Computers & Operations Research 36 (2009) 3097 – 3106

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/cor

Multiple allocation hub-and-spoke network design under hub congestion

R.S. de Camargo[a,*], G. Miranda Jr.[a], R.P.M. Ferreira[b], H.P. Luna[c]

[a]Department of Industrial Engineering, Federal University of Minas Gerais, Brazil
[b]Department of Computer Science, Pontifical Catholic University of Minas Gerais, Brazil
[c]Computer Science Institute, Federal University of Alagoas, Brazil

ARTICLE INFO

Available online 21 October 2008

Keywords:
Hub-and-spoke networks
Benders decomposition
Large scale optimization
Mixed integer non-linear programming

ABSTRACT

The multiple allocation hub-and-spoke network design under hub congestion problem is addressed in this paper. A non-linear mixed integer programming formulation is proposed, modeling the congestion as a convex cost function. A generalized Benders decomposition algorithm has been deployed and has successfully solved standard data set instances up to 81 nodes. The proposed algorithm has also outperformed a commercial leading edge non-linear integer programming package. The main contribution of this work is to establish a compromise between the transportation cost savings induced by the economies of scale exploitation and the costs associated with the congestion effects.

1. Introduction

Hub-and-spoke systems have been largely employed in the telecommunication and transportation areas (see [1–5]). This class of systems arises when commodities (passenger, cargo parcels, telecommunication packets) of a set of origins must be sent to a set of destinations. Instead of establishing a direct connection for eachorigin–destination pair, facilities that serve as switching, transshipment, sorting and distribution nodes, designated as hubs, are used as the only valid intermediate points in a path from an origin to a destination. Flows from different origins are gathered at these hub facilities prior to be routed to an intermediate hub or to be delivered to their final destinations.

The employment of these hub facilities and the routing of consolidated flows through inter-hub links allow the centralization of commodity handling and the transportation cost per unit of flow to be less expensive than directly shipping via a nonhub network structure. In other words, the hub networks take advantage of scale economies on inter-hub connections [6,7].

Usually, the main concerns of hub network design are the location of hub facilities and the allocation of nonhub nodes to hubs at the least expensive network structure, accounting for that the installation and the transportation costs. This is done by selecting node candidates as hubs and assigning origin and destination nodes to these hubs, given the flow volumes between each origin–destination node pair, the installation costs of hubs and the transportation costs between nodes of the network. While hubs are in general fully connected among them, nonhub nodes can be single allocated, meaning that a nonhub can be assigned to exactly one hub only; and multiple allocated, meaning that a nonhub can be connected to more than one hub of the network. Alternative configurations can be considered in order to ensure the application of a specific protocol [8]. For further basic notions, different considerations and formulations of hub networks see the comprehensive surveys of Campbell [9,10] and of Campbell et al. [5] and of Alumur and Kara [11].

The minimization of the installation and transportation costs approach generates solutions that tend to overload a small number of hubs, forcing the congestion effects into consideration. The hub literature has been dealing with the congestion cost (CC) effects on hub networks implicitly, when these effects are represented by constraints, and explicitly, when the CCs are expressed on the objective function.

Grove and O'Kelly [12] are among the first authors to study the effects of congestion on hub networks. They have demonstrated how the schedule delays of airline systems are influenced by the amount of flow at the hubs by simulating a single assignment hub network with fixed hub locations.

A great variety of researchers has tackled the congestion effects restricting the amount of flow transiting through a hub by means of capacity constraints. Aykin [13] has devised a Lagrangian relaxation approach, while Ernst and Krishnamoorthy [14] have used simulated annealing and random descent methods. Ebery et al. [15] have imposed capacities only on incoming flows. They have developed a branch-and-bound based on a shortest path heuristic. Solutions obtained by their method may have flows from a hub to itself routed via another hub as explained by Campbell et al. [5]. Campbell et al.

* Corresponding author. Tel.: +55 313 225 4027.
E-mail addresses: rcamargo@dep.ufmg.br (R.S. de Camargo), miranda@dep.ufmg.br (G. Miranda Jr.), poley@pucminas.br (R.P.M. Ferreira), pacca@tci.ufal.br (H.P. Luna).

0305-0548/$ - see front matter
doi:10.1016/j.cor.2008.10.004

[16–18] have proposed the hub arc location problem, in which the hub network is seen as a three-layered network. The bottom one has the origin nodes of the demand, the middle layer has the hub nodes, and the top one has the destination nodes of the demand. Besides the flow balancing done during the selection of the arcs, their model can incorporate congestion effects at the middle layer (hub node layer).

Yaman and Carello [19] have addressed the problem of solving single allocation hub location problems with modular link capacities. They have proposed a non-linear formulation where the size of inter-hub connections are treated as stepwise, instead of the linear approach proposed by Labbé et al. [20] to the quadratic capacitated hub location problem with single assignment. In order to accomplish that, capacity restrictions have been imposed on the amount of traffic flowing through the hubs rather than only considering the incoming traffic. The problem has been solved by means of a branch-and-cut algorithm based on a linearization with an exponential number of constraints and a two-level local search method. Both methods have been then combined in a heuristic concentration (see [21,22]) scheme. Marín [23] also has addressed the capacitated multiple allocation hub location problems, but assumed that the flow between a given origin–destination pair can be split into several routes.

Another interesting work is the one presented by Kara and Tansel [24]. They have modeled the transient times at hubs in addition to the travel times, having as objective function the minimization of the latest-arrival times, and consequently, the excessive delays at the hubs. Yaman et al. [25] have presented a similar formulation for a Turkish cargo delivery system where transient times have been taken into account as well as journey times. The transient times arises from the unloading, loading and sorting operations in a hub and can be a significant portion of the total transportation time or cost. They have also considered the presence of stopovers prior to the gathering of flows on the hubs. The network routing structure is divided into main lines and express lines. The main line routes may connect many stopovers before being routed to a hub and are responsible for collecting or delivering parcels. Express lines are inter-hub connections where no stopovers are allowed. Valid inequalities have been also employed in order to strengthen the proposed formulation.

A work that stands out is the one of Marianov and Serra [26]. They have modeled the hub network as an M/D/c queuing network, proposing capacity constraints based on the probability of waiting customers in the system. Due to the computational complexity of these constraints, they have linearized them and then solved the resulting model by a tabu search algorithm. They have also proposed a model for allocating servers to each installed hub. A similar work has been proposed by Rodriguez et al. [27], but a simulated annealing algorithm has been used.

Costa et al. [28] have presented a multi-criteria formulation to the capacitated single allocation hub location problem. Besides the traditional cost minimizing function, they have considered alternatively the minimization of the time the hubs take to process the flow or the minimization of the maximum service time of the hubs. For instance, Rodriguez-Martin and Salazar-Gonzalez [29] have proposed a Benders decomposition algorithm [30] to tackle a capacitated hub location problem based on a multi-commodity formulation where the arcs connecting the hubs are not assumed to create a complete graph. Their model is similar to the one presented by Sridhar and Park [31] for the fixed-charge capacitated network design problem.

Modeling congestion through capacity constraint on the flows does not mimic the exponential nature of the congestion effects. Elhedhli and Hu [32] have been the first ones to consider the costs of the congestion effects explicitly on the objective function. Using a convex cost function that increases exponentially as more flows go through the hubs, they have proposed a non-linear model to a single assignment hub location problem. The congestion convex cost function is a power-law function of airport usage relative to its capacity and it has been widely used to estimate delay costs [33]. They have linearized their model and then solved it using Lagrangian relaxation.

In this paper, we explore further the congestion effects written as a convex cost function similarly to [32], but addressing the multiple allocation hub location problem. We propose a non-linear mixed integer programming problem based on a classical formulation [34] due to its linear programming bound quality when compared to others [9,35–37]. On our formulation the number of hubs on a route is limited to two, even if there is a route with a lower cost using more than two hubs.

Although, one of the main strategies to handle non-linear problems is to linearize them, we have developed a generalized Benders decomposition (GBD) [38] to cope with our non-linear mixed integer program. The problem is decomposed in two smaller problems: at a higher level, named as master problem (MP), the location decisions are made; while at an inferior level, known as subproblem (SP), the flow balance and the congestion are handled. The MP is an integer programming problem, while the SP is a non-linear convex transportation problem. Using our approach, we have been able to solve large problems of the CAB and AP standard data sets to optimality.

This paper is organized as follows: in Section 2, the model formulation and the notation used are introduced. In Section 3, the GBD is developed to solve the problem and it is also demonstrated how the SP can be optimally solved. Computational experiments and conclusion remarks are presented in Sections 4 and 5, respectively.

2. Model formulation

In this section the congested multiple allocation hub location problem is presented. The basic components of the model are the following sets: let N be the set of node locations that exchange traffic and let K be the set of node candidates to become hubs, $K \subseteq N$. For any pair of nodes i and j ($i,j \in N$), we have w_{ij}, the flow from origin i to destination j that is routed through either one or two installed hubs. Usually, we have $w_{ij} \neq w_{ji}$.

Further, let a_k be the fixed cost of installing a hub at node $k \in K$ and let c_{ijkm} be the transportation cost per unit of flow from node i to j routed via hubs k and m ($i,j \in N$ and $k,m \in K$). This transportation cost is the composition of three cost segments: $c_{ijkm} = c_{ik} + \alpha c_{km} + c_{mj}$, where c_{ik} and c_{mj} are the standard transportation cost per unit from location i (j) to hub k (m), and αc_{km} is the discounted transportation cost between hubs k and m. The discount factor $0 \leqslant \alpha \leqslant 1$ represents the scale economies on the inter-hub connection. If only one hub is used in a given route, we have $k=m$ and no discount factor is applied.

The following decision variables are defined:

$$y_k = \begin{cases} 1 & \text{if hub } k \in K \text{ is installed} \\ 0 & \text{otherwise} \end{cases}$$

$x_{ijkm} \geqslant 0$ is the flow from origin i to destination j ($i,j \in N$) that is routed through hubs k and m ($k,m \in K$) in that order.

In order to start the development of the congested version of the uncapacitated multiple allocation hub location problem (UMAHLP), the formulation due to Hamacher et al. [34] is stated as

$$\min \quad \sum_k a_k y_k + \sum_i \sum_j \sum_k \sum_m c_{ijkm} x_{ijkm} \tag{1}$$

$$\text{s.t.} \quad \sum_k \sum_m x_{ijkm} = w_{ij} \quad \forall i,j \in N \tag{2}$$

$$\sum_m x_{ijkm} + \sum_{m \neq k} x_{ijmk} \leqslant w_{ij} y_k \quad \forall i,j \in N, \; k \in K \tag{3}$$

$$x_{ijkm} \geqslant 0 \quad \forall i,j \in N, \; k,m \in K \tag{4}$$

$$y_k \in \{0,1\} \quad \forall k \in K \tag{5}$$