Improved Benders Decomposition for Capacitated Hub Location Problem with Incomplete Hub Networks

Yifan Xu*, Weibin Dai†, Xiaoqian Sun‡ and Sebastian Wandelt§
Beijing Key Laboratory for Network-based Cooperative ATM
School of Electronic and Information Engineering, Beihang University
Beijing, China
Email: *yifan94125@sina.com, †daiweibin@buaa.edu.cn, ‡sunxq@buaa.edu.cn, §wandelt@informatik.hu-berlin.de

***Abstract*—The hub location problem (HLP) has been studied by researchers for many years. A number of model variants and solution techniques for solving the problem have been proposed. Most researchers consider the uncapacitated HLP(UHLP), given the difficulty in computation that comes with capacity constraints. Particularly, together with incomplete hub networks, capacity constraints have shown to be highly intractable. We develop a novel, efficient Benders decomposition algorithm to solve the CHLP with incomplete hub networks. In order to explore the impact of capacity constraints on hubs and backbone arcs, the CAB dataset is used as a case study. In addition, we compare the performance of our improved algorithm to the classical one. We find that capacity constraints on hubs and backbone links tend to render a robust network with more fully connected hub node pairs and flexible linking structure. In addition, the computation time is significantly reduced, up to one order of magnitude, compared with the state-of-the-art. We believe that our work lays the foundation for solving more realistic hub location problems.**

***Index Terms*—Network design, Capacitated hub location problem, Benders decomposition**

I. Introduction

Since the pioneering work by O'Kelly [1], hub location problems (HLPs) have gained popularity among various research areas such as air transportation [2] [3], telecommunication [4] and trucking industries [5]. The role of hubs is especially evident in terms of air transportation where it is uneconomical to serve the demand by assigning a flight between every city pair. Through intermediate hubs, traffic can be redirected, aggregated and disaggregated [6], which in turn can reduce operational costs and exploits economies of scale in inter-hub links.

The main decisions of HLPs involve the location of hub nodes and paths for sending flows between origin-destination (OD) pairs. While the topology of a network constituted by both hub nodes and non-hub nodes depends on applications, there are some common assumptions that characterize most HLPs, such as the eight protocols proposed by O'Kelly [7] based on three options:

- Multiple/Single allocation: whether each non-hub node can be connected to more than one hub.
- Direct/Indirect OD arcs: whether arc between every two non-hub OD nodes can be established.
- Complete/Incomplete hub network: whether all hub nodes are fully interconnected.

Apart from these mathematical assumptions, one important property is the capacity restriction which divides HLPs into Uncapacitated HLPs (UHLPs) and Capacitated HLPs (CHLPs). Although CHLPs are more pragmatic, additional capacity constraints make it difficulty to solve the problem. Models for CHLPs must be delicately formulated and efficient algorithms are needed. Recent works like [8] studied the CHLP with the single assignment, where they used Lagrangian relaxation to solve an instance of over 200 nodes. The model presented by [9] also assumes the single allocation. They used memory structure to devise heuristic methods for CHLP with modular links which are originally proposed by Yaman [10]. Correia [11] took demand uncertainty into account for the multi-period stochastic CHLP with the multiple assignment. The readers are referred to [12] for more previous papers on CHLPs.

For most papers on CHLPs, the subgraph constructed by hub nodes is complete. While this assumption exploits the possible lowest transportation cost, it could be counterproductive in some cases. Specifically, when great circle routes do not follow triangular inequality and few demand needs to be transported through two distant hubs, using intermediate hubs to serve the flow may be better. Therefore, incomplete hub networks allows for more practical linking structure. As for CHLP with incomplete hub network, Rodríguez-Martín [13] proposed a model and two algorithms for exact and heuristic solution. Based on this model, Kratica [14] modified the formulations to get a more compact model and two evolutionary algorithms are devised for large-scale instances. In the incomplete UHLP presented by Alumur [15], decision variables are related to the installation of links can be extended to CHLP with constraints on inter-hub links. The multi-product CHLP model proposed by Correia [16] features an incomplete hub network aggregated by the complete network for each product. Although the assumption about complete hub network is relaxed in these studies, their models didn't incorporate fixed cost for installing hub and arcs which could result in a more flexible and generalized network structure in response to cost pressures [17].

In particular, this paper targets at CHLP with an incomplete hub network. The objective function of our model considers a set of fixed and variable cost components which was introduced by O'Kelly [17]. The capacity limitation is imposed on both hubs and inter-hub links to restrict the number of non-hub nodes allocated to each hub and the number of installed inter-hub links. These features not only enable airline operators to devise appropriate flight schedule by calibrating cost parameters but also make the mathematical formulation difficult to solve, when there are more integer decision variables and fewer constraints on network topology.

Despite great efforts have been made, it remains challenging to solve CHLPs. This difficulty is boosted when a more general model with capacity constraints is considered. Because our model's formulations are amenable to decomposition method, we use the Benders decomposition algorithm to solve this problem. This algorithm has been proposed by Benders [18] to solve problems with complicated variables. In general, this technique partitions the whole problem into a master problem (MP) and a sub problem (SP). The algorithm proceeds with iterations between MP and SP until the gap between lower bound (LB) and upper bound (UB) reaches a given threshold.

Since the introduction of Benders Decomposition, it has achieved success in many fields such as production planning [19] and airline scheduling [20]. While most researches aim at its application, some general improvements have been made to enhance its efficiency [21]. Because the convergence rate is closely related to the optimality or feasibility cuts generated from SP, several methods are explored to enhance these cuts. Magnanti [22] proposed the Pareto-optimality cut to find the strongest cut for degeneracy. Thereafter, Mercier [23] and Papadakos [24] relaxed the restriction for using this cut. In contrast, Mercier [25] and Camargo [26] proposed theories and models for feasibility cuts. We use a problem based on Mercier's [25] model for stable and efficient Benders cuts.

The remainder of the paper is organized as follows. Section II provides definition and notations for CHLP. In section III, an improved Benders decomposition scheme is devised to speed up the overall convergence rate. Section IV compares the optimal topology under different capacity scenarios and reports the computational result of the improved Benders decomposition algorithm with the classical one. Finally, conclusions are made in Section V.

II. Model Formulation

In this section, we formulate the CHLP features incomplete hub network. In addition, non-hub nodes can be interconnected and capacity constraints are imposed on both hub nodes and inter-hub links. The cost coefficients for constructing the network and routing OD demand which is proposed by O'Kelly [17] are used as a starting point.

Given N as the number of demand nodes including hub nodes and non-hub nodes, three kinds of arcs are established in the network. Specifically, direct arcs are only allowed to connect and serve the demand between two non-hub nodes. Non-hub nodes are allocated to hub nodes through tributary arcs, and we differentiate arcs starting or ending at non-hub nodes as starting tributary or ending tributary arcs. While the flow between two hub nodes is transmitted on backbone arcs. Symbols *0,1,2,3* are used to decide these aforementioned arcs: direct, starting tributary, ending tributary and backbone arcs.

The demand from origin i to destination j is denoted by w_{ij}. The distance between node i and node j is represented by c_{ij}. Some other parameters are defined as follows: f_u^H is the fixed cost for installing a hub facility in node u, f^0, f^1, f^2, f^3 and v^0, v^1, v^2, v^3 are fixed cost and variable cost for four arcs respectively. Parameter A_{ij} refers to the arc specific fixed cost to connect i and j. $\tilde{\Gamma}_u^H$ and $\tilde{\Gamma}_{ij}^L$ denote capacity restriction on hub node u and backbone arc *(i,j)*.

Considering that customers' dissatisfaction may be incurred during transfer flight, we define S as the maximum steps for traversing from origin to destination. Although it is common in previous papers that specify the number p of hubs to be installed to ensure at most $p+1$ steps for each OD path, such assumption may result in poor economic performance with more backbone arcs. To clearly explain our model, an OD pair *(i,j)* is illustrated in Fig 1. In this case, S is set to 5. The OD flow starts from non-hub node i, goes through hub nodes u and v, and ends at non-hub node j. For normalization purpose, we establish two dummy hub nodes v' and v'' to stretch the path length to S. Because the distance of *(v,v')* and *(v',v")* is zero, these dummy hub nodes do not ichange the objective value of the model. In this way, we propose the

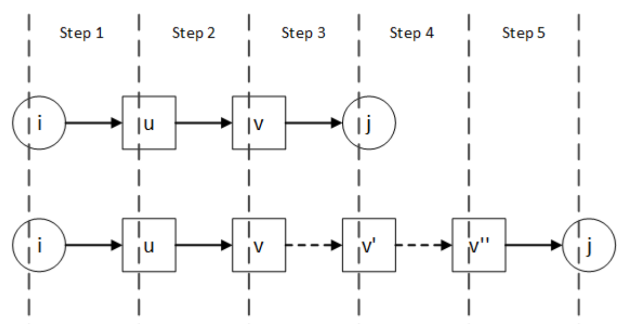


Fig. 1. Dummy graph with S=5

model for CHLP over incomplete hub network. The decision variables are $z_k, y_{ij}^0, y_{iu}^1, y_{vj}^2, y_{uv}^3$, binary variables that decide whether to install hubs, direct arcs, starting tributary arcs, ending tributary arcs, backbone arcs and x_{uvs}^{ij}, continuous variable that indicates the fraction of flow going through node u, v at step s for OD pair *(i,j)*. With some adaptation from [26] the problem is formulated as follows:

$$\min \sum_{u\in N} f_u^H z_u + \sum_{i\in N}\sum_{\substack{j\in N\\ i\neq j}} \hat{c}_{ij}^0 y_{ij}^0 + \sum_{i\in N}\sum_{\substack{u\in N\\ i\neq u}} \hat{c}_{iu}^1 y_{iu}^1 + \sum_{v\in N}\sum_{\substack{j\in N\\ v\neq j}} \hat{c}_{vj}^2 y_{vj}^2 + \sum_{u\in N}\sum_{\substack{v\in N\\ u\neq v}} \hat{c}_{uv}^3 y_{uv}^3 + \sum_{i\in N}\sum_{\substack{j\in N\\ i\neq j}} w_{ij}\Big(\sum_{u\in N} \check{c}_{iu}^1 x_{iu1}^{ij} + \sum_{v\in N} \check{c}_{vj}^2 x_{vjS}^{ij} + \sum_{u\in N}\sum_{\substack{v\in N\\ u\neq v}} \sum_{s=2}^{(S-1)} \check{c}_{uv}^3 x_{uvs}^{ij}\Big) \quad (1)$$

$$s.t. \sum_{v\in N} x_{vjS}^{ij} + y_{ij}^{0} = 1 \qquad \forall i,j \in N : i \neq j \tag{2}$$

$$x_{uv1}^{ij} \leq y_{uv}^{1} \qquad \forall i,j,u,v \in N : i \neq j, u \neq v \tag{3}$$

$$x_{uvS}^{ij} \leq y_{uv}^{2} \qquad \forall i,j,u,v \in N : i \neq j, u \neq v \tag{4}$$

$$\sum_{s=2}^{(S-1)} x_{uvs}^{ij} \leq y_{uv}^{3} \qquad \forall i,j,u,v \in N : i \neq j, u \neq v \tag{5}$$

$$x_{ii1}^{ij} = z_i \qquad \forall i,j \in N : i \neq j \tag{6}$$

$$\sum_{\substack{v\in N\\ v\neq u}} \sum_{s=1}^{(S-1)} x_{vus}^{ij} \leq z_u \qquad \forall i,j,u \in N : i \neq j, u \neq i, u \neq j \tag{7}$$

$$x_{jjS}^{ij} = z_j \qquad \forall i,j \in N : i \neq j \tag{8}$$

$$\sum_{\substack{i\in N\\ i\neq u}} \sum_{\substack{j\in N\\ i\neq j}} \Big(x_{iu1}^{ij} + \sum_{\substack{v\in N\\ v\neq u}} \sum_{s=2}^{S-1} x_{vus}^{ij} \Big) w_{ij} \leq \tilde{\Gamma}_{u}^{H} \qquad \forall u \in N \tag{9}$$

$$\sum_{i\in N} \sum_{\substack{j\in N\\ i\neq j}} \sum_{s=2}^{S-1} x_{uvs}^{ij} w_{ij} \leq \tilde{\Gamma}_{uv}^{L} \qquad \forall u,v \in N : u \neq v \tag{10}$$

$$x_{ii1}^{ij} = \sum_{v\in N} x_{iv2}^{ij} \qquad \forall i,j \in N : i \neq j \tag{11}$$

$$x_{iu1}^{ij} = \sum_{\substack{v\in N\\ v\neq i}} x_{uv2}^{ij} \qquad \forall i,j,u \in N : i \neq j, u \neq i, u \neq j \tag{12}$$

$$x_{ij1}^{ij} = x_{jj2}^{ij} \qquad \forall i,j \in N : i \neq j \tag{13}$$

$$x_{iis}^{ij} = \sum_{v\in N} x_{ivs+1}^{ij} \qquad \forall i,j \in N, s \in \{2\ldots(S-2)\} : i \neq j \tag{14}$$

$$\sum_{\substack{v\in N\\ v\neq j}} x_{vus}^{ij} = \sum_{\substack{v\in N\\ v\neq i}} x_{uvs+1}^{ij} \qquad \forall i,j,u \in N, \quad s \in \{2\ldots(S-2)\} : i \neq j, u \neq i, u \neq j \tag{15}$$

$$\sum_{v\in N} x_{vjs}^{ij} = x_{jjs+1}^{ij} \qquad \forall i,j \in N, s \in \{2\ldots(S-2)\} : i \neq j \tag{16}$$

$$x_{iiS-1}^{ij} = x_{ijS}^{ij} \qquad \forall i,j \in N : i \neq j \tag{17}$$

$$\sum_{\substack{v\in N\\ v\neq j}} x_{vuS-1}^{ij} = x_{ujS}^{ij} \qquad \forall i,j,u \in N : i \neq j, u \neq i, u \neq j \tag{18}$$

$$\sum_{v\in N} x_{vjS-1}^{ij} = x_{jjS}^{ij} \qquad \forall i,j \in N : i \neq j \tag{19}$$

$$z_i, y_{ij}^{0}, y_{ij}^{1}, y_{ij}^{2}, y_{ij}^{3} \in \{0,1\} \qquad \forall i,j \in N : i \neq j \tag{20}$$

$$x_{uvs}^{ij} \geq 0 \qquad \forall i,j,u,v \in N, s \in \{1\ldots S\} : i \neq j, u \neq v \tag{21}$$

where $\hat{c}_{ij}^{0} = c_{ij}(f^0 + A_{ij} + v^0 w_{ij})$, $\hat{c}_{iu}^{1} = c_{iu}(f^1 + A_{iu})$, $\hat{c}_{vj}^{2} = c_{vj}(f^2 + A_{vj})$, $\hat{c}_{uv}^{3} = c_{uv}(f^3 + A_{uv})$, $\check{c}_{iu}^{1} = c_{iu}v^1$, $\check{c}_{vj}^{2} = c_{vj}v^2$, $\check{c}_{uv}^{3} = c_{uv}v^3$. Objective function (1) aims to minimize the total cost including fixed installation cost for arcs and hubs, variable cost for transporting demand through arcs. All the flow paths are divided into *S* steps. In *s=1*, flow starts from its origin *i* and end at its destination *j* when *s=S*.

In addition, dummy hubs would be established to ensure that exact *S* arcs are constructed for every OD path. Constraints (2)–(8) describe the necessary installation of hubs and arcs for feasible paths. Without loss of generality, we specify that flow can stay at origin *i* for *s=1* or destination *j* only when *i* and *j* are hubs. Constraints (9)–(10) are the capacity constraints for hub nodes and backbone links. Constraints (11)–(19) state all the possible coherent connection for *s=1*, *s=2*… *S-1* and *s=S* respectively. To be specific, once a flow departures from its origin *i* to other nodes (instead of *i*) or arrives at its ultimate destination *j*, it is impossible for the flow to return to *i* or leave *j* (for other nodes rather than *j*). In other words, those circumstances are irreversible. Besides, flow paths are always successive: from *u* to *v* in step *s*, and leave *v* for other nodes in step *s+1*.

After the introduction of the five-index decision variable x_{uvs}^{ij} and infrastructure constraints (2)–(8), we are able to propose reasonable linear capacity constraints (9)–(10) as compared to the quadratic constraint on hub capacity proposed by [10]. On the other hand, the model contains $(4N^2 - 3N)$ integer variables, $(N^4 S - N^3 S)$ continuous variables and $\big(3n^4 + (S-6)n^3 + (5-S)n^2 - n\big)$ constraints, which is a challenge for traditional branch-and-cut algorithm. As decomposition methods are effective for those large-scale mixed integer programs (MIP), an improved Benders decomposition algorithm is presented to solve this NP-hard problem.

III. Methodology for Benders Decomposition

Benders decomposition has been successfully applied to combinatorial optimization problems. Its main idea is to decompose a complex model into a master problem (MP) and sub-problem (SP). Because the variables of MP can be treated as known number in constraints when solving SP, the dual problem of SP features an objective function containing variables from MP which can be used as the Benders cuts to be added to MP's constraint pool. In this way, a convincing optimal result can be obtained from the iterative interaction between the solution from MP and DSP. Many variables in the original problem are replaced by constraints with only part of the complicated variables. In this section, the original CHLP with incomplete hub network model is partitioned into an MP and an SP whose dual problem and related Benders cuts are presented here. Benders decomposition algorithm is further improved through optimality Benders cuts and feasibility Benders cuts with a bounded dual SP which generates feasibility cuts of high quality over the traditional practice.

A. The Master Problem

The MP of our CHLP model aims to calculate decision variables of the original problem. MP contains variables for installation of hubs and arcs (z, y^0, y^1, y^2, y^3). Some additional constraints are added to speed up the convergence. Moreover, a new variable η is introduced to estimate the objective value of SP. In this way, MP is formulated as follows:

$$\min \sum_{i\in N}\sum_{\substack{j\in N\\ i\neq j}} \eta_{ij} + \sum_{u\in N} f_u^H z_u + \sum_{i\in N}\sum_{\substack{j\in N\\ i\neq j}} \hat{c}_{ij}^0 y_{ij}^0 + \sum_{i\in N}\sum_{\substack{u\in N\\ i\neq u}} \hat{c}_{iu}^1 y_{iu}^1 + \sum_{v\in N}\sum_{\substack{j\in N\\ v\neq j}} \hat{c}_{vj}^2 y_{vj}^2 + \sum_{u\in N}\sum_{\substack{v\in N\\ v\neq u}} \hat{c}_{uv}^3 y_{uv}^3 \quad (22)$$

$$s.t.\ y_{ij}^0 + z_i \leq 1 \qquad \forall i,j \in N : i\neq j \quad (23)$$

$$y_{ij}^0 + z_j \leq 1 \qquad \forall i,j \in N : i\neq j \quad (24)$$

$$y_{iu}^1 + z_i \leq 1 \qquad \forall i,u \in N : i\neq u \quad (25)$$

$$y_{iu}^1 \leq z_u \qquad \forall i,u \in N : i\neq u \quad (26)$$

$$y_{vj}^2 \leq z_v \qquad \forall v,j \in N : v\neq j \quad (27)$$

$$y_{vj}^2 + z_j \leq 1 \qquad \forall v,j \in N : v\neq j \quad (28)$$

$$y_{uv}^3 \leq z_u \qquad \forall u,v \in N : u\neq v \quad (29)$$

$$y_{uv}^3 \leq z_v \qquad \forall u,v \in N : u\neq v \quad (30)$$

$$\sum_{\substack{j\in N\\ i\neq j}} \left(y_{ji}^0 + y_{ij}^0 + y_{ji}^1 + y_{ij}^1 + y_{ji}^2 + y_{ij}^2 + y_{ji}^3 + y_{ij}^3\right) \geq 2 \qquad \forall i \in N \quad (31)$$

$$y_{ij}^0 + \sum_{\substack{u\in N\\ i\neq u}} \left(y_{iu}^1 + y_{iu}^3 + y_{ij}^2\right) \geq 1 \qquad \forall i,j \in N : i\neq j \quad (32)$$

$$y_{ij}^0 + y_{ij}^1 + \sum_{\substack{v\in N\\ v\neq j}} \left(y_{vj}^2 + y_{vj}^3\right) \geq 1 \qquad \forall i,j \in N : i\neq j \quad (33)$$

Constraints (23)–(24) specify that direct arcs are established between two non-hub nodes. Likewise, constraints (25)–(30) set the condition for installing tributary and backbone arcs. With constraint (31), each node is connected with at least 2 arcs. Constraints (32)–(33) guarantee that there exist arcs to connect its origin and destination for each OD pair.

B. The Sub Problem

MP is a relaxed form of the original CHLP model, an LB is expected to be obtained when MP is solved. The solution which contains the values of complicating variables $(\overline{z}, \overline{y}^0, \overline{y}^1, \overline{y}^2, \overline{y}^3)$ will be transferred to the SP for further calculation. Moreover, since DSP has an objective function with these complicated variables which benefits the generation of Benders cuts, a DSP formulation is presented:

TABLE I
VARIABLES FOR DSP FROM CHLP

Variables	Domain	Description
α_{us}^{ij}	$\mathbb{R}$	$\forall i,j,u \in N, s\in\{1\dots S\} : i\neq j$ dual variables from constraints (2), (11)–(19)
β_{uv}^{ij}	$[0,+\infty)$	$\forall i,j,u,v \in N : i\neq j, u\neq v$ dual variables from constraint (3)
γ_{uv}^{ij}	$[0,+\infty)$	$\forall i,j,u,v \in N : i\neq j, u\neq v$ dual variables from constraint (4)
δ_{uv}^{ij}	$[0,+\infty)$	$\forall i,j,u,v \in N : i\neq j, u\neq v$ dual variables from constraint (5)
ϵ_u^{ij}	$[0,+\infty)$	$\forall i,j,u \in N : i\neq j, u\neq i, u\neq j$ dual variables from constraint (7)
θ_{ij}	$\mathbb{R}$	$\forall i,j \in N : i\neq j$ dual variables from constraint (6)
ζ_{ij}	$\mathbb{R}$	$\forall i,j \in N : i\neq j$ dual variables from constraint (8)
ρ_u	$[0,+\infty)$	$\forall u \in N$ dual variables from constraint (9)
φ_{uv}	$[0,+\infty)$	$\forall u,v \in N : u\neq v$ dual variables from constraint (10)

$$\max \sum_{i\in N}\sum_{\substack{j\in N\\ i\neq j}} \Big((1-\overline{y}_{ij}^0)\alpha_{jS}^{ij} - \overline{z}_i\theta_{ij} - \sum_{\substack{u\in N\\ u\neq i\\ u\neq j}} \overline{z}_u\epsilon_u^{ij} - \overline{z}_j\zeta_j^{ij} - \tilde{\Gamma}_{ij}^L\varphi_{ij} - \sum_{u\in N}\sum_{\substack{v\in N\\ u\neq v}} (\overline{y}_{uv}^1\beta_{uv}^{ij} + \overline{y}_{uv}^2\gamma_{uv}^{ij} + \overline{y}_{uv}^3\delta_{uv}^{ij})\Big) - \sum_{u\in N}\tilde{\Gamma}_u^H\rho_u \quad (34)$$

$$s.t.\ \alpha_{i1}^{ij} - \theta^{ij} \leq 0 \qquad \forall i,j \in N : i\neq j \quad (35)$$

$$\alpha_{u1}^{ij} - \epsilon_u^{ij} - \beta_{iu}^{ij} - w_{ij}\rho_u \leq \check{c}_{iu}^1 \qquad \forall i,j,u \in N : i\neq j, u\neq i, u\neq j \quad (36)$$

$$\alpha_{j1}^{ij} - \beta_{ij}^{ij} - w_{ij}\rho_j \leq \check{c}_{ij}^1 \qquad \forall i,j \in N : i\neq j \quad (37)$$

$$\alpha_{jS}^{ij} - \alpha_{iS-1}^{ij} - \gamma_{ij}^{ij} \leq \check{c}_{ij}^2 \qquad \forall i,j \in N : i\neq j \quad (38)$$

$$\alpha_{jS}^{ij} - \alpha_{uS-1}^{ij} - \gamma_{uj}^{ij} \leq \check{c}_{uj}^2 \qquad \forall i,j,u \in N : i\neq j, u\neq i, u\neq j \quad (39)$$

$$\alpha_{jS}^{ij} - \alpha_{jS-1}^{ij} - \zeta_{ij} \leq 0 \qquad \forall i,j \in N : i\neq j \quad (40)$$

$$\alpha_{vs}^{ij} - \alpha_{us-1}^{ij} - \epsilon_v^{ij} - \delta_{uv}^{ij} - w_{ij}\rho_v - w_{ij}\varphi_{uv} \leq \check{c}_{uv}^3 \quad \forall i,j,u,v \in N, s\in\{2\dots(S-1)\} : i\neq j, v\neq i, v\neq j, u\neq j, u\neq v \quad (41)$$

$$\alpha_{js}^{ij} - \alpha_{us-1}^{ij} - \delta_{uj}^{ij} - w_{ij}\rho_j - w_{ij}\varphi uj \leq \check{c}_{uj}^3 \qquad \forall i,j,u \in N, s\in\{2\dots(S-1)\} : i\neq j, u\neq j \quad (42)$$

$$\alpha_{us}^{ij} - \alpha_{us-1}^{ij} \leq 0 \qquad \forall i,j,u \in N, s\in\{2\dots(S-1)\} : i\neq j \quad (43)$$

The decision variables which are associated to constraints in the original problem and they are illustrated in Table I. During Benders decomposition, DSP is solved iteratively. Whenever a bounded solution status is achieved, an optimality cut will be added to the MP's constraints pool:

$$\eta_{ij} \geq (1-y^0_{ij})\overline{\alpha}^{ij}_{jS} - z_i\overline{\theta}_{ij} - \sum_{\substack{u\in N\\ u\neq i\\ u\neq j}} z_u\overline{\epsilon}^{ij}_u - z_j\overline{\zeta}_{ij} - [\tilde{\Gamma}^H_i/(N-1)]\overline{\rho}_i - \tilde{\Gamma}^L_{ij}\overline{\varphi}_{ij} - \sum_{u\in N}\sum_{\substack{v\in N\\ u\neq v}}\Big(y^1_{uv}\overline{\beta}^{ij}_{uv} + y^2_{uv}\overline{\gamma}^{ij}_{uv} + y^3_{uv}\overline{\delta}^{ij}_{uv}\Big) \qquad \forall i,j\in N: i\neq j \tag{44}$$

When DSP turns out to be unbounded, a feasibility cut will be generated from the extreme ray:

$$0 \geq (1-y^0_{ij})\overline{\alpha}^{ij}_{jS} - z_i\overline{\theta}_{ij} - \sum_{\substack{u\in N\\ u\neq i\\ u\neq j}} z_u\overline{\epsilon}^{ij}_u - z_j\overline{\zeta}_{ij} - [\tilde{\Gamma}^H_i/(N-1)]\overline{\rho}_i - \tilde{\Gamma}^L_{ij}\overline{\varphi}_{ij} - \sum_{u\in N}\sum_{\substack{v\in N\\ u\neq v}}\Big(y^1_{uv}\overline{\beta}^{ij}_{uv} + y^2_{uv}\overline{\gamma}^{ij}_{uv} + y^3_{ij}\overline{\delta}^{ij}_{uv}\Big) \qquad \forall i,j\in N: i\neq j \tag{45}$$

Where the vector $(\overline{\alpha}^{ij}_{jS}, \overline{\theta}_{ij}, \overline{\epsilon}^{ij}_u, \overline{\zeta}_{ij}, \overline{\rho}_i, \overline{\varphi}_{ij}, \overline{\beta}^{ij}_{uv}, \overline{\gamma}^{ij}_{uv}, \overline{\delta}^{ij}_{uv})$ in constraints (44)–(45) represents the extreme point and the direction of extreme ray respectively. With the result of bounded SP and MP, a UB is obtained. The gap between LB and UB is checked for each iteration and this determines the occasion to terminate the algorithm.

C. Performance Improvement for Benders Decomposition

The speed of Benders decomposition algorithm depends, to a large extent, on the quality of optimality cuts along with feasibility cuts. Good cuts enable the global optimality to be reached within fewer iterations. Frequently, DSP is solved to be bounded or unbounded which is caused by the value of complicated variable *y*. When the feasible region of DSP is bounded, there may be multiple extreme points leading to the optimal result (i.e. DSP is degenerate). In contrast, if an unbounded solution occurs, it is quite challenging to find the direction of the extreme ray from a bunch of rays. This paper presented a problem for obtaining Pareto-optimality cuts and extreme ray direction. For the sake of simplicity, we rewrite the MIP as $\min\ c'x + f'y : Ax + By \geq b, x \geq 0, y \in \mathbf{Y}$. Here the size of all vectors and matrices are matchable. Introducing a dual variable α, one can get the DSP: $\max(b - B\overline{y})'\alpha : A'\alpha \leq c, \alpha \geq 0$. To overcome the aforementioned difficulties, [22] proposed the well-known Magnanti-Wong point to pick up the strongest optimality cuts (i.e. Pareto-optimality cuts) for DSP's degeneracy. While Papadakos [24] provided a more practical definition of Magnanti-Wong point and thus changed the direction of cuts with Pareto-optimal SP: $\max\ (b - By^{MW})'\alpha : A'\alpha \leq c, \alpha \geq 0$. Here y^{MW} is the Magnanti-Wong point which can be updated during every iteration: $y^{MW} = (1-\lambda)y^{MW} + \lambda\overline{y}$. If $\overline{y}$ leads to a bounded solution of SP, then 0.5 is the most effective value for λ. Obviously, once an unbounded solution is achieved for a y^{MW}, it is proved to be invalid for generating sequential Magnanti-Wong points [27]. Thus an auxiliary SP has been proposed for getting an appropriate λ:

$$\max\ \lambda \tag{46}$$

$$s.t.\ Ax \geq b - B[(1-\lambda)y^{MW} + \lambda\overline{y}] \tag{47}$$

$$0 \leq \lambda \leq 0.5 \tag{48}$$

Algorithm 1 Benders decomposition with Branch-and-Cut.

```
Initialize UB=+∞,LB=0, GAP=+∞,y^MW.
while GAP>0 do
    Solve Benders Pareto-Optimality SP(dual)
    Add optimality BCs (44) to MP
    LB←Solve MP
    DSP* ← Solve Benders DSP
    if DSP*optimal, then
        λ* = 0.5
        Add optimality BCs (44) to MP
        Update UB
        GAP=(UB−LB)/LB
    else
        (λ*, Ray)←Solve Benders BDSP
        Add feasibility BCs (45)to MP
        λ =min{0.5, λ*}
    end if
    Update Magnanti-Wong point
end while
```

In this paper, we transform the model proposed by [27] to match the Bounded DSP (BDSP) model presented by [25]. Without loss of generality, the domain of λ is relaxed:

$$\min\ -\lambda \tag{49}$$

$$s.t.\ Ax + B(\overline{y} - y^{MW})\lambda \geq b - By^{MW} \tag{50}$$

$$0 \leq \lambda \leq 1 \tag{51}$$

As Mercier [25] proved that the BDSP is always bounded and its extreme point corresponds to exactly the direction of the extreme ray of DSP, given that a unique variable is inserted into each constraint artificially. In constraint (50), the term $B(\overline{y} - y^{MW})\lambda$ is the artificial variable we introduced to the SP. By solving its dual problem, both the value of λ and the extreme ray's direction can be measured explicitly. The BDSP takes the form like:

$$\max\ (b - By^{MW})'\alpha - \beta \tag{52}$$

$$s.t.\ A'\alpha \leq 0 \tag{53}$$

$$[B(\overline{y} - y^{MW})]'\alpha - \beta \leq -1 \tag{54}$$

$$\alpha, \beta \leq 0 \tag{55}$$

Based on the improvements on generating feasibility cuts and updating Magnanti-Wong point with BDSP, an improved Benders decomposition is presented in Algorithm 1. Magnanti-Wong point is initialized with a convex combination of *N+1* feasible solutions [27]. During every iteration, Pareto-optimality SP is first solved with Magnanti-Wong points for Pareto-optimality cuts. Then the values of variables in MP are used to update the DSP's objective function. If DSP has optimal solutions(i.e. extreme points), λ^* is assigned with *0.5* for further updating Magnanti-Wong points. While DSP has unbounded solutions, BDSP will be invoked for calcualting

λ^* and the direction of extreme rays. Optimality cuts and feasibility cuts are generated from extreme points and extreme rays respectively. This iterating process terminates when *GAP* reaches *0*.

IV. Computational Experiments

This section reports the experiment results for solving CHLPs to explore the impact of hub and backbone arc capacity constraints on the optimal network topology. The quality of feasibility cuts generated from BDSP and traditional practice are also compared. All the experiments are carried out on a Core-i7 6500U processor and CPLEX 12.6 is used as the solver. We test the mathematical formulation and Benders decomposition algorithm on the Civil Aeronautics Board (CAB) dataset with *N*=15,20,25. The CAB dataset contains Euclidean distance and air passenger traffic flow between 25 cities in the U.S. Because cost and capacity values are not provided in CAB dataset as compared to AP(Australia Post) dataset, we assign the value for f, b, A randomly. Finally, to avoid the possible unscaled infeasibility error when solving DSP with modern solvers, it is also important to rescale parameters in capacity constraints.

A. Topology Comparison

The first experiment aims to explore the different topology under varying constraints. Four categories of CHLP are presented:

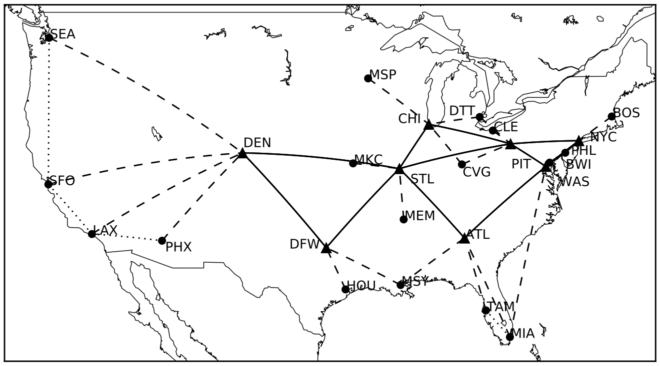
- without capacity constraints
- with hub capacity constraints
- with backbone arc capacity constraints
- with hub and backbone arc capacity constraints.

Here the CAB dataset with 25 nodes is used and the maximum number of steps is *5*. Fix and variable costs are set to be [2500,2500,2500,2500],[0.04,0.04,0.04,0.04] respectively. As for arc specific cost, $A_{ij} = 1$ if $c_{ij} \leq 1500$ and *2* otherwise. Fixed cost for installing hubs is set to be *3118240*. Capacity values for hub and backbone arc are *32e5* and 9.2e5, respectively. The results are shown in Fig 2. For all these figures, the bold triangles and circles denote hub and non-hub nodes, separately. Solid lines are backbone arcs, dash lines are tributary arcs and dot lines are direct arcs.

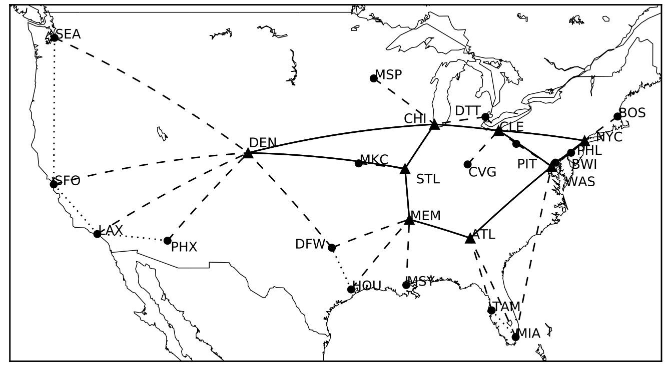
Fig 2(a) is the optimal topology without capacity constraints, under this circumstance, Saint Louis(STL) and Pittsburgh(PIT) receive much more traffic than other hub nodes. Meanwhile, backbone arcs between New York(NYC) and PIT serve the majority traveling demand in northeast America. Comparing Fig 2(a) with Fig 2(b), the role of constraints on hub capacity is clearly demonstrated as restricting the number of nodes connected to a hub. For instance, Backbone arcs between Denver(DEN) and Chicago(CHI) are established to alleviate the traffic burden of STL and decreases the number of nodes allocated to it. Cleveland(CLE) replaces the role of PIT as the hub nodes. While backbone arc capacity constraints tend to render a more fully-connected hub network and more tributary arcs to disaggregate flow on backbone arcs, as is shown in Fig 2(c), the great volume of traffic in east coast leads to more tributary arcs towards hubs like PIT and Baltimore(BWI). Direct arcs are also established for transportation in Boston(BOS) and non-hub node NYC. For the last graph Fig 2(d), the joint influence of these capacity constraints reduces the number of hubs in central America and increases that in eastern America, indicating the discrepancy in traveling demand between eastern and western regions. The STL turns out to be a non-hub node in this case. So traffic from west coast to east coast actually has more alternative paths to choose rather than choosing STL as the only intermediate hub. In this way, capacity constraints indirectly improve the robustness of the CHLP network and incorporate more practical factors in contrast to UHLP. Overall, the role of hub and backbone capacity constraints is non-negligible under the four scenarios.

B. Solution Methods Comparison

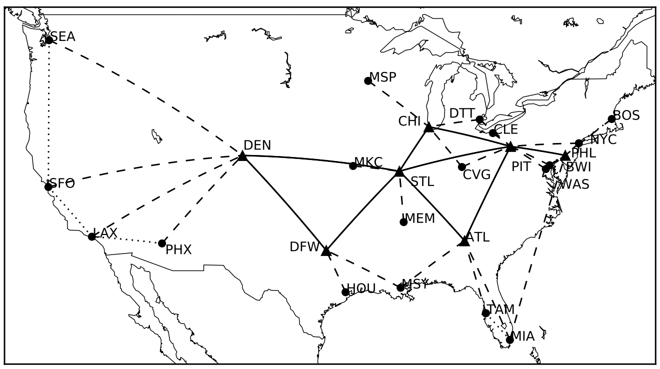
The second computational experiment shows the effectiveness of the BDSP in updating generating feasibility cuts, when DSP is unbounded in contrast to the traditional feasibility cuts from solvers (e.g. get_ray function in Cplex Python API). McDaniel [28] showed that valid initial Benders cuts can be obtained through warm start phase, which relaxes the integral restriction of MP. All tests are carried out in warm-start phase and integer phase. Algorithm1 is executed for both Benders decomposition with BDSP and Benders decomposition with traditional feasibility cuts except for the slight difference in *Ray*. The results for CAB dataset of *15, 20, 25* nodes are shown in Table II. The maximum number of steps for all the tests is set to be 5 and hub installation cost are set to be *0, 3248280, 3118240* for instances with *15, 20, 25* nodes respectively. The results in Table II includes objective function value(Obj), relative gap between UB and LB(Gap), CPU time in seconds(CPU times) and the number of iterations(#iter). The *inf* denotes a value of *4e7*, and it is the maximum possible flow for all the hubs and arcs, this is uncapacitated situations. From the results, it is clear that it takes more efforts for Cplex to deal with CHLP with large scale networks and more capacity constraints. This is especially obvious for instances with 25 nodes. This testbed also shows that BDSP generates higher quality and more stable feasibility cuts, when compared to the traditional methods which reduces the overall computing time and iterations. The disparity of CPU time for the two method is enlarged, in the worst case, it can take much longer time. This might be unacceptable in terms of computational efficiency. The disparity is further enlarged when both hub capacity constraints and backbone arc constraints are considered. As fixed costs and capacity constraints are considered in the CHLP, infeasible SP (i.e. unbounded DSP) are expected to occur more frequently than normal HLPs to avoid fixed costs and satisfy capacity limitations. This shows the pivotal role of BDSP in finding proper extreme rays in improving the overall performance of Benders decomposition, when addressing more practical CHLPs.



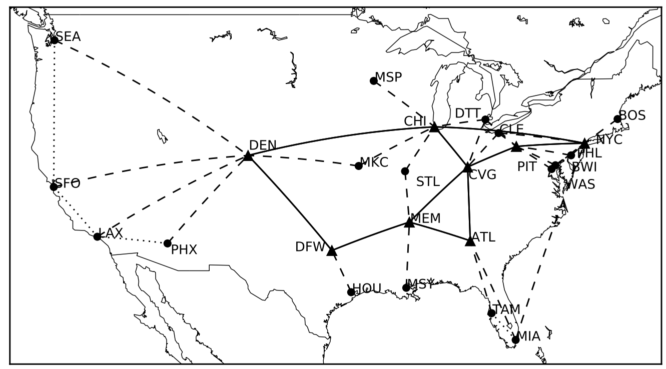
(a)Without capacity constraints
obj=451,792,379



(b)With hub capacity constraints
obj=451,992,919



(c)With backbone arc capacity constraints
obj=451,940,173



(d)With hub and backbone arc capacity constraints
obj=462,528,368

Fig. 2. The comparison of four CHLPs with and without capacity constraints in the CAB dataset

TABLE II
COMPARISON OF FEASIBILITY CUTS FROM BDSP AND TRADITIONAL METHOD

#Nodes	Fixed cost (*1e3)	Variable cost (*1e-2)	Capacity (*1e5)	Feasibility cuts from BDSP				Feasibility cuts from traditional method			
				Obj	Gap (%)	CPU time (s)	#iter	Obj	Gap (%)	CPU time (s)	#iter
15	[1,3,3,6]	[6,4,4,3]	[13,4.4]	145472763	0.0	171.9	16	145472763	0.0	2700.7	48
			[12,inf]	145472763	0.0	194.9	17	145472763	0.0	430.0	35
			[inf,4.4]	144790014	0.0	242.5	16	146361318	-1.085	2644.7	43
20	[2.5,2.5,2.5,5]	[5,4,4,3]	[28,12]	286314633	0.0	2189.6	22	286314633	0.0	3610.9	35
			[28,inf]	286314633	0.0	2377.2	23	286314633	0.0	3586.5	35
			[inf,11]	285591833	0.0	2154.3	23	285591833	0.0	3230.9	36
25	[2.5,2.5,2.5,2.5]	[4,4,4,4]	[32,9.2]	462528368	0.41	18443.7	28	462635377	0.39	31454.1	32
			[29,inf]	462421909	0.0	19763.1	22	488055839	-5.543	85502.3	29
			[inf,9.2]	451940173	0.0	6611.1	20	451940173	0.0	7701.0	25

V. CONCLUSION

In this study, we tackled a capacitated hub location problem with incomplete hub network. The optimal solution to this problem is correlated to fixed cost, the maximum length of a path and capacity constraints on hubs and backbone arcs. As these assumptions bring great difficulty for computation, an improved Benders decomposition algorithm was proposed. In order to explore the impact of capacity constraints of hubs and backbone arcs, the CAB dataset was used as a case study. In addition, the performance of the improved algorithm is compared with the classical one in the evaluation. Our work provided a more reasonable and robust network structure [29]. Computational challenge has been tackled to a certain extent. Further research could improve the performance for further model generalization. We also hope that our methodology can be adapted to other (re)location problems, e.g. [30]–[32], or other abstraction levels, e.g. to improve the air route network structure [33].

ACKNOWLEDGMENT

This study is supported by the Research Fund from National Natural Science Foundation of China (Grant Nos. 61650110516,. 61601013, 61521091, and 91538204)

REFERENCES

[1] M. E. O'Kelly, "The location of interacting hub facilities," *Transportation Science*, vol. 20, no. 2, pp. 92–106, 1986.

[2] X. Sun, S. Wandelt, M. Hansen, and A. Li, "Multiple airport regions based on inter-airport temporal distances," *Transportation Research Part E: Logistics and Transportation Review*, vol. 101, pp. 84–98, 2017.

[3] X. Sun, S. Wandelt, and X. Cao, "On node criticality in air transportation networks," *Networks and Spatial Economics*, pp. 1–25, 2017.

[4] E. Gourdin, M. Labb, and H. Yaman, "Telecommunication and location," *Facility Location*, pp. 275–305, 2001.

[5] S. R. Mohan and S. K. Neogy, "A genetic algorithm for the problem of configuring a hub-and-spoke network for a ltl trucking company in brazil. eur j oper res," *European Journal of Operational Research*, vol. 179, no. 3, pp. 747–758, 2007.

[6] J. F. Campbell and M. E. O'Kelly, *Twenty-Five Years of Hub Location Research*. INFORMS, 2012.

[7] M. E. O'Kelly and H. J. Miller, "The hub network design problem : A review and synthesis," *Journal of Transport Geography*, vol. 2, no. 1, pp. 31–40, 1994.

[8] I. Contreras, J. A. Daz, and E. FernŁndez, "Branch and price for large-scale capacitated hub location problems with single assignment," *Informs Journal on Computing*, vol. 23, no. 1, pp. 41–55, 2011.

[9] A. Hoff, J. Peiró, Á. Corberán, and R. Martí, "Adaptive memory programming for the capacitated modular hub location problem," *Technical Report*, 2016.

[10] H. Yaman and G. Carello, "Solving the hub location problem with modular link capacities," *Computers & Operations Research*, vol. 32, no. 12, pp. 3227–3245, 2005.

[11] I. Correia, S. Nickel, and F. Saldanha-da Gama, "A stochastic multi-period capacitated multiple allocation hub location problem: Formulation and inequalities," *Omega*, 2017.

[12] S. Alumur and B. Y. Kara, "Network hub location problems: The state of the art," *European Journal of Operational Research*, vol. 190, no. 1, pp. 1–21, 2008.

[13] I. Rodríguez-Martín and J. J. Salazar-González, "Solving a capacitated hub location problem," *European Journal of Operational Research*, vol. 184, no. 2, pp. 468–479, 2008.

[14] J. Kratica, M. Milanović, Z. Stanimirović, and D. Tošić, "An evolutionary-based approach for solving a capacitated hub location problem," *Applied Soft Computing*, vol. 11, no. 2, pp. 1858–1866, 2011.

[15] S. A. Alumur, B. Y. Kara, and O. E. Karasan, "The design of single allocation incomplete hub networks," *Transportation Research Part B: Methodological*, vol. 43, no. 10, pp. 936–951, 2009.

[16] I. Correia, S. Nickel, and F. Saldanha-da Gama, "Multi-product capacitated single-allocation hub location problems: formulations and inequalities," *Networks and Spatial Economics*, vol. 14, no. 1, pp. 1–25, 2014.

[17] M. E. O'Kelly, J. F. Campbell, R. S. D. Camargo, and G. D. M. Jr, "Multiple allocation hub location model with fixed arc costs," *Geographical Analysis*, vol. 47, no. 1, p. 7396, 2015.

[18] J. F. Benders, "Partitioning procedures for solving mixed-variable programming problems," *Numerische Mathematik*, vol. 4, no. 1, pp. 238–252, 1962.

[19] Y. Adulyasak, J.-F. Cordeau, and R. Jans, "Benders decomposition for production routing under demand uncertainty," *Operations Research*, vol. 63, no. 4, pp. 851–867, 2015.

[20] J.-F. Cordeau, G. Stojković, F. Soumis, and J. Desrosiers, "Benders decomposition for simultaneous aircraft routing and crew scheduling," *Transportation science*, vol. 35, no. 4, pp. 375–388, 2001.

[21] R. Rahmaniani, T. G. Crainic, M. Gendreau, and W. Rei, "The benders decomposition algorithm: A literature review," *European Journal of Operational Research*, 2016.

[22] T. L. Magnanti and R. T. Wong, "Accelerating benders decomposition: Algorithmic enhancement and model selection criteria," *Operations Research*, vol. 29, no. 3, pp. 464–484, 1981.

[23] A. Mercier, *A computational study of Benders decomposition for the integrated aircraft routing and crew scheduling problem*. Elsevier Science Ltd., 2005.

[24] N. Papadakos, "Practical enhancements to the magnanti-wong method," *Operations Research Letters*, vol. 36, no. 4, pp. 444–449, 2008.

[25] A. Mercier, "A theoretical comparison of feasibility cuts for the integrated aircraft-routing and crew-pairing problem," *Transportation Science*, vol. 42, no. 1, pp. 87–104, 2008.

[26] R. S. de Camargo, G. de Miranda, M. E. OKelly, and J. F. Campbell, "Formulations and decomposition methods for the incomplete hub location network design problem with and without hop-constraints," *Applied Mathematical Modelling*, vol. 51, pp. 274 – 301, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0307904X17304274

[27] E. M. de Sá, R. S. de Camargo, and G. de Miranda, "An improved benders decomposition algorithm for the tree of hubs location problem," *European Journal of Operational Research*, vol. 226, no. 2, pp. 185–202, 2013.

[28] D. Mcdaniel and M. Devine, "A modified benders' partitioning algorithm for mixed integer programming," *Management Science*, vol. 24, no. 3, pp. 312–319, 1977.

[29] S. Wandelt, X. Sun, and X. Cao, "Computationally efficient attack design for robustness analysis of air transportation networks," *Transportmetrica A: Transport Science*, vol. 11, no. 10, pp. 939–966, 2015. [Online]. Available: http://dx.doi.org/10.1080/23249935.2015.1089953

[30] A. K. Singh, N. Subramanian, K. S. Pawar, and R. Bai, "Cold chain configuration design: location-allocation decision-making using coordination, value deterioration, and big data approximation," *Annals of Operations Research*, pp. 1–25, 2016.

[31] D. Kress and E. Pesch, "Sequential competitive location on networks," *European Journal of Operational Research*, vol. 217, no. 3, pp. 483–499, 2012.

[32] M. Caserta, S. Schwarze, and S. Voß, "A mathematical formulation and complexity considerations for the blocks relocation problem," *European Journal of Operational Research*, vol. 219, no. 1, pp. 96–104, 2012.

[33] S. Wandelt and X. Sun, "Efficient compression of 4d-trajectory data in air traffic management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 844–853, April 2015.