



## An Estimated Formulation for the Capacitated Single Allocation $p$ -hub Median Problem with Fixed Costs of Opening Facilities

Rasul Esmailbeigi<sup>a</sup>, Bahman Naderi<sup>b</sup>, Alireza Arshadikhamseh<sup>b</sup>, Parvaneh Loni<sup>b</sup>

<sup>a</sup> Department of Industrial Engineering, University of Newcastle, Australia

<sup>b</sup> Department of Industrial Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran

### Abstract

In this paper, we consider the capacitated single allocation  $p$ -hub median problem generalized with fixed costs of opening facilities. The quadratic mathematical formulation of this problem is first adapted and then linearized. The typical approaches of linearization result in a high size complexity, i.e., having a large number of variables. To downsize the complexity, variables of the formulation are analyzed and some preprocessing approaches are defined. An estimated formulation is then developed to approximately solve large instances of the problem by commercial optimization solvers. The basic idea of this formulation is mapping the linearized formulation of the problem to a new formulation with fewer variables and a modified objective function. The efficacy of this formulation is shown by a computational study, where the estimated formulation is compared to a modified genetic algorithm from the literature. Results of computational experiments indicate that the estimated formulation is capable of generating good solutions within reasonable amount of time.

### Keywords:

Capacitated Single Allocation  $p$ -hub Median Problem; Mathematical Formulation; Linearization; Preprocessing.

### 1. Introduction

The hub location problem is first introduced by O'Kelly (1986). From that time until now, this problem is applied in several fields like transportation, geographical science, economics, operations research, network design, etc. In a hub location problem  $n$  locations (nodes) exist. If  $N = \{1, 2, \dots, n\}$  be the set of these points, for each  $i \in N$  and for each  $j \in N$  there is a flow  $w_{ij} \geq 0$  which must be transferred from  $i$  to  $j$ . When  $i \neq j$ ,  $w_{ij}$  is not necessarily equal to  $w_{ji}$ . Moreover,  $w_{ii}$  can be a positive value. These flows can be passengers, goods, information, etc. The collection and distribution of these flows can only be done in some special nodes called hub. We denote the set of all hub nodes by  $H$ . Sometimes the number of hubs is not predetermined, and it is found by the problem. However, the number of hubs,  $p = |H|$ , can be given a priori.

The underlying assumption is that the transference between two nodes  $i$  and  $j$  is not allowed when  $i \in N \setminus H$  and  $j \in N \setminus H$ . Generally, it is assumed that the incoming and outgoing traffic for a non-hub node can be routed through at most  $r$  hubs ( $1 \leq r \leq |H|$ ). When  $r = 1$ , the problem is referred to as the *single allocation*. In single allocation, each node can receive and send flow through exactly one hub. O'Kelly (1987) presented the single allocation  $p$ -hub median problem in which the number of hubs  $p = |H|$  is given a priori. When  $r = |H|$ , the problem is called *multiple allocation*.

\*Corresponding author email address: bahman.naderi@khu.ac.ir

in which routing through all hub nodes is allowed for each non-hub node. The single and multiple allocation hub location problems are classical versions of the problem. The general version of the problem is the  $r$ -allocation  $p$ -hub median problem which is first introduced by Yaman (2011).

The hub location problems can also be classified in terms of capacity for hubs. In the *capacitated* version of the problem, the amount of traffic passed through a hub is restricted. The capacity constraints can be considered for both input and output flow for the hub nodes. In typical hub location problems the objective is to minimize total cost of transportation. The total cost can be classified into three categories; collection cost (from non-hub nodes to hub nodes), transfer cost (between hub node), and distribution cost (from hub nodes to non-hub nodes). Let  $c_{ij}$  denote the transportation cost of a unit of flow from  $i$  to  $j$ . It is assumed that  $c_{ii} = 0$  for each  $i \in N$ . Moreover, the triangular inequality is not necessarily hold among parameters  $c_{ij}$ . The cost of transferring a unit of  $w_{ij}$  from node  $i$  to hub  $k$ , from hub  $k$  to hub  $l$ , and then from hub  $l$  to node  $j$  can be represented as follows:  $C_{iklj} = \alpha.c_{ik} + \beta.c_{kl} + \gamma.c_{lj}$

In the relation,  $\alpha$ ,  $\beta$ , and  $\gamma$  are nonnegative coefficients. The parameter  $\beta$  which is referred to as *discount factor* cannot exceed 1, i.e.  $0 \leq \beta \leq 1$ . It is also assumed that  $\beta \leq \gamma$ , and  $\beta \leq \alpha$ .

Sometimes in addition to total transferring cost, fixed costs are considered for establishing the hubs. O'Kelly (1992) introduced the hub location with fixed costs. He took into account the single allocation hub location problem in which the number of hubs were a decision variable in the problem. In the  $p$ -hub median problem, yet the fixed costs of opening facilities are ignored (see Alumur and Kara (2008)).

In this paper, we consider the capacitated single allocation  $p$ -hub median problem with fixed costs of opening facilities. This problem is a generalization of the capacitated single allocation  $p$ -hub median problem in which fixed costs of opening facilities are taken into account. In this problem, exactly  $p$  hub nodes must be selected from  $n$  nodes for opening facilities, and the objective is to minimize the sum of total transportation cost and total fixed costs. It is assumed that fixed costs of opening facilities may differ from node to node.

The literature includes several mathematical formulations for the single allocation hub location problems, but they cannot solve large instances of the problem. O'Kelly (1987) presented a binary quadratic mathematical formulation for the single allocation  $p$ -hub median problem. As is explained in Section 1.2, several linearizations for this quadratic formulation can be found in the literature. However, typical approaches of linearization result in a high size complexity, i.e., having a large number of variables. To downsize the complexity, we consider linearization of the quadratic formulation with a preprocessing approach. Then a mixed integer linear programming formulation is developed to approximately solve larger instances of the problem by commercially available solvers. The idea of this formulation is based on mapping the linearized formulation to a new formulation with fewer variables and a modified objective function. The introduction of estimated formulation is important for two main reasons: first, solving mathematical programming formulations with commercially available solvers is rather easier than using purpose-built algorithms. Second, due to continuous improvement in integer programming software systems, the ability of solving large scale instances are enhanced. As is pointed out by Bixby (2002), over the past decade, the problem-solving speed of mathematical programs has increased by a factor of more than 1,000,000, thanks to both the hardware and the software improvements. So, mathematical programming formulations will play a prominent role in optimally or approximately solving large scale instances. In this paper, we present an estimated formulation to approximately solve larger instances of the problem using commercial solvers. The estimated formulation presented in this paper has the following three desirable characteristics:

- It is easily solved within reasonable amount of time even for larger instances of the problem with 200 nodes. Also the solutions generated by this formulation are fairly good.
- By slight variations in this formulation, it can be used to solve a variety of the single allocation hub location problems, specially, those with various constraints.
- It considers capacity constraints as well as other constraints of the formulation simultaneously. However, during typical heuristic methods, this is usually done in complicated procedures. Specially, for instances with tight constraints that finding feasible solutions for the heuristic procedures is a hard task, the estimated formulation can find good solutions easily. Moreover, in the case that the problem is infeasible, it can easily detect infeasibility, but the original formulation may seeks to find a feasible solution within a lot of time.

To the best of our knowledge no heuristic procedure is developed for directly solving this optimization problem. However, the contribution of our research is developing the concept of estimated formulation to solve larger instances of a variety of single allocation hub location problems. The rest of the paper is organized as follows. In Section 1.2, we review relevant literature including mathematical programming formulations and heuristic procedures. Section 2 describes the new approach, where we develop a mixed integer linear programming formulation based on a

preprocessing. In Section 3, we present results of a computational study to show the efficacy of the presented formulation. Finally, we conclude the paper in Section 4.

## 1.2. Related literature

In this section, we review some heuristic procedures as well as mathematical programming formulations developed to solve single allocation hub location problems.

### 1.2.1. Heuristic procedures

Several heuristic and meta-heuristic approaches have been developed by researchers to solve single allocation hub location problems. O'Kelly (1987) was the first researcher who used heuristic approaches to solve single allocation  $p$ -hub median problem. Later Klincewicz (1991, 1992) presented several heuristic procedures to solve this problem. Skorin-Kapov and Skorin-Kapov (1994) developed a tabu search heuristic for the problem and compared it with heuristics of O'Kelly (1987) and Klincewicz (1992). Ernst and Krishnamoorthy (1996) presented a simulated annealing heuristic to solve this problem. Later, Ernst and Krishnamoorthy (1999) presented two heuristic procedures for the capacitated single allocation hub location problem. Kratica et al. (2007) developed two genetic algorithms (GAHUB1 and GAHUB2) for the uncapacitated single allocation  $p$ -hub median problem. They concluded that GAHUB2 performs better than GAHUB1 since it reached or improved all best known solutions of the benchmark data set. Chen (2007) developed a hybrid heuristic based on simulated annealing and tabu search for the uncapacitated single allocation hub location problem. Silva and Cunha (2009) presented a number of heuristics for this problem and showed that their approach is six times faster than the hybrid heuristic of Chen (2007). Ilić et al. (2010) developed a general variable neighborhood search heuristic for the uncapacitated single allocation  $p$ -hub median problem. They showed that their method outperform the best-known heuristics in terms of solution quality and computational effort. The interested readers are referred to surveys by Alumur and Kara (2008), Campbell and O'Kelly (2012), and Farahani et al. (2013) for details. In the following, we review mathematical programming formulations of the problem.

### 1.2.2. Mathematical programming formulations

Over the past two decades, modeling different mathematical formulations for the hub location problem has attracted considerable attention. The first mathematical formulation for the hub location problem is presented by O'Kelly (1987). He formulated the single allocation  $p$ -hub median problem as a binary quadratic programming formulation. In this formulation, he used the following binary variables:

$$z_{ik} = \begin{cases} 1 & \text{If } k \text{ is a hub node, and node } i \text{ is assigned to it} \\ 0 & \text{Otherwise} \end{cases}$$

In the case that  $z_{kk} = 1$  for some  $k \in N$ , node  $k$  is selected as a hub for establishing a facility. The formulation is expressed as follows:

$$\min \sum_{i \in N} \sum_{j \in N} w_{ij} \{ \alpha \sum_{k \in N} c_{ik} \cdot z_{ik} + \beta \sum_{k \in N} \sum_{l \in N} c_{kl} \cdot z_{ik} \cdot z_{jl} + \gamma \sum_{l \in N} c_{lj} \cdot z_{jl} \} \quad (1)$$

$$\text{s.t. } \sum_{i \in N} z_{ik} \leq (n - p + 1) \cdot z_{kk} \quad \forall k \in N, \quad (2)$$

$$\sum_{k \in N} z_{ik} = 1 \quad \forall i \in N, \quad (3)$$

$$\sum_{k \in N} z_{kk} = p, \quad (4)$$

$$z_{ik} \in \{0,1\} \quad \forall i, k \in N. \quad (5)$$

In this formulation, objective function (1) minimizes the total transportation cost. Constraints (2) ensure that a non-hub node is only assigned to a hub node. According to constraints (3), each non-hub node is assigned to exactly one hub node. Constraint (4) ensures that exactly  $p$  hub nodes are selected. Constraints (5) specify domains of decision variables.

O'Kelly (1992) also presented a binary quadratic programming formulation for the single allocation hub location problem with fixed costs. Using  $z_{ik}$  variables, his formulation (with  $\alpha = \gamma = 1$ ) can be expressed as follows:

$$\min \sum_{i \in N} \sum_{k \in N} c_{ik} \cdot (\alpha \cdot O_i + \gamma \cdot D_i) \cdot z_{ik} + \sum_{i \in N} \sum_{k \in N} z_{ik} \sum_{l \in N} \sum_{j \in N} (\beta \cdot w_{ij} \cdot c_{kl}) \cdot z_{jl} + \sum_{k \in N} F_k \cdot z_{kk} \quad (6)$$

s.t. (3), (5), and:

$$z_{ik} \leq z_{kk} \quad \forall i \in N, k \in N \setminus \{i\}. \quad (7)$$

In this formulation,  $O_i = \sum_{j \in N} w_{ij}$  represents the total amount of flow originating from node  $i$ , and  $D_j = \sum_{i \in N} w_{ij}$  represents the total amount of flow destined to node  $i$ . In objective function (6), part one considers the collection and distribution costs, part two considers the transferring cost between hubs, and part three considers fixed costs of establishing facilities in hub nodes. Constraints (7) ensure that a non-hub node is only assigned to a hub node. It can be seen from the formulation that the number of hubs is determined by the formulation itself.

Campbell (1994) presented the first linear programming formulation for the single allocation  $p$ -hub median problem. A better formulation for this problem is developed by Ernst and Krishnamoorthy (1996). Later, Ebery (2001) presented a mixed integer linear programming formulation for the problem with fewer variables and constraints. However, this formulation gives poor performance in comparison to the formulation presented by Ernst and Krishnamoorthy (1996). Ernst and Krishnamoorthy (1999) presented an integer linear programming formulation for the capacitated single allocation hub location problem. In this formulation binary  $z_{ik}$  variables are used as well as a set of continuous variables. Let  $x_{kl}^i$  be the total amount of flow originated from node  $i$ , and routed between hubs  $k$  and  $l$ . The formulation is then expressed as follows.

$$\min \sum_{i \in N} \sum_{k \in N} c_{ik} \cdot (\alpha \cdot O_i + \gamma \cdot D_i) \cdot z_{ik} + \sum_{i \in N} \sum_{k \in N} \sum_{l \in N} \beta \cdot c_{kl} \cdot x_{kl}^i + \sum_{k \in N} F_k \cdot z_{kk} \quad (8)$$

s.t. (3), (5), (7), and:

$$\sum_{l \in N} x_{kl}^i - \sum_{l \in N} x_{lk}^i = O_i \cdot z_{ik} - \sum_{j \in N} w_{ij} \cdot z_{jk} \quad \forall i, k \in N, \quad (9)$$

$$\sum_{i \in N} O_i \cdot z_{ik} \leq \Gamma_k \cdot z_{kk} \quad \forall k \in N, \quad (10)$$

$$x_{kl}^i \geq 0 \quad \forall i, k, l \in N. \quad (11)$$

In this formulation, objective function (8) minimizes the sum of total transportation cost and total fixed costs. Constraints (9) are flow balance equations at hub node  $k$  for the flow originating from node  $i$ . Constraints (10) make sure that the total input of hub  $k$  does not exceed its capacity  $\Gamma_k$ . Constraints (11) specify domains of the variables.

Correia et al. (2010) pointed out that in the absence of triangular inequality among parameters  $c_{ij}$ , this formulation may result in infeasible solutions. They completed the formulation by the following  $n^2$  inequalities:

$$\sum_{\substack{l \in N \\ l \neq k}} x_{kl}^i \leq O_i \cdot z_{ik} \quad \forall i, k \in N. \tag{12}$$

These inequalities ensure that variable  $x_{kl}^i$  take 0 if  $z_{ik}$  is 0. They also showed that adding these inequalities can reduce the average solution time of the formulation.

More recently, Merakli and Yaman (2017) study a capacitated hub location problem under hose demand uncertainty. Puerto et al. (2016) consider the ordered median hub location problems with capacity constraints. Rostami et al. (2016) develop a lower bounding procedures for the single allocation hub location problem. Stanojević et al. (2015) propose a hybridization of an evolutionary algorithm and a parallel branch and bound for solving the capacitated single allocation hub location problem. Rabbani et al. (2017) propose a metaheuristic algorithms for solving a hub location problem: application in passive optical network planning.

## 2. Estimated formulation

In Section 1, we describe the problem and review its related literature. In this Section, a linearization of the quadratic formulation of the problem is considered together with a preprocessing approach. Then the estimated formulation is developed based on this preprocessing.

The capacitated single allocation  $p$ -hub median problem with fixed costs of opening facilities can be easily formulated by merging the presented formulations in Section 1. The quadratic formulation of this problem is expressed as minimization of objective function (6) subject to constraints (3)-(5) and (10). Also, to solve small instances of the problem to optimality, the linear formulation of the problem can be expressed as minimization of objective function (8) subject to constraints (3)-(5) and (9)-(12). Note that in the existence of capacity constraints (10), inequalities (2) and (7) can be dropped. Besides, for an asymmetric distance matrix, the first part of objective functions (6) and (8) which considers the collection and distribution costs must be wrote as  $\sum_{i \in N} \sum_{k \in N} (\alpha \cdot O_i \cdot c_{ik} + \gamma \cdot D_i \cdot c_{ki}) \cdot z_{ik}$ .

To linearize the quadratic version of the problem, we define binary variable  $y_{iklj}$  as follows:

$$y_{iklj} = \begin{cases} 1 & \text{If } z_{ik} = z_{jl} = 1 \\ 0 & \text{Otherwise} \end{cases}$$

These variables are used to linearize the quadratic part of objective function (6), i.e.  $y_{iklj} = z_{ik} \cdot z_{jl}$ . Due to this definition a couple of preprocessing can be performed to eliminate some of these variables.

**Proposition 1.** *There is no need for variables  $y_{ikkj}$  to be considered in the linearized formulation of the problem.*

**Proof.** Note that due to the assumptions  $c_{kk} = 0$  for each  $k \in N$ . So, variables  $y_{ikkj}$  should not be considered.

**Proposition 2.** *In the linearized formulation of the problem, all variables  $y_{ikli}$  can be eliminated for  $k \neq l$ .*

**Proof.** Due to constraints (3) each non-hub node is exactly assigned to one hub node. So, all variables  $y_{ikli}$  will be equal to zero when  $k \neq l$ .

**Proposition 3.** *In the linearized formulation of the problem, all variables  $y_{ikli}$  can be eliminated.*

**Proof.** According to Proposition 2, among  $y_{ikli}$  variables just variables of form  $y_{ikk}$  can take one. So, due to Proposition 1 variables  $y_{ikli}$  should not be generated.

**Proposition 4.** *In the linearized formulation of the problem, all variables  $y_{ikij}$  and  $y_{ijlj}$  can be eliminated.*

**Proof.** It is obvious that a node either is a hub or a non-hub node. So, among variables  $y_{ikij}$  just variables of form  $y_{iij}$  can be positive. However, due to Proposition 1 these variables can be eliminated. A similar proof for variables  $y_{ijlj}$  indicates that these variables can also be eliminated.

Now by using Propositions (1)-(4) a linearized formulation for the problem can be stated as follows. All indices of variables in this formulation are elements of  $N$ .

$$\min \sum_i \sum_{k \neq i} (\alpha.O_i.c_{ik} + \gamma.D_i.c_{ki}).z_{ik} + \beta.\sum_i \sum_{j \neq i} w_{ij} \sum_{k \neq j} \sum_{\substack{l \neq k \\ l \neq i}} c_{kl} \cdot y_{iklj} + \sum_k F_k \cdot z_{kk} \quad (13)$$

s.t. (3)-(5), (10), and:

$$y_{iklj} \geq z_{ik} + z_{jl} - 1 \quad \forall i \neq j, k \neq l, i \neq l, j \neq k, \quad (14)$$

$$y_{iklj} = y_{jtki} \quad \forall i \neq j, k \neq l, i \neq l, j \neq k, \quad (15)$$

$$0 \leq y_{iklj} \leq 1 \quad \forall i \neq j, k \neq l, i \neq l, j \neq k. \quad (16)$$

In the linearized formulation, constraints (14) assure that variable  $y_{iklj}$  takes one if  $z_{ik} = z_{jl} = 1$ . Valid inequalities (15) are included so as to strengthen the formulation. Constraints (16) specify domains for  $y_{iklj}$  variables. Note that in this formulation we considered these variables as continuous variables since in the optimal solution for the problem they will be binary.

The number of variables and constraints in the presented formulation is bounded by  $O(n^4)$ . Obviously, this formulation cannot be used for solving larger instances of the problem to optimality. However, we apply it to develop the estimated formulation.

As is shown by Propositions (1)-(4), some form of variables  $y_{iklj}$  can be eliminated. Now, we define some parameters based on Propositions (1)-(4) and circumstances of the nodes as follows:

$$i \text{ hub \& } j \text{ nun-hub } (y_{iij}): \quad s_{ij} = f(c_{il}) \quad \forall i \neq j \\ l \notin \{i, j\}$$

$$i \text{ nun-hub \& } j \text{ hub } (y_{ikj}): \quad t_{ij} = f(c_{kj}) \quad \forall i \neq j \\ k \notin \{i, j\}$$

$$i \text{ nun-hub \& } j \text{ nun-hub } (y_{iklj}): \quad u_{ij} = f(c_{kl}) \quad \forall i \neq j \\ l \neq k \\ l \notin \{i, j\} \\ k \notin \{i, j\}$$

In these definitions  $f$  is a function, which its inputs are some elements of the distance matrix. Although, various functions can be applied, for the sake of simplicity we consider  $f$  as the *mean* or average of the inputs. Now in the objective function (13) we estimate transferring costs between hubs as follows:

$$B = \beta.\sum_i \sum_{j \neq i} w_{ij} \sum_{k \neq j} \sum_{\substack{l \neq k \\ l \neq i}} c_{kl} \cdot y_{iklj} \approx$$

$$\hat{B} = \beta.\sum_i \sum_{j \neq i} w_{ij} \cdot \{c_{ij} \cdot z_{ii} \cdot z_{jj} + s_{ij} \cdot z_{ii} \cdot (1 - z_{jj}) + t_{ij} \cdot (1 - z_{ii}) \cdot z_{jj} + u_{ij} \cdot (1 - z_{ii}) \cdot (1 - z_{jj})\}$$

The expression can be simply linearized by considering  $h_{ij} = z_{ii} \cdot z_{jj}$ . After linearization, the following mixed integer linear programming formulation is achieved. We refer to this formulation as *estimated formulation*.

$$\min \sum_i \sum_{k \neq i} (\alpha.O_i.c_{ik} + \gamma.D_i.c_{ki}).z_{ik} + \beta.\sum_i \sum_{j \neq i} w_{ij} \cdot \{c_{ij} \cdot h_{ij} + \\ s_{ij} \cdot (z_{ii} - h_{ij}) + t_{ij} \cdot (z_{jj} - h_{ij}) + u_{ij} \cdot (h_{ij} - z_{ii} - z_{jj} + 1)\} + \sum_k F_k \cdot z_{kk} \quad (17)$$

s.t. (3)-(5), (10), and:

$$h_{ij} \leq z_{ii} \quad \forall i < j, \quad (18)$$

$$h_{ij} \leq z_{jj} \quad \forall i < j, \quad (19)$$

$$h_{ij} \geq z_{ii} + z_{jj} - 1 \quad \forall i < j, \quad (20)$$



$$h_{ij} = h_{ji} \quad \forall i < j, \tag{21}$$

$$z_{ik} + h_{ik} \leq z_{kk} \quad \forall i \neq k, \tag{22}$$

$$h_{ij} \geq 0 \quad \forall i \neq j, \tag{23}$$

Objective function (17) minimizes the estimated value of total cost. Constraints (18)-(21) are added for linearization. Constraints (22) are a set of valid inequalities for the formulation. According to these constraints at most one of the variables  $z_{ik}$  and  $h_{ik}$  can take one when  $i \neq k$ . Without these constraints the formulation remains valid, but they are added so as to strengthen the formulation. Constraints (23) specify domains of constraints. Evidently, according to the formulation, variables  $h_{ij}$  take binary values in the optimal solution.

The number of variables and constraints in the estimated formulation is bounded by  $O(n^2)$ . Note that just the solution provided by the estimated formulation is used and the objective value of estimated formulation cannot be considered as a valid index of quality of the solution. The performance of this formulation is examined in Section 3.

### 3. Computational results

In this section, we evaluate performance of the estimated formulation. All the experiments were run on a computer with a two-core 2.66 GHz Intel processor and 4.0 GB RAM. We used MATLAB (R2009a) to code algorithms and calculate input parameters of the estimated formulation. To solve the formulation, we used solver CPLEX of GAMS-22.

Our computational experiments have carried out using the AP data set. This well-known data set consists of 200 nodes and includes fixed costs and capacities for the nodes. Two types of fixed costs, loose and tight are considered. In the data set with tight fixed costs, nodes with larger flows have higher fixed costs. So in general it is harder to decide which nodes should be hubs in such cases. Similarly two types of capacities, loose and tight are considered. In the AP data set it is assumed that  $\alpha = 3$ ,  $\beta = 0.75$ , and  $\gamma = 2$ .

We considered GAHUB2 and modified it to solve the problem. GAHUB2 has been developed by Kratica et al. (2007) to solve the uncapacitated single allocation  $p$ -hub median problem. However, it can be easily used to solve this problem by including fixed costs and penalty of violating constraints in the objective function (so called constraint relaxation). We consider following approach to make GAHUB2 suitable for solving this problem. Let  $J_k$  be the set of non-hub nodes assigned to hub  $k$  in a solution of GAHUB2. Then total input flow for hub  $k$  is  $I_k = \sum_{l \in J_k} O_l + O_k$ . Now for each hub  $k$

calculate  $d_k = I_k - \Gamma_k$ , and constitute set  $D^+ = \{k \mid d_k > 0\}$ . Then the necessary condition for a feasible solution is that  $|D^+| = 0$ . If the solution is feasible update the objective value as follows. In this relation  $obj$  represents the total transportation cost (the objective value for the original GAHUB2).

$$new\_obj = (obj + \sum_{k \in H} F_k)$$

Otherwise, calculate  $\pi = \sum_{k \in D^+} \frac{d_k}{\Gamma_k}$  and update the objective value as follows.

$$new\_obj = (1 + \lambda \cdot \pi) \times (obj + \sum_{k \in H} F_k)$$

In this relation  $\lambda$  is the weight for scaling the penalty of infeasibility. This parameter can be either a constant value or a variable. According to our computational tests, adjusting this weight dynamically, i.e., using  $\lambda = \frac{|D^+|}{p}$ , make GAHUB2 perform well. So, we used this parameter in our computational experiments.

To compare our approach with the modified GAHUB2 (MGAHUB2), five classes of instances for smaller instances of the problem and nine classes for larger instances are considered. Each Class is denoted by  $Cn$  in which  $n$  represents the number of nodes for the instances of the class. Smaller instances of the problem include classes  $C10$ ,  $C20$ ,  $C25$ ,

C40, and C50. These classes are solved to optimality by mathematical programming formulations of the problem. Larger instances of the problem include classes C60, C70, C75, C90, C100, C125, C150, C175, and C200. We consider for each class, three different levels of number of hubs, L1, L2, and L3 that are shown in Table 1.

**Table 1.** Different levels of number of hubs for each class

Class	C10	C20	C25	C40	C50	C60	C70	C75	C90	C100	C125	C150	C175	C200
L1	2	3	4	5	7	8	9	10	12	13	17	19	22	25
L2	3	5	7	10	13	15	18	19	23	25	32	39	44	50
L3	5	10	13	20	25	30	35	38	45	50	63	75	88	100

Also, for each level of number of hubs, four problem instances exist in terms of levels of capacity and fixed costs (tight or loose). We consider following approach for comparing two methods: we first computed input parameters of the estimated formulation and then solved it by GAMS. Then we recorded the actual objective value of its solution as well as the total computational time for computing input parameters and solving the formulation by GAMS. Finally, we solved the MGAHUB2 in the same time limit as our method takes. Note that in the cases that MGAHUB2 could not find a feasible solution in the specified time limit, we give it more time to find a feasible solution. We solved MGAHUB2 ten times, and for each instance, the average gap of ten objective values is calculated. We also solved smaller instances of the problem to optimality, and used optimal values to calculate gap for the heuristics. So, for smaller instances, the gap is calculated as follows:

$$\text{gap} = \frac{\text{Objective value} - \text{Optimal value}}{\text{Optimal value}} \times 100\%$$

Since for the larger instances of the problem the optimal solution cannot be achieved by solving mathematical formulations, we computed relative gaps for heuristics as follows:

$$\text{gap} = \frac{\text{Objective value} - \min(\text{Objective values})}{\min(\text{Objective values})} \times 100\%$$

Table 2 summarizes results of different classes of instances. The average amount of gap and average amount of computational times for 12 instances in each class are reported in the table for the estimated formulation (MIP) and MGAHUB2 (GA).

**Table 2.** Comparing the new approach with the GAHUB2

Class	Small instances						Large instances							
	C10	C20	C25	C40	C50	C60	C70	C75	C90	C100	C125	C150	C175	C200
GA (%)	2.64	6.50	3.35	6.74	6.32	3.75	4.20	4.27	3.58	8.18	7.37	2.44	10.81	4.92
MIP (%)	1.11	0.22	0.66	1.51	2.08	0.57	0.73	0.05	0.48	0.00	0.16	0.31	0.05	0.78
CPU (s)	2.85	2.92	3.32	3.58	4.04	4.51	5.65	6.23	8.21	9.45	14.78	27.28	45.72	75.60

From Table 2 we can make the following two observations:

- (1) In all 14 classes of instances, the average gap of MGAHUB2 is significantly larger than that of the new approach. The difference between two methods is more apparent in larger classes, where we computed relative gaps for them. Evidently, results indicate that solutions to the estimated formulation are fairly good.
- (2) The computational times in the table say that the estimated formulation is capable of solving all instances within reasonable amount of time. Moreover, the solution time of estimated formulation for different classes of instances has not much difference. It is noteworthy to say that such computational time for larger classes is very good as a mixed integer linear programming formulation with the number of variables and constraints bounded by  $O(n^2)$ .

Now we examine methods in terms of levels of number of hubs. Numerical results for this comparison are provided in Table 3. For each level of number of hubs, the average values associated with each level are reported. From the table we can make following two observations:

- (1) For larger instances, it can be seen that there is a trend in gaps when the number of hubs is changed. In MGAHUB2, the average amount of gaps becomes larger when the number of hubs is decreased but for the new approach, it is the other way around. So, for fewer hubs, the estimated formulation gives better solutions than the MGAHUB2. For small instances of the problem, there is no trend, but the new approach shows better performance anyway.



- (2) For larger instances, the computational times are decreased for a greater number of hubs. Nevertheless, the changes are small enough to say that the solution time for the estimated formulation is roughly independent of number of hubs. Results of small instances also indicate the assertion.

**Table 3.** Results in terms of levels of number of hubs.

Level of $p$	Small instances			Large instances		
	GA (%)	MIP (%)	CPU (s)	GA (%)	MIP (%)	CPU (s)
$L1$	2.79	0.90	3.34	9.84	0.01	23.19
$L2$	6.70	1.16	3.36	4.00	0.08	21.50
$L3$	5.84	0.92	3.32	2.66	0.95	21.11

We now examine results in terms of condition of fixed costs and capacities. Consider Table 4 in which digits 0 and 1 represent the condition of fixed costs and capacities respectively. A digit 0 represents that the capacity or fixed cost is tight and a digit 1 represents that it is loose. Each combination zero-one includes 15 and 27 instances for small and large instances respectively. From this table we can make the following observations and insights:

- (1) Consider combination 1,0 in which fixed costs are loose but capacities are tight. In this condition, for small instances, both methods show worst performance in comparison to other conditions. However, for large instances, it can be seen that the average gap for the estimated formulation in this condition is significantly smaller than that of other conditions. It means that in this condition (fixed costs loose but capacities tight) the performance of estimated formulation is far better than that of MGAHUB2. Note that gaps in smaller instances represent the difference between the solution and the optimal solution, but for larger instances they represent the difference between two methods.
- (2) Although, in all four conditions, the estimated formulation gives better solutions than MGAHUB2, the difference is more apparent when capacities are tight. This is true for both small and large instances of the problem. Specially, for large instances, results indicate that when capacity constraints are tight, solving the estimated formulation can be very effective. In the cases that the capacity is tight, the estimated formulation results in better solutions than GAHUB2.
- (3) As is expected, when fixed costs and capacities are loose (combination 1,1), both MGAHUB2 and the estimated formulation shows better performance in comparison to other conditions. This can be understood from small instances where objective values are compared with the optimal solutions. For both small and large instances, when fixed costs become tight, the gaps for both methods are increased.
- (4) The solution time of the estimated formulation for tight fixed costs (constraints) is more than loose fixed costs (constraints). The difference is yet significantly small, so it can be ignored.

**Table 4.** Results in terms of condition of fixed costs and capacities.

Condition	Small instances			Large instances		
	GA (%)	MIP (%)	CPU (s)	GA (%)	MIP (%)	CPU (s)
$0,0$	5.51	1.07	3.27	8.72	0.13	23.43
$0,1$	4.54	0.97	3.35	3.42	0.68	21.51
$1,0$	6.78	1.09	3.35	7.12	0.02	21.96
$1,1$	3.61	0.84	3.40	2.75	0.56	20.85

#### 4. Conclusion

In this study, we generalized the capacitated single allocation  $p$ -hub median problem by considering fixed costs of opening facilities. We first considered a linearization of the quadratic formulation of this problem with a preprocessing approach. Then a mixed integer linear programming formulation, the estimated formulation, is presented to approximately solve larger instances of this problem by commercially available solvers. It was shown that the solution time to the estimated formulation is fairly small even for larger instances of the problem with 200 nodes. Moreover, for tight constraints, it was shown that the use of the estimated formulation can be very effective in solving large scale instances. This formulation can also be used (by slight variations) to solve a variety of the single allocation hub location problems, specially, those with various constraints.

#### References

- Alumur, S. and B. Y. Kara (2008). Network hub location problems: The state of the art. *European Journal of Operational Research*, Vol. 190(1), pp. 1-21.
- Bixby, R. E. (2002). Solving Real-World Linear Programs: A Decade and More of Progress. *Operations Research*, Vol. 50(1), pp. 3-15.
- Campbell, J. F. (1994). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research*, Vol. 72(2), pp. 387-405.

- Campbell, J. F. and M. E. O'Kelly (2012). Twenty-Five Years of Hub Location Research. *Transportation Science*, Vol. 46(2), pp. 153-169.
- Chen, J.-F. (2007). A hybrid heuristic for the uncapacitated single allocation hub location problem. *Omega*, Vol. 35(2), pp. 211-220.
- Correia, I., S. Nickel, and F. Saldanha-da-Gam. (2010). The capacitated single-allocation hub location problem revisited: A note on a classical formulation. *European Journal of Operational Research*, Vol. 207(1), pp. 92-96.
- Ebery, J. (2001). Solving large single allocation p-hub problems with two or three hubs. *European Journal of Operational Research*, Vol. 128(2), pp. 447-458.
- Ernst, A. T. and M. Krishnamoorthy (1996). Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location Science*, Vol. 4(3), pp. 139-154.
- Ernst, A. T. and M. Krishnamoorthy (1999). Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research*, Vol. 86(0), pp. 141-159.
- Farahani, R. Z., M. Hekmatfar, A., Boloori, and A.E. Nikbakhsh (2013). Hub location problems: A review of models, classification, solution techniques, and applications. *Computers & Industrial Engineering*, Vol. 64(4), pp. 1096-1109.
- Ilić, A., D. Urošević, J. Brimberg, and N. Mladenović (2010). A general variable neighborhood search for solving the uncapacitated single allocation p-hub median problem. *European Journal of Operational Research*, Vol. 206(2), pp. 289-300.
- Klincewicz, J. (1992). Avoiding local optima in the p-hub location problem using tabu search and GRASP. *Annals of Operations Research*, Vol. 40(1), pp. 283-302.
- Klincewicz, J. G. (1991). Heuristics for the p-hub location problem. *European Journal of Operational Research*, Vol. 53(1), pp. 25-37.
- Kratka, J., Z. Stanimirović, D. Tošić, and V. Filipović, (2007). Two genetic algorithms for solving the uncapacitated single allocation p-hub median problem. *European Journal of Operational Research*, Vol. 182(1), pp. 15-28.
- Merakli, M. and H. Yaman (2017). A capacitated hub location problem under hose demand uncertainty. *Computers & Operations Research*, Vol. 88, pp. 58-70.
- O'Kelly, M. (1992). Hub facility location with fixed costs. *Papers in Regional Science*, Vol. 71(3), pp. 293-306.
- O'Kelly, M. E. (1986). The Location of Interacting Hub Facilities. *Transportation Science*, Vol. 20(2), pp. 92-106.
- O'Kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research*, Vol. 32(3), pp. 393-404.
- Puerto, J., A.B. Ramos, A.M. Rodríguez-Chía, and M.C. Sánchez-Gil (2016). Ordered median hub location problems with capacity constraints. *Transportation Research Part C: Emerging Technologies*, Vol. 70, pp. 142-156.
- Rabbani, M., M. Ravanbakhsh, H. Farrokhi-Asl, M. Taheri, (2017). Using metaheuristic algorithms for solving a hub location problem: application in passive optical network planning." *International Journal of Supply and Operations Management*, In-press.
- Rostami, B., C. Buchheim, J.F. Meier and U. Clausen (2016). Lower Bounding Procedures for the Single Allocation Hub Location Problem. *Electronic Notes in Discrete Mathematics*, Vol, 52, pp. 69-76.
- Silva, M. R. and C. B. Cunha (2009). New simple and efficient heuristics for the uncapacitated single allocation hub location problem. *Computers & Operations Research*, Vol. 36(12), pp. 3152-3165.
- Skorin-Kapov, D. and J. Skorin-Kapov (1994). On tabu search for the location of interacting hub facilities. *European Journal of Operational Research*. Vol. 73(3), pp. 502-509.
- Stanojević, P., M. Marić, Z. Stanimirović, (2015). A hybridization of an evolutionary algorithm and a parallel branch and bound for solving the capacitated single allocation hub location problem. *Applied Soft Computing*, Vol. 33, pp. 24-36.
- Yaman, H. (2011). Allocation strategies in hub networks. *European Journal of Operational Research*, Vol. 211(3), pp. 442-451.