European Journal of Operational Research 220 (2012) 48–57

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

New formulation and a branch-and-cut algorithm for the multiple allocation p-hub median problem

Sergio García [a], Mercedes Landete [b,*], Alfredo Marín [c]

[a] Departamento de Estadística, Universidad Carlos III de Madrid, Spain
[b] Departamento de Estadística, Matemáticas e Informática, Centro de Investigación Operativa, Universidad Miguel Hernández de Elche, Spain
[c] Departamento de Estadística e Investigación Operativa, Universidad de Murcia, Spain

ARTICLE INFO

Article history:
Received 31 May 2011
Accepted 20 January 2012
Available online 28 January 2012

Keywords:
Hub location
Discrete location
Integer programming

ABSTRACT

This article deals with the uncapacitated multiple allocation p-hub median problem, where p facilities (hubs) must be located among n available sites in order to minimize the transportation cost of sending a product between all pairs of sites. Each path between an origin and a destination can traverse any pair of hubs.

For the first time in the literature, an integer programming formulation with $O(n^2)$ variables has been devised to approach this problem. Based on this formulation, a branch-and-cut algorithm has been developed which allows to solve larger instances than those previously solved in the literature. The proposed algorithm performs specially well for relatively large values of p.

1. Introduction

1.1. Hub location problems

In a typical hub location problem, a set of n locations—points—is given. Let $N = \{1,\ldots,n\}$ be the set of indexes of these locations. For each ordered pair of points (i,j) there is a quantity $W_{ij} \geqslant 0$ of product that needs to be sent from i to j. The key decision in this problem is which locations will be used as *hubs*, i.e., transshipment points where the product will be collected and distributed.

In the multiple allocation version of the problem, once the hubs have been fixed, each origin point i will be connected to each destination point j, $i, j \in N$, through an independent path that can traverse one or two hubs. Moreover, if neither the origin nor the destination are hubs, then the path will contain one or two *intermediate* hubs; if either the origin or the destination (but not both of them) is a hub, the flow between them either will be directly sent or will traverse only one additional hub; finally, if both i and j are hubs, then the flow will be directly sent from i to j without using any other hub. Single allocation problems, where every point must send and receive the product to and from a unique hub, have also been broadly studied in the literature, but are out of the scope of this paper.

The transportation costs of the system are calculated in the following way. Any unit of product with origin in point i and destination in point j that traverses hubs k and m in this order costs $c_{ijkm} \geqslant 0$. When $k = m$, i.e., if there is only one hub in the path, the cost is analogously denoted by c_{ijkk}. Particular cases like that in which the origin i is a hub (with associated cost c_{ijim}) do not deserve more attention.

In the *uncapacitated multiple allocation hub location problem* (HL) a fixed cost f_k associated with the transformation of point k into a hub has to be payed. The aim is to find the set of points to be transformed into hubs and the paths associated with all the origin–destination pairs such that the total cost (fixed plus transportation) is minimized. The problem we are going to deal with, the *multiple allocation p-hub median problem* (pH), only takes into account the transportation costs whereas the number of hubs must be exactly p. These two problems are closely related; in particular, most of the integer programming formulations used for one of them can be easily adapted to solve the other.

In many applications, costs c_{ijkm} are calculated using the formula

$$c_{ijkm} = d_{ik} + \alpha d_{km} + d_{mj}, \quad (1)$$

where $0 < \alpha < 1$ is a discount factor between hubs (economies of scale) and d_{uv} are costs associated with the links between points. The more general formula $c_{ijkm} = \beta d_{ik} + \alpha d_{km} + \gamma d_{mj}$, with $\beta \geqslant \alpha$ and $\gamma \geqslant \alpha$ fixed amounts associated with collection and distribution, respectively, is also frequently used in the literature. We will assume here that costs are calculated using these formulas. Moreover, we will suppose that costs d are distances or, at least, that they satisfy the triangle inequality, $d_{ii} = 0\ \forall i$ and $d_{ij} \geqslant 0\ \forall i, j$. As a consequence, the limit of two hubs per path must not be imposed

* Corresponding author. Tel.: +34 626 961780.
E-mail addresses: sergio.garcia@uc3m.es (S. García), landete@umh.es (M. Landete), amarin@um.es (A. Marín).

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2012.01.042

since it is always satisfied by at least one optimal solution of the problem.

1.2. Literature

Many papers in the literature of discrete location are devoted to the resolution of hub location problems. For an introduction to the application fields and a review of hub location, we refer the reader to [5,1], respectively. In the following we will only cite the papers that are needed for a better understanding of this article.

The first integer programming formulations for both problems, HL and pH, were presented by Campbell [3,4]. Binary variables y_k take value 1 if a hub is located at point k, whereas variables x_{ijkm} measure the fraction of W_{ij} sent through hubs k and m (in this order). The formulation for HL is

$$(\mathrm{C}) \quad \min \quad \sum_i \sum_j \sum_k \sum_m W_{ij} c_{ijkm} x_{ijkm} + \sum_k f_k y_k$$

$$\text{s.t.} \quad \sum_k \sum_m x_{ijkm} = 1 \quad \forall i,j, \tag{2}$$

$$x_{ijkm} \leqslant y_k \quad \forall i,j,k,m \tag{3}$$

$$x_{ijkm} \leqslant y_m \quad \forall i,j,k,m \tag{4}$$

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

$$x_{ijkm} \geqslant 0 \quad \forall i,j,k,m. \tag{6}$$

Since formulation (C) contains $n^4 + n$ variables and $2n^4 + n^2$ linear constraints and its linear relaxation is weak, using it becomes a hard task (see e.g. [13] for checking the difficulties when using dual methods based on (C)). Formulation (C) can be adapted to pH just by fixing $f_k = 0$ $\forall k$ and adding the constraint

$$\sum_k y_k = p. \tag{7}$$

We will call (C_p) this formulation for pH. Note that in all sums and sets without any other specification, the indexes will vary from 1 to n.

Skorin-Kapov et al. [16] and O'Kelly et al. [15] tightened constraints (3) and (4) in the context of pH, replacing them by

$$\sum_m x_{ijkm} \leqslant y_k \quad \forall i,j,k, \tag{8}$$

$$\sum_k x_{ijkm} \leqslant y_m \quad \forall i,j,m. \tag{9}$$

By doing this replacement in formulation (C), the number of linear constraints reduces from $2n^4 + n^2$ to $2n^3 + n^2$ and much better lower bounds can be obtained using the linear relaxation associated with the new formulation.

Cánovas et al. [7] and Hamacher et al. [12] carried out the definitive reduction of constraints (8) and (9), replacing them by

$$\sum_k x_{ijkm} + \sum_{k \neq m} x_{ijmk} \leqslant y_k \quad \forall i,j,k, \tag{10}$$

so reducing the number of linear constraints to $n^3 + n^2$ and getting even better bounds. We will call (CH) this formulation and (CH_p) the corresponding version for pH. Recent results on the resolution of HL by means of (CH) can be consulted in [6,14], and very recent results are available in [8], which succeeds in solving large scale instances of HL using a Benders decomposition approach based on (CH).

Concerning pH, it was formulated in [10,11] using three-index variables. Their formulation, designed for transportation costs calculated using (1), is

$$(\mathrm{EK}) \quad \min \quad \sum_i \sum_k d_{ik} z_{ik} + \sum_i \sum_k \sum_m \alpha d_{km} h_{ikm} + \sum_i \sum_j \sum_m d_{mj} n_{ijm}$$

$$\text{s.t.} \quad \sum_k z_{ik} = O_i \quad \forall i,$$

$$\sum_m n_{ijm} = W_{ij} \quad \forall i,j,$$

$$\sum_m h_{i\ell m} + \sum_j n_{ij\ell} = z_{i\ell} + \sum_k h_{ik\ell} \quad \forall i,\ell,$$

$$z_{ik} \leqslant O_i y_k \quad \forall i,k,$$

$$\sum_i n_{ijm} \leqslant D_j y_m \quad \forall j,m,$$

$$\sum_k y_k = p,$$

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

$$n_{ijm}, h_{ikm}, z_{ik} \geqslant 0 \quad \forall i,j,k,m,$$

where $O_i = \sum_j W_{ij},\ i \in N$, are the units of product which must be sent from i, $D_j = \sum_i W_{ij},\ j \in N$, are the units of product which must be sent to j, and the three families of new variables have the following meaning:

- z_{ik} is the amount of product going from origin i directly to hub k,
- h_{ikm} is the amount of product going from origin i directly to hub k and then directly to point m, and
- n_{ijm} is the amount of product going from origin i to any hub and then to hub m and finally to destination j.

1.3. The challenge

Formulations for HL and pH based on (C_p) use $O(n^4)$ variables and, at least, $O(n^3)$ constraints. Although reduced versions of (C_p) like (CH_p) provide us with very good lower bounds and can be used to solve medium sized instances (especially in the case of HL), they are huge and require large amounts of memory and computational effort. On the other hand, formulations based on three-index variables derived from (EK), while smaller, give poorer lower bounds. The aim of the research line that gave rise to this article was to develop, improve and use small formulations to solve pH, problem which has not been solved in the literature so efficiently as HL. In particular, we wanted to obtain a formulation with $O(n^2)$ variables that could be used to solve large instances. In this paper we present such a formulation. Using it, we have been capable to solve to optimality: (i) instances previously solved in the literature with up to 40 points, many of them needing similar computational times, (ii) larger instances with 50 points and any value of p, (iii) very large instances with up to 200 points when the number of hubs is large enough.

To the best of our knowledge, this is the first time that a formulation with $O(n^2)$ variables has been applied to the multiple allocation p-hub location problem. However, there has been previous work on a similar formulation for the single allocation p-hub median problem (see [9]).

The rest of the paper is organized as follows. In Section 2 we present a new reduced formulation for pH. Based on the valid inequalities generated in Section 3, we strengthen and reduce even more the initial formulation in Section 4. All these improvements are embedded into a branch-and-cut algorithm in Section 5. The solution method is tested in Section 6 and, finally, some conclusions are drawn in Section 7.