive, taking into account journey times as well as the transient times at hubs. We first focus on a typical cargo delivery firm operating in Turkey and observe that stopovers are essential components of a ground-based cargo delivery system. The existing formulations of the hub location problem in the literature do not allow stopovers since they assume direct connections between demand centers and hubs. In this paper, we propose a generic mathematical model, which allows stopovers for the latest arrival hub location problem. We improve the model using valid inequalities and lifting. We present computational results using data from the US and Turkey.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Hub location; Modeling; Stopovers; Sequential lifting; Valid inequalities

1. Introduction

In this paper, we provide a general model for ground-transportation based parcel-delivery services; the modeling requirements are obtained from a study of major parcel delivery firms operating in Turkey. One dominant fact that emerged from this study is that all of the investigated firms deliver cargo primarily using ground transportation systems, with supplementary air transportation taking place only occasionally and under rather exceptional circumstances. It seems reasonable to expect ground-based delivery systems in smaller or mid-sized countries or for regional shipments in larger sized countries where delivery over large territories may be more economical or quicker by air while restriction to regional shipments may justify the use of ground transportation systems.

^{*} Corresponding author. Fax: +90 312 2664054.

E-mail address: hyaman@bilkent.edu.tr (H. Yaman).

To give the problem some perspective, most cargo delivery firms operate on a hub based system where cargo items collected from various demand centers are consolidated, sorted, and rerouted at hubs to eventually arrive at their destinations (see, e.g., Sigafos and Easson, 1988). The problem we describe in this paper also fits this basic structure with one difference being that Turkish delivery systems rely mostly on ground transportation while hub-based distribution systems discussed in the literature predominantly focus on air transportation. In a hub based system, flows from origins to destinations follow a transportation route that consists of three legs: from the origin to its serving hub; from the serving hub to the hub that serves the destination; and finally from the second hub to the destination. Transportation costs between hubs are typically scaled by a discount factor α . The major design issue for hub based systems is the locations of hubs and the allocations of demand centers to hubs. For basic notions and models in hub location problems, the reader is referred to O'Kelly (1986, 1987), Campbell (1994, 1996), Ernst and Krishnamoorthy (1996, 1998), Skorin-Kapov et al. (1996) and Kara and Tansel (2000, 2003). A recent and extensive survey of hub location problems is given in Campbell et al. (2002), chapter 12 of Drezner and Hamacher (2002). Among recent studies are Boland et al. (2004), Carello et al. (2004), Ernst et al. (2002), Hamacher et al. (2004), Labbe and Yaman (2004), Labbe et al. (2005), Marin (2005), Mayer and Wagner (2002), Racunica and Wynter (2005), Tan and Kara (2007), Topcuoglu et al. (2005) and Yaman and Carello (2005).

There are a number of routing-related issues that are specific to ground transportation based delivery systems. These issues have not been addressed in the previous literature except by Kuby and Gray (1993) who introduce the notion of stopovers and propose a model for fixed hub locations that allows at most two stopovers in any given nonhub-to-hub journey. With this restriction, they are able to enumerate possible paths with at most two stopovers and include each such path as a possible decision. In our case, we allow an arbitrary number of stopovers on any given path in which case the number of paths is prohibitively too many to permit enumeration. We include such decisions in our model without relying on a-priori enumeration.

Kara and Tansel (2001) have worked on a problem that has been generally overlooked in the literature – the issue of unavoidable waiting times at hubs arising from a lack of synchronization of arriving and departing vehicles. Kara and Tansel (2001) defined the *Latest Arrival Hub Location Problem* which incorporates waiting times at hubs. Wagner (2004) observed that the standard p -Hub Center model yields the same solution as does the Latest Arrival Hub Location model because the bottleneck route in the latter model is a route for which the waiting times at hubs are necessarily zero. Despite that, we prefer to work with the Latest Arrival Hub Location problem due to a number of modeling insights and artifacts that are not available in the p -Hub Center model. With the realism and insights available in the Latest Arrival Hub Location Model, it is much easier, for example, to incorporate different real-world requirements such as the consideration of the “next-day-vehicle” (Tan and Kara, 2007). In this paper, we incorporate the issues related to multiple stopovers. Thus, we give a comprehensive model for hub-based parcel delivery systems that goes beyond standard location and allocation decisions to address issues relating to multiple stopovers, design of vehicle routes and inevitable waiting times at hubs.

We now give an overview of the paper. Section 2 is devoted to an analysis of the operating characteristics of ground-transportation based parcel delivery systems. In Section 3, we develop a general model which incorporates the main features of ground transport delivery systems. In Section 4, we provide some valid inequalities and lifting results to strengthen the model. Computational results are presented in Section 5. The paper ends with concluding remarks in Section 6.

2. Operating characteristics of ground transportation

We base our description of the main features of the operating structure of Turkish cargo delivery systems on the delivery network of one of the companies that we have interviewed. Other Turkish cargo delivery firms operate essentially in the same way with minor variations. The delivery network consists of two types of nodes: branch offices and operation centers. These nodes are connected by the routes of a ground transportation system. Branch offices are the points of initial contact with the customers; here the cargo is picked up from the customer and handled in trust of the company until it is picked up by the consignee. Operation centers are processing centers where incoming parcels are collected and rerouted to their destinations; these centers also act as branch offices.

When a customer wishes to send a parcel, she calls the nearest branch office to request pick-up or takes the parcel to the branch office herself. At the branch office, a bar code is placed on the parcel that designates the specific operation center through which the parcel will get delivered to the destination branch office. The journey of the parcel from the originating branch office to the destination branch office consists of three legs: the first leg from the originating branch office to its serving operation center; the second leg from this operation center to the final operation center which is the one designated on the bar code; and the third leg from this final operation center to the destination branch office. Sometimes, the first and second operation centers may be the same. In the first leg, the parcel, together with other parcels originating at the same branch office, are collectively delivered to an operation center that handles the deliveries from this branch office. Because the firm operates on a single-assignment basis, there is only one such operation center for each branch office. During its journey from the branch office to the assigned operation center, the delivery truck stops at various other branch offices to pick up their parcels. We refer to each such stop as a *stopover*. When the truck arrives at the assigned operation center, all parcels are unloaded and this completes the first leg of the journey.

The unloaded parcels at the operation center are sorted into groups according to the names of the operation centers printed on the bar codes. The second leg of the journey consists of the direct delivery from this operation center to the next operation center. The delivery between the two operation centers is done by means of larger trucks operating on express lines. No stopovers are allowed on express lines. An express line truck departing from an operation center waits for the arrival of all incoming trucks so as to receive all cargo whose bar codes designate the outbound operation center. The second leg of the journey is completed when the express line truck arrives at its designated operation center.

Parcels arriving at the same operation center on express line trucks are unloaded and sorted according to their final branch offices. As in the case of express line trucks, a delivery truck departing from an operation center towards branch offices waits for the arrival of all incoming trucks into that operation center, so as to receive all cargo that will be delivered by this truck. Incoming trucks may be delivery trucks from other branch offices or express line trucks from other operation centers. The third leg of the parcel's journey is completed when the delivery truck arrives at the destination branch office. The destination branch office may be a stopover for the truck in which case the truck continues its trip to other branch offices. When the parcel under consideration arrives at its destination branch office, it is either delivered to the address of the consignee or a call is placed to notify the customer.

For convenience, we refer to trips that connect branch offices to operation centers as *main lines* and to trips that connect two operation centers as *express lines*.

To summarize, the company uses a delivery network consisting of main lines and express lines. Main lines connect branch offices to operation centers and may visit many branch offices on the way to collect or deliver parcels. Express lines connect pairs of operation centers and no stopovers are allowed on such lines. Transient times occur at operation centers in order for departing trucks to receive all their outgoing parcels. Design issues relating to this network include the selection of locations of operation centers, the assignment of branch offices to operation centers, and the routing structures of main lines.

The operation centers may be viewed as hubs that collect and disseminate flows (parcels) between pairs of branch offices. It is natural to view branch offices as demand points that generate flows.

3. Problem definition and model development

One of the main concerns of delivery firms is to deliver the parcels to the consignees within a reasonable amount of time. Minimizing the longest delivery time approximates the objective of interest. This calls for a minimax objective. The problem is then to design a delivery network for which the associated maximum delivery time is minimized.

The Latest Arrival Hub Network Design Problem is the problem of determining the locations of hubs, the allocation of nonhubs to hubs, and the associated routing structure between nonhubs and hubs with multiple stopovers so as to minimize the arrival time of the latest arriving cargo at destinations. The model proposed in this paper gives a complete resolution of this problem.

We may assume that the delivery firm has a fleet composed of two types of vehicles with type 0 operating on express lines and type 1 on main lines. Type 1 vehicles are medium sized and operate between a nonhub and a

hub node. These vehicles may make multiple stops on each trip. Type 0 vehicles are larger sized and usually somewhat faster trucks that operate between two hubs. They are not allowed to make intermediate stops on their trips.

Let q_0 and q_1 be the number of trucks available in the fleet for express lines and main lines, respectively. Because each pair of hubs requires a type 0 vehicle, the number of hubs, p , is such that $q_0 \geq p(p - 1)$.

Let $N = \{1, 2, \dots, n\}$ be the set of nodes (branch offices). Let t_{kl} be the shortest travel time from node $k \in N$ to node $l \in N$ using the ground transportation network by a type 1 vehicle. We assume $t_{kl} = t_{lk}$. The travel time by a type 0 vehicle is taken as αt_{kl} . Note here that in this application α is used to capture the increased speed of the vehicles used between two hub nodes, in contrast to its traditional use as a cost discount.

Let a *trip* mean a maximal journey (path) traversed in the same vehicle beginning at a node where the empty vehicle is initially loaded and ending at a node where the vehicle is fully unloaded. Our formulation will be based on the foregoing definition of a trip. In the model, return trips are not used with the understanding that a truck completing its trip will follow the same route on the way back visiting the same stopovers in reverse order. Note that the travel times are symmetric, so this approach is reasonable.

Let r_i be the ready time of cargo at branch office $i \in N$. At a branch office, a certain amount of time is required to load or unload the truck. We define δ_j to be the time required for such processes at nonhub nodes $j \in N$.

Define h_i to be a zero-one variable with $h_i = 1$ if node $i \in N$ is a hub and 0 otherwise.

Define also

$$a_{ij} = \begin{cases} 1 & \text{if there is a main line trip starting at node } i \in N \text{ and terminating at hub } j \in N, \\ 0 & \text{otherwise.} \end{cases}$$

Note here that, in our model we only keep track of main line trips from branch offices towards hubs, i.e., from origins to hubs, and model the distribution to destinations as reverse trips.

Also define

$$s_i = \begin{cases} 1 & \text{if node } i \in N \text{ is a stopover,} \\ 0 & \text{otherwise.} \end{cases}$$

We have the following constraints:

Assignment of branch offices to trips: Each branch office should be assigned to a trip if it is not a hub itself. This is guaranteed via

$$\sum_{j \in N} a_{ij} + s_i + h_i = 1 \quad \forall i \in N. \tag{1}$$

That is, any node $i \in N$, is either a hub, a stopover, or starts a trip. Additionally, mainline trips should end in hubs, and this is guaranteed via

$$a_{ij} \leq h_j \quad \forall i \in N, j \in N. \tag{2}$$

Fleet composition: The number of assigned vehicles cannot exceed the available number

$$\sum_{i \in N} \sum_{j \in N} a_{ij} \leq q_1. \tag{3}$$

The following constraint forces the number of hubs opened to p .

$$\sum_{i \in N} h_i = p. \tag{4}$$

Determination of paths of trips: The routing structure of the main lines and express lines are to be determined by the model. Express lines are automatically determined when locations of hubs are specified. They are the least costly routes between pairs of hubs that permit the passage of an express line truck. The same cannot be said for main lines as these routes contain an unknown number of stopovers that will be determined by the model.

The routes (the sequence of arcs) followed by mainline trucks are defined in the model by means of flow variables x_{ij} which takes on the value 1 if there is a trip in which node $i \in N$ precedes node $j \in N \setminus \{i\}$ and 0 otherwise. These x_{ij} 's will form the arcs of the paths of the trips. We need to impose the following constraints so that the arcs constitute paths:

$$\sum_{j \in N \setminus \{i\}} x_{ij} = \sum_{j \in N} a_{ij} + s_i \quad \forall i \in N, \tag{5}$$

$$\sum_{j \in N \setminus \{i\}} x_{ji} = \sum_{j \in N} a_{ji} + s_i \quad \forall i \in N. \tag{6}$$

The right-hand-side of (5) is actually $1 - h_i$ due to (1). If node i is not a hub, then (5) guarantees that, there is an arc (in one trip) that leaves from node i . On the other hand, if i is a hub, no leaving arcs are permitted (due to the one-way definition of the trips). Due to constraint (6) if node i is a stopover, there is exactly one arc entering it. If node i is a hub, then the number of arcs entering it is equal to the number of main line trips ending at this node.

Arrival times: In what follows, we give the necessary machinery to compute the arrival times at the final destinations and to avoid subtours. Let D'_i be the time at which the truck on a main line trip arrives at node $i \in N$. Observe that, node i can be the starting point of the trip, a stopover, or the ending hub node. The variable D'_i should satisfy the following constraints:

$$D'_i \geq r_i \quad \forall i \in N, \tag{7}$$

$$D'_i \geq (D'_j + t_{ji} + \delta_j)x_{ji} \quad \forall j \in N, i \in N \setminus \{j\}. \tag{8}$$

The arrival time at any intermediate or final node i of a trip will be determined by the arrival time at the preceding node, j , on the path plus the handling time δ_j at that node plus the traveling time, t_{ji} . This is satisfied via constraint (8). Notice that due to constraint (8) subtours are eliminated.

Let D_j be the departure time of a main line truck on the way back from hub j (i.e., from the hub towards branch offices). Note here that, even though one may think of making this departure time dependent on the destination index, it will be evident shortly that such dependence need not be emphasized due to the so called “index independence property” (Kara and Tansel, 2001). The departure time of a vehicle from hub j occurs, at the earliest, when all incoming cargo at hub j is processed to identify the outbound cargo. The incoming cargo arrives at hub j either by mainline trips or by express line trucks. Recall that D'_j is the arrival time of a mainline truck at j and mainline trips end at hubs. Thus, D'_j of a hub node will determine the earliest possible departure time from that hub towards other hubs. The departure time of a mainline truck from hub j towards nonhub destinations is determined by the last vehicle to arrive there, i.e.,

$$D_j \geq (D'_i + \alpha t_{ij})h_i \quad \forall i \in N, j \in N, \tag{9}$$

$$D_j \geq D'_j \quad \forall j \in N. \tag{10}$$

Constraint (9) keeps the time of the latest arriving express line trucks at hub j and (10) keeps the time of the latest arriving mainline trucks at hub j . Observe that allowing $j = i$ in (9) actually provides (10). We remark here that, the expressions of the departure times of mainline trucks in (9) and (10) do not depend on the destination indices.

Let T_i be the time to reach $i \in N$ through all paths coming from nonhub nodes (independent of ready times)

$$T_j \geq (T_i + t_{ij} + \delta_i)x_{ij} \quad \forall i \in N, j \in N \setminus \{i\}, \tag{11}$$

$$T_j \geq 0 \quad \forall j \in N.$$

As can be seen from (11), the arrival time at any node j is determined by the arrival time at the preceding node i along the path plus the processing time at i and the travel time from i .

Objective function: The minimization of the latest arrival time results in the following objective function:

$$\min_{i \in N} \max(D_i + T_i)h_i,$$

which can be equivalently expressed as

$$\begin{aligned} \min \quad & Z, \\ \text{s.t.} \quad & Z \geq (D_i + T_i)h_i \quad \forall i \in N. \end{aligned} \tag{12}$$

The complete model is as follows:

$$\begin{aligned} \min \quad & Z, \\ \text{s.t.} \quad & \sum_{j \in N} a_{ij} + s_i + h_i = 1 \quad \forall i \in N, \end{aligned} \tag{1}$$

$$a_{ij} \leq h_j \quad \forall i \in N, j \in N, \tag{2}$$

$$\sum_{i \in N} h_i = p, \tag{3}$$

$$\sum_{i \in N} \sum_{j \in N} a_{ij} \leq q_1, \tag{4}$$

$$\sum_{j \in N \setminus \{i\}} x_{ij} = \sum_{j \in N} a_{ij} + s_i \quad \forall i \in N, \tag{5}$$

$$\sum_{j \in N \setminus \{i\}} x_{ji} = \sum_{j \in N} a_{ji} + s_i \quad \forall i \in N, \tag{6}$$

$$D'_i \geq r_i \quad \forall i \in N, \tag{7}$$

$$D'_j \geq (D'_i + t_{ij} + \delta_i)x_{ij} \quad \forall i \in N, j \in N \setminus \{i\}, \tag{8}$$

$$D_i \geq (D'_j + \alpha t_{ji})h_j \quad \forall i \in N, j \in N, \tag{9}$$

$$T_j \geq (T_i + t_{ij} + \delta_i)x_{ij} \quad \forall i \in N, j \in N \setminus \{i\}, \tag{11}$$

$$Z \geq (D_i + T_i)h_i \quad \forall i \in N, \tag{12}$$

$$a_{ij} \in \{0, 1\} \quad \forall i \in N, j \in N,$$

$$x_{ij} \in \{0, 1\} \quad \forall i \in N, j \in N \setminus \{i\},$$

$$s_i, h_i \in \{0, 1\} \quad \forall i \in N,$$

$$D_i, D'_i, T_i \geq 0 \quad \forall i \in N.$$

If we assume that the stopovers manage their ready times so that the trucks do not wait at any stopover, then we can change (7) to

$$D'_i \geq r_i \sum_{j \in N} a_{ij} \quad \forall i \in N,$$

so that we only wait for the ready times of the starting nodes.

The proposed model is a nonlinear mixed integer program with $2n^2 + n$ binary variables and $3n$ real variables. The number of constraints is $4n^2 + 3n + 2$. The nonlinearity is due to constraints (8), (9), (11), and (12).

Constraints (8) and (11) can be linearized as follows:

$$D'_j \geq D'_i + (t_{ij} + \delta_i)x_{ij} - M(1 - x_{ij}) \quad \forall i \in N, j \in N \setminus \{i\}, \tag{13}$$

$$T_j \geq T_i + (t_{ji} + \delta_i)x_{ij} - M(1 - x_{ij}) \quad \forall i \in N, j \in N \setminus \{i\}, \tag{14}$$

where M is a large number.

Constraints (9) can be linearized as follows:

$$D_i \geq D'_j + \alpha t_{ji}h_j \quad \forall i \in N, j \in N. \tag{15}$$

If $h_j = 1$ then (9) and (15) yield the same right hand sides. If $h_j = 0$ (i.e., node j is not a hub), then there is a main line trip that goes through node j . This trip ends at some node $m \in N$ which is a hub. Then constraint (13) implies that $D'_m \geq D'_j$. So, constraint (15) for node j is dominated by the same constraint for node m .

To linearize constraints (12), we first replace constraints (9) with the following constraints:

$$D_i \geq D'_j + \alpha t_{ji}(h_i + h_j - 1) \quad \forall i \in N, j \in N. \tag{16}$$

As we do not need the value of D_i if i is not a hub, this new model will be valid. Then, we can linearize constraints (12) as follows:

$$Z \geq D_i + T_i \quad \forall i \in N. \tag{17}$$

If node $i \in N$ is not a hub, then there exists an optimal solution in which $D_i \leq \max_{j \in N} D'_j$. As i is not a hub, the mainline trip that goes through node i ends at some node $m \in N$ which is a hub. Constraints (16) imply that $D_m \geq \max_{j \in N} D'_j$. So $D_m \geq D_i$ in such an optimal solution. Constraints (14) imply that $T_m \geq T_i$. So, constraint (17) for node i is dominated by the same constraint for node m .

So, a linear formulation can be obtained by replacing constraints (8), (9), (11), and (12) with (13), (14), (16), and (17).

In our test instances, $r_i = 0$ for all $i \in N$. This simplifies the formulation. With zero ready times, there exists an optimal solution where $T_i = D'_i$ for each $i \in N$ and T_i can be removed. Also, D_i 's can be removed from the formulation. Thus, we have

$$Z \geq D'_i + D'_j + \alpha t_{ji}(h_i + h_j - 1) \quad \forall i \in N, j \in N \setminus \{i\}, \tag{18}$$

$$Z \geq 2D'_i \quad \forall i \in N. \tag{19}$$

We also aggregate constraints (2) to decrease the size of the formulation and get the following constraints:

$$\sum_{i \in N} a_{ij} \leq q_1 h_j \quad \forall j \in N. \tag{20}$$

The resulting formulation is as follows:

$$\begin{aligned} \min \quad & Z, \\ \text{s.t.} \quad & \sum_{j \in N} a_{ij} + s_i + h_i = 1 \quad \forall i \in N, \end{aligned} \tag{1}$$

$$\sum_{i \in N} a_{ij} \leq q_1 h_j \quad \forall j \in N, \tag{20}$$

$$\sum_{i \in N} h_i = p, \tag{3}$$

$$\sum_{i \in N} \sum_{j \in N} a_{ij} \leq q_1, \tag{4}$$

$$\sum_{j \in N \setminus \{i\}} x_{ij} = \sum_{j \in N} a_{ij} + s_i \quad \forall i \in N, \tag{5}$$

$$\sum_{j \in N \setminus \{i\}} x_{ji} = \sum_{j \in N} a_{ji} + s_i \quad \forall i \in N, \tag{6}$$

$$D'_i \geq D'_j + (t_{ji} + \delta_j)x_{ji} - M(1 - x_{ji}) \quad \forall i \in N, j \in N \setminus \{i\}, \tag{13}$$

$$Z \geq D'_i + D'_j + \alpha t_{ji}(h_i + h_j - 1) \quad \forall i \in N, j \in N \setminus \{i\}, \tag{18}$$

$$Z \geq 2D'_i \quad \forall i \in N, \tag{19}$$

$$a_{ij} \in \{0, 1\} \quad \forall i \in N, j \in N,$$

$$x_{ij} \in \{0, 1\} \quad \forall i \in N, j \in N \setminus \{i\},$$

$$s_i, h_i \in \{0, 1\} \quad \forall i \in N,$$

$$D'_i \geq 0 \quad \forall i \in N.$$

The final version has $2n^2 + n$ binary variables and n real variables. The number of constraints is $2n^2 + 3n + 2$.

4. Lifting results and valid inequalities

In this section, we propose several ways of improving the strength of our formulation. First, we use lifting to strengthen two families of constraints. Then, we derive a new family of valid inequalities using lifting. Let F

be the set of feasible solutions. First we lift variables x_{ij} and x_{ji} in constraints (18) and obtain the following inequalities.

Proposition 1. For $i \in N$ and $j \in N \setminus \{i\}$, the inequality

$$Z \geq D'_i + D'_j + \alpha t_{ji}(h_i + h_j - 1) + (t_{ji} + \delta_j)x_{ji} + (t_{ij} + \delta_i)x_{ij} \tag{21}$$

is valid for F .

Proof. If $x_{ji} = 0$ and $x_{ij} = 0$, then the inequality is the same as constraint (18). If $x_{ji} = 1$, then $x_{ij} = 0$, $h_j = 0$ and the inequality is $Z \geq D'_i + D'_j + \alpha t_{ji}(h_i - 1) + t_{ji} + \delta_j$. As $Z \geq 2D'_i$ by constraint (19), $D'_i \geq D'_j + t_{ji} + \delta_j$ by constraints (18) and $h_i \leq 1$, the inequality is satisfied. The case where $x_{ij} = 1$ is similar. So, the inequality (21) is valid. \square

It is easy to see that inequality (21) dominates constraint (18).

Next, we lift variable x_{ij} in constraint (19).

Proposition 2. For $i \in N$, the inequality

$$Z \geq 2 \left(D'_i + \sum_{j \in N \setminus \{i\}} (t_{ij} + \delta_i)x_{ij} \right) \tag{22}$$

is valid for F .

Proof. If $x_{ij} = 1$ for some $j \in N \setminus \{i\}$, then $D'_j \geq D'_i + t_{ij} + \delta_i$. As $Z \geq 2D'_j$ by constraint (19), $Z \geq 2(D'_i + t_{ij} + \delta_i)$. If $\sum_{j \in N \setminus \{i\}} x_{ij} = 0$, then the inequality is the same as constraint (19). \square

Observe that inequality (22) dominates constraint (19).

Finally, we derive a new family of valid inequalities which give lower bounds on Z .

Proposition 3. For $i \in N$ and $j \in N \setminus \{i\}$, define

$$\begin{aligned} \lambda_{ij}^0 &= \alpha t_{ij}, \\ \lambda_{ij}^i &= \min_{m \in N \setminus \{i\}} (t_{im} + \delta_i + \alpha t_{mj}), \\ \lambda_{ij}^j &= \min_{m \in N \setminus \{j\}} (\alpha t_{im} + t_{mj} + \delta_j), \\ \lambda_{ij} &= \min_{m \in N \setminus \{i\}} \min_{l \in N \setminus \{j\}} (t_{im} + \delta_i + \alpha t_{ml} + t_{lj} + \delta_j). \end{aligned}$$

The inequality

$$Z \geq \lambda_{ij}^0 + (\lambda_{ij}^j - \lambda_{ij}^0)(1 - h_j) + \min\{\lambda_{ij} - \lambda_{ij}^j, \lambda_{ij}^i - \lambda_{ij}^0\}(1 - h_i) \tag{23}$$

is valid for F .

Proof. Inequality $Z \geq \lambda_{ij}^0$ is valid when $h_i = h_j = 1$. We first lift h_j and then h_i . For inequality $Z \geq \lambda_{ij}^0 + \omega(1 - h_j)$ to be valid, we need $\omega \leq Z - \lambda_{ij}^0$ for all possible solutions such that $h_i = 1$ and $h_j = 0$. So, $\omega = \min_{F: h_i=0, h_j=1} Z - \lambda_{ij}^0$. It is easy to see that $\min_{F: h_i=0, h_j=1} Z = \lambda_{ij}^j$ as triangle inequality is satisfied and $0 \leq \alpha \leq 1$. So $\omega = \lambda_{ij}^j - \lambda_{ij}^0$.

Next, we lift variable h_i . For inequality $Z \geq \lambda_{ij}^0 + (\lambda_{ij}^j - \lambda_{ij}^0)(1 - h_j) + \omega(1 - h_i)$ to be valid, we need $\omega \leq Z - \lambda_{ij}^0 - (\lambda_{ij}^j - \lambda_{ij}^0)(1 - h_j)$ for all possible solutions with $h_i = 0$. We can compute the tightest value of ω as $\min_{F: h_i=0} Z - \lambda_{ij}^0 - (\lambda_{ij}^j - \lambda_{ij}^0)(1 - h_j)$. When $h_j = 1$, minimum value for Z is λ_{ij}^i and when $h_j = 0$, minimum value for Z is λ_{ij} . So, $\omega = \min\{\lambda_{ij} - \lambda_{ij}^j, \lambda_{ij}^i - \lambda_{ij}^0\}$. \square

In the following section, we present computational results which illustrate the effect of lifting and valid inequalities in solving sample problems.

5. Computational results

We test our models with two different data sets: the CAB data set and a Turkish network. For computational tests we use CPLEX 8.1 on an AMD Opteron 252 processor (2.6 GHz) with 2 GB of RAM. The CAB data set is considered to be a benchmark by most researchers in the hub location area. The set is generated from the Civil Aeronautics Board Survey of 1970 passenger data in the United States (O’Kelly, 1987). It contains passenger flows and distances between 25 cities. In our experimental design, we take $p \in \{2, 3, 4\}$. In standard hub location models, it is customarily assumed that there is cost savings due to bulk transportation between two hubs. This savings is incorporated into the models by multiplying the cost parameters with an “economies of scale factor, α ”, which is between 0 and 1. When time is a concern, there will not be any time savings due to bulk transportation. Our investigation of the cargo companies in Turkey revealed that they use more specialized and faster trucks between hub nodes. The trucks used in express lines drive on the average 1.11 times faster than ordinary trucks. Thus we take $\alpha = 0.9$. The CAB data set is based on airlines with an unlimited fleet size assumption. Since our models relax that assumption, we need the fleet size as an additional parameter and we take $q_1 \in \{4, 5\}$.

We first want to investigate the effects of valid inequalities and lifting and for that purpose we focus on the cases with $n = 10$ and 15. Table 1 summarizes the results for $n = 10$ and Table 2 gives the results of $n = 15$. We put a time bound of one hour on CPLEX. Both of the tables have four main parts. The first two columns give the problem parameters of the instance which are the number of hubs, p , and the fleet size, q_1 . The next three columns give the optimum solution: optimal locations of hubs, the stopovers in the solution, and the optimal objective function value (the maximum time to reach any of the demand centers). In the next two columns of Table 1 are the CPLEX performance measures of the model without valid inequalities and lifting. We refer to this model as Model 1. Here, we report the number of nodes in the branch and cut tree and the cpu time in seconds. The last two columns of this table report the same performance measures obtained using the model, referred to as Model 2, where constraints (18) and (19) are replaced with inequalities (21) and (22), respectively, and inequalities (23) are included.

For $n = 10$, we could solve both models 1 and 2 to optimality within seconds. Model 2 outperforms Model 1 for all instances.

For $n = 15$, we could not get the optimum solutions within an hour with Model 1. Thus, in Table 2, the columns related to Model 1 report the final upper (ub) and lower bounds (lb) and the final % gap, i.e., $\frac{ub-lb}{ub} * 100$. In the columns related to Model 2, we report the initial % gap, i.e., $\frac{opt-db}{opt} * 100$ where opt is the

Table 1
Results for 10-node problems in CAB data, with and without valid inequalities and lifting

p	q_1	Optimal			Model 1		Model 2	
		Hubs	Stopovers	Obj. value	No. of nodes	cpu	No. of nodes	cpu
2	4	4,7	2,5,6,9	2046	37715	13.63	19194	6.43
2	5	4,10	5,6,9	1854	8808	3.28	5517	1.95
3	4	4,8,10	5,6,9	1760	9756	4.16	3497	1.61
3	5	6,7,8	5,9	1730	31542	9.32	3198	1.40
4	4	3,4,7,8	5,9	1630.7	1031	1.17	526	0.27
4	5	3,4,7,8	5,6	1630.7	632	1.05	560	0.28

Table 2
Results for 15-node problems in CAB data, with and without valid inequalities and lifting

p	q_1	Optimal			Model 1			Model 2		
		Hubs	Stopovers	Obj. value	Final ub	Final lb	Final % gap	Init. % gap	No. of nodes	cpu
3	7	3,12,13	4,5,9,10,11	2345.4	2419.4	2021.2	16.46	0.23	134306	167.20
3	9	3,12,13	1,4,9,11	2345.4	2412.4	2230	7.56	0.23	147026	171.27
4	7	3,11,12,13	4,5,6,7	2340	2340	2132	8.89	0.00	12897	8.85
4	9	3,11,12,14	4,6,10,13,15	2340	2340	1992.1	14.87	0.00	533655	532.81

optimal value and db is the optimal value of the linear programming relaxation, the number of nodes in the branch and cut tree and the cpu time in seconds.

Observe from Table 2 that, with Model 2 we get the optimum within minutes whereas with Model 1 we could not get any optimal results within an hour. We also remark here that the initial lower bound is improved substantially with the inclusion of the valid inequalities. Thus, for the Turkish network we utilised Model 2.

In our application with the Turkish network, we take the most populated and industrialized 16 cities as the nodes (Fig. 1). These 16 cities are geographically separate except for nodes 10 and 12, Izmir and Manisa, which are close by. These two cities are included because they are in the top 10% of industrialized cities of Turkey.

We again take $p \in \{2,3,4\}$ and $\alpha = 0.9$. The size of the fleet will affect the number of stopover cities. Observe that, if the fleet size is large enough, the model would assign a dedicated trip for each node. For each p value we wanted to have two cases: at least 2 nodes served as stopovers and at least 4 nodes served as stopovers corresponding respectively to about 10% and 25% of the nodes. Since the number of nodes is 16, we need to allocate $16 - p$ nodes. For $p = 2$, if the fleet size is 12, the remaining two nodes will be served as stopovers. With similar reasoning, we set the fleet size parameter q_1 . Table 3 depicts the results of the model with the Turkish data. The first two columns of the table provide the p and q_1 parameters of each instance. The optimal hub set, the stopover cities and the objective function value are given in columns 3–5. The last three columns provide the performance of CPLEX: the initial % gap, the number of nodes and the cpu time in seconds.

As can be seen from the table, for fixed $p + q_1$, the initial % gap decreases as p increases. For fixed p , it increases as q_1 decreases. Notice that the initial % gaps are much higher compared to the gaps we obtained with the CAB data for $n = 15$ and it takes a longer time to close this gap.

Observe from Table 3 that, node 14 is among the optimal hubs for all the instances and node 2 is among the optimal hubs for five of the instances. The main reason for this may be that these cities are more centrally located than all others.

For $p = 2$ and $q_1 = 10$, the longest delivery time is between origin-destination pair 10 and 16 (Izmir and Urfa). When p is increased to 3, for the same $p + q_1$ combination, the longest delivery time is between nodes 10 and 5, Izmir and Diyarbakir. We provide the allocation structure of three hubs with nine trucks in Fig. 2.

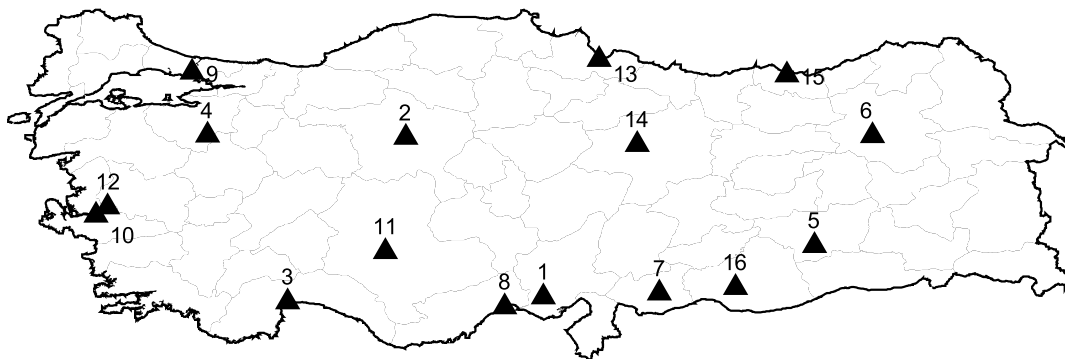


Fig. 1. The 16 cities over the Turkey map.

Table 3
Results for Turkish data

p	q_1	Optimal			Init. % gap	No. of nodes	cpu
		Hubs	Stopovers	Obj. value			
2	12	2,14	1,11	1505.8	11.66	1639924	4002.10
2	10	2,14	1,4,11,12	1549.8	14.17	2188234	6377.06
3	11	1,2,14	7,8	1476.8	10.97	1427932	3012.60
3	9	2,8,14	1,7,11,12	1504.8	12.63	4928150	5112.90
4	10	2,8,12,14	1,7	1438.7	8.98	2835159	3470.48
4	8	1,9,11,14	7,8,12,16	1483	11.70	4952607	3049.33

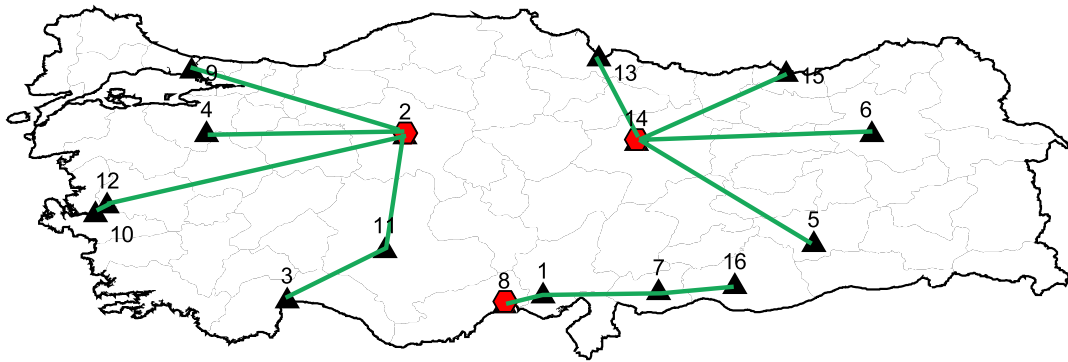


Fig. 2. The allocation structure of three hubs and nine trucks.

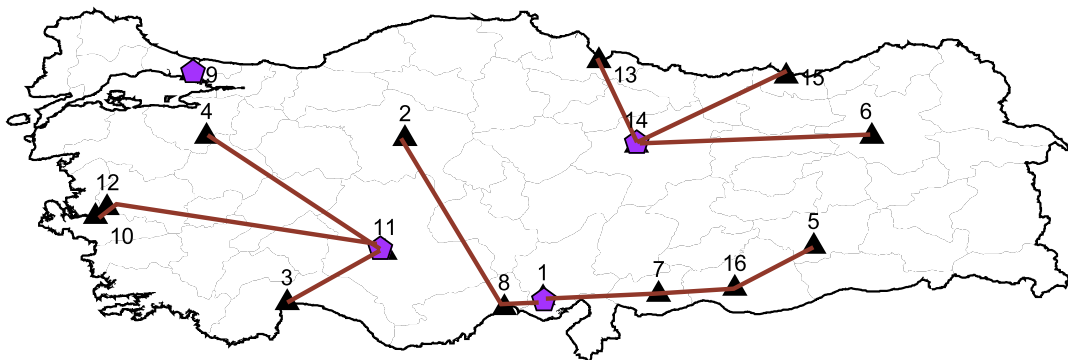


Fig. 3. The allocation structure of four hubs and eight trucks.

For $p = 4$ and $q_1 = 8$, the hubs are located in Istanbul (9), Sivas (14), Konya (11), and Adana (1). The longest delivery time is between nodes 10 (Izmir) and 6 (Erzurum). The allocation structure of four hubs with eight trucks is given in Fig. 3.

For all the cases with $p + q_1 = 12$, since nodes 10 and 12 are very close to each other, they are served with the same truck, with node 12 being the stopover.

5.1. Cost/service quality trade-off

Even though the primary focus of the proposed model is the minimization of the latest delivery time, additional insights can be gained by a comparison to a cost related measure. A solution that yields the minimum latest delivery time does not generally yield the minimum total cost. In fact, these two measures of performance are in fair amount of conflict and any gain in one is expected to be accompanied by a loss in the other.

To see the trade-offs between these two measures, we have conducted a computational study by selecting a problem from our test bed that has three hubs and nine trucks (the test problem in line 4 of Table 3) and minimizing the total travel time for different amounts of deviations from the minimum latest delivery time. We may assume that the total travel time is a reasonable approximation of the total cost (up to a multiplicative constant) since transportation costs are generally incurred in direct proportion to travel times in city-to-city cargo delivery networks. The minimum latest delivery time rounded up to the closest integer is 1505 for the selected problem which we take as a base value. Deviations from the base in increments of 5% are allowed on the latest delivery time and the corresponding minimizing total travel time is computed for each such deviation by solving a modified version of Model 2 where the objective function is replaced by the sum of travel

times while new constraints on delivery times with an upper bound of $Z^u = 1505(1 + \beta)$ rounded to the closest integer are added to the original constraint set. Here, β refers to the percent deviation allowed from the base value. Additional variables are also introduced as necessary to make the modified model linear and consistent with the new structure. In particular, we replace the objective function with $\min 2\sum_{i \in N} \sum_{j \in N} t_{ij}x_{ij} + \sum_{i \in N} \sum_{j \in N} \alpha t_{ij}y_{ij}$ and we add the constraints $Z \leq Z^u$, $y_{ij} \geq h_i + h_j - 1$ and $y_{ij} \geq 0$ for all $i \in N$ and $j \in N$.

The results of the trade-off study are tabulated in Table 4. The columns of the table give the allowed percent increases in latest delivery times and the corresponding latest delivery times, hub locations, total travel times, and percent savings in total travel time.

An inspection of the last column indicates that a 5% deviation from the base is accompanied by slightly less than 5% savings in total travel time. When the deviation from the base is increased from 5% to 10%, the percent savings in total time goes up from about 5% to 10.5%. Further increase from the base value leads to a somewhat steady but rather slow improvement in total time, ending in about 22% savings at a deviation of 100% from the base. Consequently, the savings in total time by deviating 10% from the base is almost as much as the additional savings obtained by deviating 100% (10.5% versus 22%). If we interpret the deviation from the base as an indicator of deterioration of service quality, we may conclude that a cost conscious company may justifiably allow up to 10% deterioration while additional savings in costs is not well justified unless the company is prepared to remedy the consequences of substantial amounts of loss in service quality.

Additional insights are gained from a study of the changes in hub locations. The third column of Table 4 gives the hub locations (as defined by city numbers shown in Fig. 1). The initial solution corresponding to the minimum possible latest delivery time places the three hubs at cities 2, 8, and 14 (Ankara, Mersin, and Sivas). The latest delivery time for this solution is defined by the city pair 10 and 5 where city 10 is connected to the hub at city 2 via stopover at city 12 and city 5 is connected to the hub at city 14 directly. At 5% deviation from the base, the hub locations at cities 2 and 14 survive while the remaining hub location switches from city 8 (Mersin) to city 7 (Gaziantep). The latest delivery time for this solution is defined by the city pair 9 and 5 (Istanbul and Diyarbakir) where city 9 is connected to hub at city 2 via stopover at city 4 and city 5 is connected to the new hub at city 7 directly. Note also that the allocation of city 5 is switched from hub 14 in the initial solution to hub 7 in the new solution. The critical path in the first solution that defines the latest delivery time via the route 10–12–2–14–5 is replaced now by a longer critical path in the new solution defined by the route 9–4–2–7–5, thereby leading to a cost reduction of about 5% by allowing a deterioration of close to 5% in service quality. At 10% deviation from the base, the hub locations at cities 2 and 14 continue to survive while the third hub switches now from city 7 to city 1 (Adana). The critical path in this solution is defined by the route 9–4–2–1–7–16–5 which is identical to the previous critical path in the portion 9–4–2 but differs from it in the remainder. Note also the addition of stopovers at cities 7 and 16 to the route that connects city 5 to its hub

Table 4
Latest delivery time/total travel time trade-off

% Increase in latest time β	Latest time	Hubs	Total time	% Decrease in total time
	1505	2, 8, 14	10759.6	
5	1580	2, 7, 14	10226.0	4.96
10	1655	1, 2, 14	9628.0	10.52
15	1731	1, 11, 14	9435.2	12.31
20	1806	1, 2, 14	9364.0	12.97
25	1881	1, 11, 14	9225.2	14.26
30	1956	1, 11, 14	9225.2	14.26
35	2031	1, 11, 14	9063.2	15.77
40	2107	1, 11, 14	8961.2	16.71
45	2182	1, 11, 14	8853.2	17.72
50	2257	2, 13, 14	8804.0	18.18
60	2408	1, 8, 11	8651.4	19.59
70	2558	1, 8, 11	8563.4	20.41
80	2709	1, 8, 11	8563.4	20.41
90	2859	1, 8, 11	8421.4	21.73
100	3010	1, 8, 11	8389.4	22.03

at city 1. From a deviation of 15–45%, the optimal hub locations are at cities 1, 11, and 14 with the exception of a single switch from city 11 to city 2 at 20% deviation (a return to city 11 occurs at 25% deviation). Note that the hub locations 2 and 14 that have remained intact between 0% and 10% deviation from the base have lost their advantage to the hub pair 1 and 14 from this point on (up to a β value of 45%). Note also that hub 14 is the only hub that has survived throughout all switches from 0% to 50% deviation from the base. At 50% deviation from the base, a nonsurviving switch occurs from the hub triplet (1,11,14) to the triplet (2,13,14) which is immediately replaced by the triplet (1,8,11) at 60% deviation and has continued to be the active triplet up to a 100% deviation from the base. What essentially happens in deviating 15–100% from the base is that the hubs at 1 and 11 survive throughout while hub 14 that survives from 15% to 50% is replaced by hub 8 at 60% deviation. This predominant pattern is temporarily disturbed at 20% and 50% which is certainly possible due to the existence of many routes of various different lengths one of which may become active in a short range of β values thereby altering the structure of locational patterns that may be predominant outside of these ranges.

We observe that for $\beta = 60, \dots, 100$, nodes 1 and 8, which are very close to each other, are among the optimal hub locations. When we solve the same instances with $p = 2$, we observe that the total travel times are less and the optimal hub locations are nodes 1 and 11 for $\beta = 60, \dots, 90$, and nodes 8 and 11 for $\beta = 100$.

The trade-off study discussed above is based on a modeling assumption that disregards the magnitudes of flows. The latest arrival model that we formulated in the paper is based on the assumption that a single truck is assigned to each route connecting nonhub or hub nodes to hub nodes. The capacity of the assigned truck is assumed to be sufficient to carry out the pick-up and delivery operations along the route. Consider, however, a situation where the volume flowing on a given route may be so large that the load capacity of a single truck may not be sufficient. This would require the assignment of multiple trucks along the same route so that the contribution of each such route to the total cost would be the number of trucks operating along that route times the length of that route. This requires a fundamentally different modeling viewpoint (capacitated versus uncapacitated trucks) and is the subject of a different study. Nevertheless, we may address the question of how the magnitudes of flows may affect the locational pattern by means of a crude comparison based on a p -hub median model that takes magnitudes of flows into account with a latest arrival model without stopovers. We have solved a p -hub median problem for the cases of 2, 3, and 4 hubs using the same distance data and with the addition of the flow data that assigns the flow f_{ij} to the city pair (i,j) where f_{ij} is the product of the populations of cities i and j divided by a constant. The resulting hub locations for the p -hub median are nodes 1 and 2 (Adana and Ankara) for $p = 2$, nodes 2, 4, and 9 (Ankara, Bursa, and Istanbul) for $p = 3$, and nodes 1, 2, 9, and 12 (Adana, Ankara, Istanbul, and Manisa) for $p = 4$. In contrast, the hub locations for the latest arrival model without stopovers are nodes 2 and 14 (Ankara and Sivas) for $p = 2$, nodes 2, 12, and 14 (Ankara, Manisa, and Sivas) for $p = 3$, and nodes 2, 5, 14, and 16 (Ankara, Diyarbakir, Sivas, and Urfa) for $p = 4$. There is quite a bit of discrepancy between the solutions of the two models (as expected). The p -hub median model has shifted the favored hub locations to or nearby the more heavily populated cities which are Istanbul, Ankara, Izmir (nearby location is Manisa), Adana, and Bursa whereas the latest arrival model without stopovers has favored more central locations that serve to reduce the length of the longest delivery route. We note that, had the populations of the cities been taken into account in the latest arrival models, the resulting models might have proposed solutions that may be more in alignment with the p -hub median model. This is left for future work.

6. Conclusion

The main premise of this study has been to identify the operational characteristics of a ground transportation based delivery systems and to give an all-inclusive model that captures all essential components. To our knowledge, such issues have not been collectively modeled in the earlier literature. The model that has been proposed is a zero-one mixed integer programming model with four classes of nonlinear constraints. Linearizations are proposed to handle nonlinearity. Valid inequalities are developed to strengthen the proposed model. The use of these inequalities in solving the problem instances is verified using the standard CAB data set. The final model is then tested with a data set for parcel delivery in Turkey.

Acknowledgment

We are grateful to an anonymous referee for suggesting the cost/service quality trade-off analysis.

References

- Boland, N., Krishnamoorthy, M., Ernst, A.T., Ebery, J., 2004. Preprocessing and cutting for multiple allocation hub location problems. *European Journal of Operational Research* 155, 638–653.
- Campbell, J.F., 1994. Integer programming formulations of discrete hub location problems. *European Journal of Operational Research* 72, 387–405.
- Campbell, J.F., 1996. Hub location and the p -hub median problem. *Operations Research* 44, 923–935.
- Campbell, J.F., Ernst, A.T., Krishnamoorthy, M., 2002. Hub Location Problems. In: Drezner, Z., Hamacher, H. (Eds.), *Facility Location: Applications and Theory*. Springer-Verlag, Berlin.
- Carello, G., Della Croce, F., Ghirardi, M., Tadei, R., 2004. Solving the hub location problem in telecommunication network design: a local search approach. *Networks* 44, 94–105.
- Ernst, A.T., Krishnamoorthy, M., 1996. Efficient algorithms for the uncapacitated single allocation p -hub median problem. *Location Science* 4, 139–154.
- Ernst, A.T., Krishnamoorthy, M., 1998. Exact and heuristic algorithms for the uncapacitated multiple allocation p -hub median problem. *European Journal of Operational Research* 104, 100–112.
- Ernst, A.T., Hamacher, H., Jiang, H., Krishnamoorthy, M., Woeginger, G., 2002. Uncapacitated single and multi allocation p -hub center problems. Technical Report, CSIRO Mathematical and Information Sciences, Australia.
- Hamacher, H., Labbe, M., Nickel, S., Sonneborn, T., 2004. Adapting polyhedral properties from facility to hub location problems. *Discrete Applied Mathematics* 145, 104–116.
- Kara, B.Y., Tansel, B.C., 2000. On the single assignment p -hub center problem. *European Journal of Operational Research* 125 (3), 648–655.
- Kara, B.Y., Tansel, B.C., 2001. The latest arrival hub location problem. *Management Science* 47 (10), 1408–1420.
- Kara, B.Y., Tansel, B.C., 2003. The single assignment hub covering problem: models and linearizations. *Journal of the Operational Research Society* 54 (1), 59–64.
- Kuby, M.J., Gray, R.G., 1993. The hub network design problem with stopover and feeders: the case of federal express. *Transportation Research Part A* 27, 1–12.
- Labbe, M., Yaman, H., 2004. Projecting the flow variables for hub location problems. *Networks* 44, 84–93.
- Labbe, M., Yaman, H., Gourdin, E., 2005. A branch and cut algorithm for hub location problems with single assignment. *Mathematical Programming* 102, 371–405.
- Marin, A., 2005. Formulating and solving splittable capacitated multiple allocation hub location problems. *Computers and Operations Research* 32, 3093–3109.
- Mayer, G., Wagner, B., 2002. HubLocator: an exact solution method for the multiple allocation hub location problem. *Computers and Operations Research* 29, 715–739.
- O’Kelly, M.E., 1986. The location of interacting hub facilities. *Transportation Science* 20, 92–105.
- O’Kelly, M.E., 1987. A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research* 32, 393–404.
- Racunica, I., Wynter, L., 2005. Optimal location of intermodal freight hubs. *Transportation Research Part B* 39, 453–477.
- Sigafoos, R.A., Eason, R.R., 1988. *Absolutely Positively Overnight: The Unofficial Corporate History of FedEx*. St Lukes Press, Memphis.
- Skorin-Kapov, D., Skorin-Kapov, J., O’Kelly, M.E., 1996. Tight linear programming relaxations of uncapacitated p -hub median problems. *European Journal of Operational Research* 94, 582–593.
- Tan, P.Z., Kara, B.Y., 2007. A hub covering model for cargo delivery systems. *Networks* 49, 28–39.
- Topcuoglu, H., Corut, F., Ermis, M., Yilmaz, G., 2005. Solving the uncapacitated hub location problem using genetic algorithms. *Computers and Operations Research* 32, 967–984.
- Wagner, B., 2004. A note on “The latest arrival hub location problem”. *Management Science* 50, 1751–1752.
- Yaman, H., Carello, G., 2005. Solving the hub location problem with modular link capacities. *Computers and Operations Research* 32, 3227–3245.