World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:14, No:7, 2020

Constant Factor Approximation Algorithm for p-Median Network Design Problem with Multiple Cable Types

Chaghoub Soraya, Zhang Xiaoyan

Abstract—This research presents the first constant approximation algorithm to the p-median network design problem with multiple cable types. This problem was addressed with a single cable type and there is a bifactor approximation algorithm for the problem. To the best of our knowledge, the algorithm proposed in this paper is the first constant approximation algorithm for the p-median network design with multiple cable types. The addressed problem is a combination of two well studied problems which are p-median problem and network design problem. The introduced algorithm is a random sampling approximation algorithm of constant factor which is conceived by using some random sampling techniques form the literature. It is based on a redistribution Lemma from the literature and a steiner tree problem as a subproblem. This algorithm is simple, and it relies on the notions of random sampling and probability. The proposed approach gives an approximation solution with one constant ratio without violating any of the constraints, in contrast to the one proposed in the literature. This paper provides a $(21 + 2\epsilon)$-approximation algorithm for the p-median network design problem with multiple cable types using random sampling techniques.

Keywords—Approximation algorithms, buy-at-bulk, combinatorial optimization, network design, p-median.

I. Introduction

The p-median problem is one of the well studied problems in the literature. In a p-median instance we are given a graph $G = (V, E)$, a set of facilities $F \subseteq V$, a set of client $D \subseteq V$, and an assignment cost c_{ij} for assigning client j to facility i. The goal of this problem is to open at most p facilities and assign each client to its nearest open facility such that all the demands are satisfied and the total incurred cost is minimized. There exist several variants of the p-median problem such as the capacitated p-median [14], connected p-median [17], matroid median [6], knapsack median [6], p-median bilevel [5], planar p-median [20], and p-median problem with uniform penalties [8]. In this paper, we consider the p-problem which integrates with Buy-at-Bulk in network design.

The Buy-at-Bulk network design problem was introduced by Salman et al. [12]. They gave an $O(\log^2 n)$ approximation algorithm for the single sink buy-at-bulk network design problem in Euclidean graph. Awerbuch and Azar [4] also presented an $O(\log^2 n)$ for the same problem but in general metric space. Garg et al. [13] presented an $O(K)$-approximation algorithm based on LP-rounding. Guha, Meyerson and Munagala [18] presented the first constant-factor approximation algorithm, where K is the number of cables. Then, this factor was improved to 2000 by Talwar [11], who also presented an LP based rounding algorithm with a factor of 216. Gupta et al. [2] gave an 76.8 approximation algorithm which is simple and easy to analyse.

C. Soraya and Z. Xiaoyan are with School of Mathematical Science & Institute of Mathematics, Nanjing Normal University, Nanjing 210023, China (e-mail: chaghoubsoraya@yahoo.fr).

The capacitated-cable p-median in buy-at-bulk network design problem (pMNDP) which was proposed by Ravi and Sinha [16] has the two following variants:

Single cable case: We are given an undirected graph $G = (V, E)$ with non negative costs c_e on edges, and a set $D \subseteq V$ of demands with d_j representing the nonnegative weight of $j \in D$. We are also given a single type of cable with specified capacity and a cost per unit length. The capacitated-cable p-median in Buy-at-Bulk network design problem with one cable type was first studied by Ravi and Sinha [16], who gave a bifactor approximation algorithm, of factor $(\rho_{pM} + 2, 2)$, where ρ_{pM} is the factor of the approximation algorithm for the p-median problem. In general, a bifactor approximation algorithm may violate the p-median constraint.

Multiple cable case: In an instance of capacitated cable p-median problem we are given an undirected graph $G = (V, E)$ with non negative costs c_e on edges, a set $D \subseteq V$ of demands where each demand j has a nonnegative weight d_j, and a set of K cable types where each type has a specified capacity and a cost per unit length. These cables satisfy the so-called "economy of scale"; namely, there exists a break-point, beyond which it becomes more economic to use the next cable type than copy the current one. We assume that the flows among the nodes are unsplittable.

A thorough research of the relevant literature yielded no article about multiple cable types. Hence, it is assumed that this problem is only dealt with in this research paper. To the best of our knowledge, the problem with multiple cable types was not researched before and thus is it is dealt with for the first time in this paper. We present a random sampling approximation algorithm of constant factor by using some techniques in [2].

The p-median in network design problem is NP-hard, since it generalizes the well known NP-hard problem of p-median. The problem is also a generalization of the Single-Sink Buy-at-Bulk network design problem. which is a well studied problem in the literature.

The p-median in network design problem has various applications in transportation, computer network design and hierarchical design of telecommunication network [9], [16].

For the pMNDP with K different cable types, we present a $(21 + 2\epsilon)$-approximation algorithm by extending some techniques used by [2] for the Single-Sink in Buy-at-Bulk network design. The optimal solution to the pMNDP with multiple type consists of:

1) A set of p open facilities,
2) a set of cables with sufficient capacity installed on the edges to support the flow.

Our approach gives an approximation solution with constant unifactor ratio for multiple cable types that does not violate any of the constraints. We first present lower bounds of the total expected cost of installing cables incurred by p-median, p-Steiner tree, and redistributing demands. Then, we prove the performance guarantee of the given approximation algorithm.

II. p-Median in Buy-at-Bulk Network Design Problem

The p-median in network design problem (pMNDP) with multiple type is defined on a graph $G = (V, E)$, where V denotes the set of nodes with $|V| = n$, and E denotes the set of edges. There are clients at a subset $C \subseteq V$ of nodes in the graph, and each client has a unit demand. There are also facilities at a subset $F \subseteq V$ and a set of different cable types. Each cable k has a limited capacity u_k and a set-up cost γ_k per unit length. There is also a weight function $c_e \in \mathbb{Z}_{\geq 0}$ for each edge $e \in E$. Cables are indexed such that $u_1 \leq u_2 \leq ... \leq u_K, \gamma_1 \leq \gamma_2 \leq ... \leq \gamma_K$, and $\gamma_1/u_1 \geq \gamma_2/u_2 \geq ... \geq \gamma_K/u_K$. Let the flow dependent cost of cable k be $\varrho_k = \frac{\gamma_k}{u_k}$, where $1 \leq k \leq K$.

A solution to the pMNDP with multiple type consists of opening at most p facilities, assigning each client to an open facility and installing a combination of cables on the edges of the network.

For each edge, we create a pair of anti-parallel directed arcs, with the same characteristics as the original one. We denote by E the set of edges e which is the undirected version of the set of arcs $a \in A$, where A is the set of arcs. The edge between nodes l and m is $e = (l, m)$, and the arc between the same nodes is $a = (l, m)$ or $a = (m, l)$. Let $\delta^+(u) = (u, v) \in A$ and $\delta^-(v) = (u, v) \in A$. Let x_e^k indicate whether or not cable of type k is installed on edges e. Let y_i indicate whether or not facility i is opened. Let f_{uv}^j indicate if there is a flow from client j on edge (u, v) or not. We have the following integer linear programming (IP) formulation for the pMNDP with multiple cable type problem.

A. Integer Programming Formulation

$$\min \sum_{e \in E} \sum_{k \in K} \gamma^k l_e^k x_e^k$$

$$\sum_{i \in F} y_i \leq p, \tag{1}$$

$$\sum_{\bar{e} \in \delta^+(j)} f_{\bar{e}}^j \geq 1, \quad j \in D, \tag{2}$$

$$\sum_{\bar{e} \in \delta^+(j)} f_{\bar{e}}^j = \sum_{\bar{e} \in \delta^-(j)} f_{\bar{e}}^j, \quad \forall j \in D, v \in V \setminus F, v \neq j, \tag{3}$$

$$\sum_{\bar{e} \in \delta^-(i)} f_{\bar{e}}^j - \sum_{\bar{e} \in \delta^+(i)} f_{\bar{e}}^j \leq y_i, \qquad \forall j \in D, \forall i \in F, \tag{4}$$

$$\sum_{j \in D} d_j (f_{(l,m)}^j + f_{(m,l)}^j) \leq \sum_{k=1}^{K} u_k x_{(m,l)}^k, \ x_e^k \in \mathbb{Z}^+ \cup \{0\}. \tag{5}$$

The first constraint imposes that the number of open facilities does not exceed p. The second constraint ensures that at least one unit of flow leaves each client. The third constraint imposes the conservation of flow at non-facility nodes. The forth constraint states that the flow can only terminate at an opened facility. The last constraint imposes that the capacity of the cables installed on each edge is sufficient to support the flow on this edge.

III. Approximation Algorithm for p-Median in Buy-at-Bulk Network Design with Multiple Cable Types

Let us assume, w.l.o.g, that the number of demands is power of 2, also all values u_k, γ_k and d_j are powers of 2. The last assumption can be accounted by losing a factor 2 in the approximation ratio. This idea of rounding is generalized from [2].

The following lemma is the Redistribution Lemma, that allows us to divide the nodes in a tree into clusters.

Lemma 1: [15](Redistribution lemma) Let T be a tree rooted at r with each edge having capacity U. For each vertex $j \in T$, let $w(j) < U$ be the weight located at j with $\sum_j w(j)$ which is a multiple of U. Then there is an efficiently computable (random) flow on the tree that redistributes weights without violating edge capacities, so that each vertex receives a new weight $w^{'}(j)$ that is either 0 or U. Moreover, $Pr[j$ has $w^{'}(j) >= U] = w(j)/U$.

Proof: Let us replace each edge in T by two oppositely directed arcs. we first show that the Lemma holds in this bidirected tree. First, we take an Euler tour of the vertices, yielding a cycle C. We also pick a value Y drawn uniformly at random from $(0, U]$. We maintain a counter Q, which initially is set to 0.

We next go around the cycle, starting at the vertex $j_0 = r$, and visiting all the vertices $j_0, j_1, ..., j_m$ in (say) clockwise order. When we visit a vertex j_k, we set $Q \leftarrow Qw(j_k)$. Suppose the counter Q, just before reaching j_k was Q_{old}, and $Q_{new} = Q_{old} + w_{j_k}$ is the value after accounting for j_k. If $xU + Y \in (Q_{old}, Q_{new})$ for some integer xi.e., the counter crossed the point modulo then we mark j_k, and ask that it send $Q_{new} - xU + Y$ weight to the next marked vertex lying clockwise on the cycle. In the other case, we ask that the vertex send all its weight to the next marked vertex lying clockwise on the cycle. Note that the construction ensures that each arc on the cycle carries at most U units of weight; furthermore, a vertex j gets marked with probability $W(j)/U$, and this is exactly the probability that it has U units of weight at the end of the process. This process naturally induces a redistribution of weights in the original tree as well; however, since each edge of the tree was replaced by two opposite arcs a and $\bar{a}$. Suppose that both arcs carry flow, with a path from i to j

using $\overline{a}$. We can decrease the flow sent on these path by ϵ, and instead sent ϵ flow from i to j', and from i' to j. This does not change the amount of weight reaching a marked vertex, but decreases the total flow crossing e. This process stops when each edge is used in only one direction, at which point the flow crossing each edge of T is at most U, completing the proof of the lemma.

■

A. Algorithm Description

By extending the random sampling approximation algorithm in [2], we give a constant-factor approximation algorithm for pMNDP with multiple cable type. Our algorithm is randomized and contains five main steps. In the first step, it starts from the entire set of demands D. Then in each stage k of the second step, iteratively, we use only cable of type k and $k+1$ to collect the demands in D_k into a new set D_{k+1}. This demand collection is performed using Lemma 1, which ensures that the total demands in each node is a multiple of u_{k+1}. To carry out the third step, we solve a p-median instance with facility set F and client set D_K. Step 4 marks each client in D_K with a probability P_K. Finally Step 5 opens facilities such that there exist some marked clients which were assigned to this facilities in Step 3.

Definition 1: let F^* be the set of open facilities in the optimal solution and let $l(h, j)$ be the distance between nodes h and j in G. The distance between two nodes h and j is defined by the length of shortest path joining them. Denote $l(h, F) = \min_{j \in F} l(h, j)$.

B. Analysis

Let ρ_{ST} be the approximation ratio of the approximation algorithm for the Steiner tree problem. We need the following lemmas given in [2] for analyzing the above algorithm.

Lemma 2: [2] For every client $j \in D$ and Stage k, $1 \leq k \leq K$, we have $Pr[j \in D_k] = 1/U_k$.

Lemma 3: [2] The expected cost incurred in stage k is at most $(\rho_{ST}+3)\gamma_{k+1}E[C(\check{T}_k)]$, where $\check{T}_k$ is the optimal Steiner tree on D'_k.

Let C^*_k be the cost of installing cable type k in the optimal solution. The expected cost $E[B_k]$ incurred in Stage $k \in \{1, \dots, K\}$ of Step 2 of the algorithm is bounded as follows.

Lemma 4:

$$\mathbb{E}[B_k] \leq \gamma_{k+1}(\rho_{ST}+3)\left[\sum_{t>k}\frac{1}{\gamma_t}C^*_t + \frac{\varrho_k}{\gamma_{k+1}}\sum_{t\leq k}\frac{1}{\gamma_t}C^*_t\right].$$

Proof: Stage k of Step 2 of the algorithm builds a Steiner tree $\check{T}_k$ whose cost is $C(\check{T}_k)$ and the cost of the cable built on each edge of this Steiner tree is at most $\rho_{ST}.\gamma_{k+1}.C(\check{T}_k)$. Using an argument similar to the one used in Lemma 3, it can be shown that the cable type k used to collect demands on D'_k incurs a cost of at most $2\gamma_{k+1}.C(\check{T}_k)$ and the cost of the cable type $k+1$ used to send back the demands to the random node in D_k is at most $\gamma_{k+1}.C(\check{T}_k)$. Hence, the total expected cost denoted B_k is bounded as follows:

$$\mathbb{E}[B_k] \leq \gamma_{k+1}(\rho_{ST}+3)C(\check{T}_k). \qquad (6)$$

Algorithm 1 Algorithm pMNDP

Data: A graph $G = (V, E)$, a set of facilities F, a set of demands D and a set K of cable types.

Output: A number of cables x^k_e installed on the edges, a set of p opened facilities.

Step 1. Let D_1=D;

Step 2. For stage $k = 1, 2, ..., K$ *do*

1) Mark each client in D_k with probability $P_k = \frac{\gamma_k}{\gamma_{k+1}}$, and let D'_k be the set of marked clients.
2) Construct a Steiner tree T_k on D'_k, and build cable type $k+1$ on each of its edges.
3) Send the demands of each client to its nearest $j \in D'_k$ via a shortest path using cable type k. Let $d_{k(j)}$ be the demands collected at $j \in D'_k$ and $D_k(j)$ be the clients sending demands to j in stage k.
4) Redistribute demands in D'_k by applying Lemma 1, where $T = T_k$, $U = U_{k+1}$, and $w(j) = d_k(j) \mod u_{k+1}$. Demands are routed using cable of type $k+1$ built at phase 2 of the current stage.
5) Divide $D_k(j)$ into groups of U_{k+1}/U_k nodes, and each group sends an amount of U_{k+1} demands back from k to a random member of the group via the shortest paths. Build a new cable of type $k+1$, and the resulting demand location is D_{k+1}.

Step 3. We solve a p-median instance on D_K with facility set F and client set D_K, and the assignment cost is $C_{kj} = \gamma_K.l(j, k), \forall j \in D_k, k \in F$.

Step 4. Mark each client in D_K with probability $P_k = \frac{\gamma_K}{u_K}$, and D'_K is the set of marked clients, and let F_u be the set of open facilities.

Step 5. Open facility $k \in F_U$ if some of the clients assigned to it are marked. Let I be the set of open facilities.

In order to find an upper bound for $C(\check{T}_k)$, we construct a Steiner tree T with terminals set D'_k in Step 2 (2). First, we add edges with cable type $k+1$ or higher in the optimal solution OPT to T. It is clear that the cost of this subgraph is at most $\sum_{t>k}\frac{1}{\gamma_t}C^*_t$. We augment T by adding all the missing edges from paths connecting each client j to an open facility i in OPT, where these edges have only cable of type $t \leq k$. For the sake of simplicity, we assume that only one cable type $t \leq k$ is built on each of these edges. Hence by Lemma 2, $P[e \in T] \leq \frac{u_t}{u_k}.\frac{\gamma_k}{\gamma_{k+1}}$. Finally, summing over all the edges with cables of type at most k, the expected cost of edges with cable of type $t \leq k$ is bounded by $\sum_{t\leq i}\frac{u_t}{u_k}.\frac{\gamma_k}{\gamma_{k+1}}.\frac{1}{\sigma_t}C^*_t$.

Now from T we extract a tree $\check{T}_k$ spanning D'_k to obtain the following inequality.

$$\mathbb{E}[C(\check{T}_k)] \leq \mathbb{E}[c(T)] \leq \sum_{t>k}\frac{1}{\gamma_t}C^*_t + \frac{\varrho_k}{\gamma_{k+1}}\sum_{t\leq k}\frac{1}{\gamma_t}C^*_t, \qquad (7)$$

which implies the claimed bound. ■

We denote by C_U the assignment cost incurred by the solution of the p-median problem computed in Step 3. The following lemma provides a lower bound for this assignment cost.

Lemma 5:

$$\mathbb{E}[C_U] \leq \rho_{pM} \sum_{k=1}^{K} \frac{\varrho_K}{\gamma_k} C_k^*,$$

where ρ_{pM} is the approximation ratio of the approximation algorithm for the p-median problem

Proof: The expected cost of assigning each client to its nearest open facility is at most

$$\gamma_K \mathbb{E}\left[\sum_{j \in D_K} l(j, F^*)\right] = \sum_{j \in D} \varrho_K l(j, F^*) \leq \sum_{k=1,\ldots,K} \frac{\varrho_K}{\gamma_k} C_k^*. \tag{8}$$

Hence, the total cost of approximating the solution in Step 3 is at most

$$\rho_{pM} \left(\sum_{k=1}^{K} \frac{\varrho_K}{\gamma_k} C_k^*\right), \tag{9}$$

implying the desired bound. ■

Together with the above Lemmas, now we are ready to give the approximation ratio of the Algorithm pMNDP.

Theorem 1: Algorithm pMNDP is a $(21 + 2\epsilon)$-approximation algorithm for the pMNDP with multiple cable type.

Proof: The total expected cost incurred in Step 2 of the algorithm is bounded by $(\rho_{ST} + 3) \sum_{k=1}^{K} \left(\sum_{t>k} \frac{\gamma_{k+1}}{\gamma_k} + \sum_{t \leq k} \frac{\varrho_t}{\varrho_k}\right) C_k^* \leq 4(\rho_{ST} + 3) \sum_k C_k^*$. Using Lemma 5 and the fact that ϱ_k and γ_k are powers of 2, the total cost of building cables in Step 2 can be bounded by

$$4(\rho_{ST} + 3) \sum_k C_k^*. \tag{10}$$

In addition, the total cost incurred by installing cable that connects open facilities in the p-median instance can be bounded by

$$2\rho_{pM} \sum_i C_k^*. \tag{11}$$

Together bounds (10) and (11) imply the following bound to the pMNDP solution with multiple cable type:

$$(4\rho_{ST} + 12 + 2\rho_{pM}) C^*. \tag{12}$$

The currently best approximation ratio for the Steiner tree problem is $\rho_{ST} = ln(4)$ [10], and the best approximation ratio for the p-median problem is $\rho_{pM} = 1 + \sqrt{3} + \epsilon$ [19]. Hence the approximation ratio of Algorithm pMNDP is no more than $21 + 2\epsilon$. ■

IV. Conclusion

In this paper, we address the p-median in network design with multiple cable types, that has not been addressed in the literature before. We provide the first constant unifator approximation algorithm for it. Our future interesting work is to further improve the approximation factor for the problem with single and multiple cable types, respectively.

References

[1] A. Agrawal, P. Kelin and R. Ravi, *When trees collide: An approximation algorithm for the generalized steiner problem on networks.* SIAM J. Comput, 1995.

[2] A. Gupta, A. Kumar and T. Rougharden, *Simpler and better approximation algorithms for network design*, 35th ed. ACM sympos, on the theory of comput, Sna Diego, CA, 2003.

[3] A. Hassin, R. Ravi and F.S. Salman, *Approximation algorithm for capacitated network design problem.* Algorithmica, 2006.

[4] B. Awerbuch and Y. Azar, *Buy-at-bulk network design*, 38th ed. in Proc.IEEE Symposium on foundations of Computer Science (FOCS), 1997.

[5] C. Selene, M. Vallejo and J. Fernando, *Solving the p-median bilevel problem with order through a hybrid heuristic* . Applied Soft Computing, 2017.

[6] C. Swamy, *Improved approximation algorithms for matroid and knapsack median problems and applications.* ACM Transactions on Algorithms, 2016.

[7] C. Swamy and A. Kumar, *Primal dual algorithms for connected facility location problems.* Algorithmica, 2004.

[8] C. Wu, D. Du and D. Xu, *An approximation algorithm for the k-median problem with uniform penalties via pseudo-solution.* Theoretical Computer Science, 2018.

[9] G. Blelloch, A. Gupta and K. Tangwoongsan, *Parallel probabilistic tree embedding, k-median and buy-at-bulk network design.* 24th ed. ACM Symposium on Parallelism in Algorithms and Architectures-SPAA, 2012.

[10] J. Byrka and K. Aardal, *An improved LP-based approximation for the steiner tree*, 10th ed. Proceedings of the forty-second ACM symposium on Theory of computing, 2010.

[11] K. Talwar, *Single-sink buy-at-bulk LP has constat integrality gap*, 9th ed . in Proc. International Conference on integer Programming and Combinatorial Optimization (IPCO), 2002.

[12] F. S. salman, J. Cheryan, R. Ravi and S. Subramanian, *Approximating the single-sink link installation problem in network design.* SIAM J. Optim, 2000.

[13] N. Garg, R. Khandekar, G. Konjevod, R. Ravi, F. S. Salman, A. Sinha *On the integrality gap of natural formulation of the single-sink buy-at-bulk network design problem*, 8th ed. International Conference on Integer programming and Combinatorial Optimization (IPCO), 2001.

[14] P. kaveh, *Solving capacitated p-median problem by hybrid k-median clustering and FNS algorithm.* International journal of Innovation, 2010.

[15] R. Jpothi and B. Raghavachari, *Improved approximation algorithms for the single-sink buy-at-bulk network design problems.* Journal of Discrete Algorithms, 2009.

[16] R. Ravi and S. Sinha, *Approximation algorithms combining facility location and network design.* Oper. Rzs, 2006.

[17] S. C. Chang, W. C. K. Yen, Y. L. Wang and J. J. Liu, J. J, *The NP-hardness of the connected p-median problem on bipartie graphs and split graphs.* Chiang Mai J. Sci,2003.

[18] S. Guha, A. Meyerson and k. Mungala, *A constant factor approximation for the single sink edge installation problem* . SIAM J.Comput, 2009.

[19] S. Li and O. Svensson, *Approximating k-median via pseudo-approximation*, 45th ed . Proceedings of the 45th annual ACM symposium on Symposium on theory of computing-STOC, 2013.

[20] Z. Drezner, J. Brimberg, N. Mladenović and S. Salhi *Solving the planar p-median problem by variable neighborhood and concentric searches.* Journal of Global Optimization, 2015.