

Approximating connected facility location with buy-at-bulk edge costs via random sampling

Andreas Bley^{1,3} and Mohsen Rezapour^{2,3}

*Institute for Mathematics, TU Berlin, Germany
Preprint 2013/02, March 1, 2013*

Abstract

In the connected facility location problem with buy-at-bulk edge costs we are given a set of clients with positive demands and a set of potential facilities with opening costs in an undirected graph with edge lengths obeying the triangle inequality. Moreover, we are given a set of access cable types, each with a cost per unit length and a capacity such that the cost per capacity decreases from small to large cables, and a core cable type of infinite capacity. The task is to open some facilities and to connect them by a Steiner tree using core cables, and to build a forest network using access cables such that the edge capacities suffice to simultaneously route all client demands unsplit to the open facilities. The objective is to minimize the total cost of opening facilities, building the core Steiner tree, and installing the access cables. In this paper, we devise a constant-factor approximation algorithm for this problem based on a random sampling technique.

Keywords: network design, connected facility location, approximation algorithm.

1 Introduction

In this paper we study a new network design problem, *Connected Facility Location with buy-at-bulk edge costs* (BBConFL), which arises in the design of

¹ Supported by the DFG research center MATHEON–‘Mathematics for key technologies’.

² Supported by the DFG research training group ‘Methods for Discrete Structures’.

³ Emails: {bley,rezapour}@math.tu-berlin.de

fiber optic telecommunication access networks. In this context, the telecommunication network consists of a regional backbone network and several local access networks. The traffic originating from the clients is sent through the access networks towards the gateway nodes, which provide routing functionalities and access to the backbone. The backbone network consists of very high speed links of (almost) unlimited capacity between the core nodes, which are used to route the traffic further towards its destination. Designing such a network involves selecting the gateway nodes, connecting them to each other, and choosing and dimensioning the links from the clients to this core network.

We model the scenario above as the BBConFL problem. We are given an undirected graph $G = (V, E)$ with nonnegative edge lengths $c_e \in \mathbb{Z}_{\geq 0}$, $e \in E$, obeying the triangle inequality, a set $F \subset V$ of facilities with opening costs $f_i \in \mathbb{Z}_{\geq 0}$, $i \in F$, and a set of clients $D \subset V$ with demands $d_j \in \mathbb{Z}_{> 0}$, $j \in D$. We are also given K types of cables that may be used in the access network connecting clients to open facilities. A cable of type i has capacity $u_i \in \mathbb{Z}_{> 0}$ and cost (per unit length) $\sigma_i \in \mathbb{Z}_{\geq 0}$. Core links, which are used to connect the open facilities, have a cost (per unit length) of $M \in \mathbb{Z}_{\geq 0}$. The task is to find a subset $I \subseteq F$ of facilities to open, a Steiner tree $S \subseteq E$ connecting the open facilities I , and a forest $E' \subseteq E$ with a cable installation on its edges, such that E' connects all clients D to the open facilities I and the installed capacities suffice to simultaneously route all client demands to the open facilities. We are allowed to install multiple copies and types of access cables on each edge of E' . The objective is to minimize the total cost, where the cost for using edge e in the Steiner tree is Mc_e and the cost for installing a single cable of type i on edge e is $\sigma_i c_e$.

The (unsplittable) *Single-Sink Buy-at-Bulk* problem (u-SSBB) is a special case of BBConFL where $|F| = 1$. The first constant factor approximation for u-SSBB was proposed by Guha et al. [1]. Talwar [2] improved the factor to 216, using an LP rounding technique. Exploiting the random-sampling framework by Gupta et al. [3], which is also used in our approach, the factor was later reduced to 145.6 in [4], and then to 40.82 by Grandoni et al. [5].

Also, the *Connected Facility Location* problem (ConFL) can be considered as a special case of BBConFL with only one cable type of unit capacity. ConFL has received a lot of attention in the literature; see e.g., [6,7,8,9]. The current best 3.19-approximation for ConFL is due to Grandoni et al. [9].

If we omit the requirement to connect the open facilities by a core Steiner tree, then the BBConFL problem reduces to a *k-cable facility location problem*. For this problem, Ravi et al. [10] provide an $O(k)$ approximation extending the algorithm of Guha et al. [1].

In this paper we present a randomized constant factor approximation algorithm for BBConFL, which is a natural generalization of the two NP-hard problems u-SSBB and ConFL.

2 Preliminaries

Let $\delta_i = \sigma_i/u_i$ be the cost per capacity of cable type i . We can assume w.l.o.g. that $u_i < u_{i+1}$ and $\sigma_{i+1} > \sigma_i$ for all i . In addition, we assume that $\sigma_K < M$, which is natural in our and many other applications. Furthermore, we assume that all values u_i and σ_i are powers of 2. This can be easily enforced increasing the cost of any solution by factor of at most 4 and it permits to use the demand redistribution technique introduced by Jothi et al. [4].

Lemma 2.1 ([4]) *Let T be a tree with each edge having capacity U . For each vertex v in T , let $x(v) < U$ be the demand originating at v . Assume all values U and $x(v)$, $v \in T$, are powers of 2. There is an efficient randomized algorithm that computes a flow on T that respects the edge capacities and redistributes the demands (without splitting any demand) such that each vertex receives a new demand $\hat{x}(v) \in \{0, U\}$ and, moreover, $\Pr[\hat{x}(v) = U] = \frac{x(v)}{U}$ for all $v \in T$.*

We can also assume that $u_1 = \sigma_1 = 1$ and that each value d_j , $j \in D$, is a power of 2, losing another factor of at most 2 in the approximation ratio.

For the sake of simplicity, we finally assume w.l.o.g. that $d_j = 1$ for all j , replacing j by several copies of co-located unit-demand clients. The algorithm presented in this paper can easily be adapted to ensure that those demands travel together along the same path towards an open facilities; see [4] for additional details. By adding dummy demands, we can assume that also the number of demands $|D|$ is a power of 2.

3 Approximating BBConFL

Extending the algorithm proposed in [3] using techniques from [4,8], we developed an approximation algorithm for BBConFL.

Our algorithm, shown on page 4, builds the solution in a bottom-up manner in K stages. Starting with all demand nodes $D_1 = D$ in Step 1, in stage i of Step 2 it aggregates the demands from the demand nodes D_i into a smaller, randomly sampled node set D_{i+1} using only cables of type i and $i + 1$. Eventually, in Steps 3–7, the demands are aggregated from the nodes in D_k to a randomly sampled subset of the facilities, which is then connected by a core Steiner tree.

One easily verifies that the algorithm runs in polynomial time. To analyze its performance, we introduce some further notation. Let OPT denote an optimal solution with cable cost $C^* = \sum_{j=1}^K C_j^*$, where C_j^* is the amount paid for cables of type j in OPT , core Steiner cost S^* , and opening cost O^* . Let F^* and T_{core}^* denote the set of open facilities and the Steiner tree connecting them in OPT , respectively. For each $j \in D$, let $\sigma^*(j)$ be the facility j is assigned to in OPT . The following lemma has been shown in [3].

Algorithm 1

1. Let $D_1 = D$. Guess a facility r to open in the solution.
 2. For stage $i = 1, 2, \dots, K - 1$ do
 - Mark each client in D_i with prob. $\frac{\sigma_i}{\sigma_{i+1}}$. Let D'_i be the marked clients.
 - Construct a ρ_{ST} -approximate Steiner tree T_i spanning terminals $F_i = D'_i \cup \{r\}$. Install cable type $i + 1$ on each $e \in T_i$.
 - For each $k \in D_i$, send its demand to the nearest $j \in F_i$ via a shortest path, using cables of type i . Let $d_i(j)$ be demand aggregated at $j \in F_i$.
 - Redistribute demands $x(j) := d_i(j) \bmod u_{i+1}$, $j \in F_i$, within tree T_i of capacity u_{i+1} (cf. Lemma 2.1). The corresponding flow is supported by cables of type $i + 1$. Let $\bar{d}_i(j)$ be the demand aggregated at node j .
 - For each $j \in F_i$, let $D_i(j) \subseteq D_i$ be the nodes sending demand to j in stage i (including j itself, if $j \neq r$). Partition $D_i(j)$ into groups of $\frac{u_{i+1}}{u_i}$ nodes. Send the total demand u_{i+1} in each group back to a random member of that group via shortest paths, installing cables of type $i + 1$. Let D_{i+1} be the resulting demand locations.
 3. Compute a ρ_{FL} -approximate solution $U = (F_U, \sigma_U)$ to the unconnected facility location problem on D_K and the edge cost per unit length σ_K .
 4. Mark each client in D_K with prob. $\frac{\sigma_K}{M}$. Let D'_K be the marked clients.
 5. Open facility $i \in F_U$ if some client in $\sigma_U^{-1}(i)$ is marked. Let I be the set of open facilities.
 6. Compute a ρ_{ST} -approximate Steiner tree T_K on terminals $F_K = D'_K \cup \{r\}$. Construct a tree T'_K on I by adding the shortest path between every $j \in D'_K$ and the corresponding open facility $\sigma_U(j) \in I$.
 7. Connect each client in D_K to a closest open facility in $I \cup \{r\}$ using cables of type K .
-

Lemma 3.1 For every client $j \in D$ and stage i , $\Pr[j \in D_i] = 1/u_i$.

The expected cost of the optimal Steiner tree on F_i can be bounded as follows.

Lemma 3.2 Let T_i^* be the optimal Steiner tree on F_i and $c(T_i^*)$ its cost. Then

$$\mathbf{E}[c(T_i^*)] \leq \frac{1}{M} S^* + \sum_{t>i} \frac{1}{\sigma_t} C_t^* + \frac{\delta_i}{\sigma_{i+1}} \sum_{t \leq i} \frac{1}{\delta_t} C_t^*$$

Proof. We construct a feasible Steiner tree T on F_i as follows. First, add the optimal Steiner tree T_{core}^* and all edges with cable type $i + 1$ or higher in OPT

to T . The resulting subgraph has cost at most $\frac{1}{M}S^* + \sum_{t>i} \frac{1}{\sigma_t}C_t^*$.

Then, we add all missing edges from the paths P_j^* connecting each client $j \in F_i$ with $\sigma^*(j)$ in OPT . Note that these edges have cable type $t \leq i$ in OPT . For simplicity, suppose only one cable type $t \leq i$ is installed on edge e . Then $e \in T$ if and only if $j \in F_i$ for some $j \in D$ where $e \in P_j^*$. Since each $j \in F_i$ with probability $\frac{1}{u_i} \frac{\sigma_i}{\sigma_{i+1}}$, it follows from the Union Bound that $e \in T$ with probability at most $\frac{u_t}{u_i} \frac{\sigma_i}{\sigma_{i+1}}$. Thus the expected cost of adding these edges is at most $\frac{\delta_i}{\sigma_{i+1}} \sum_{t \leq i} \frac{1}{\delta_t} C_t^*$. Extracting a tree spanning F_i from T , we obtain

$$\mathbf{E}[c(T_i^*)] \leq \mathbf{E}[c(T)] \leq \frac{1}{M}S^* + \sum_{t>i} \frac{1}{\sigma_t}C_t^* + \frac{\delta_i}{\sigma_{i+1}} \sum_{t \leq i} \frac{1}{\delta_t}C_t^* \quad \square$$

Let A_i be the total cost incurred in Stage i of Step 2 of the algorithm. The following lemma, shown by Gupta et al. [3], applies also to our problem. (Again, T_i^* denotes the optimal Steiner tree on F_i .)

Lemma 3.3 $\mathbf{E}[c(A_i)] \leq \sigma_{i+1}(\rho_{ST} + 3) \mathbf{E}[c(T_i^*)]$

Let O_U be the opening costs and C_U be the connection costs of the solution to the unconnected facility location problem UFL computed in Step 3.

Lemma 3.4 $\mathbf{E}[O_U + C_U] \leq \rho_{FL}(O^* + \sum_{i=1}^K \frac{\delta_K}{\delta_i} C_i^*)$

Proof. We obtain a feasible solution for UFL by connecting each client $j \in D_K$ to its BBConFL optimal facility $\sigma^*(j) \in F^*$. Its expected cost is at most

$$O^* + \sigma_K \cdot \mathbf{E}\left[\sum_{j \in D_K} l(j, F^*)\right] = O^* + \delta_K \sum_{j \in D} l(j, F^*) \leq O^* + \sum_{i=1, \dots, K} \frac{\delta_K}{\delta_i} C_i^* .$$

The last inequality follows from the fact that $\sum_{j \in D} l(j, F^*) \leq \sum_{i=1}^K \frac{u_i}{\sigma_i} C_i^*$, where $l(j, F^*)$ is the shortest path distance from j to $\sigma^*(j) \in F^*$. \square

Lemma 3.5 *The cost of cable installation in Step 7 of the algorithm satisfies $\mathbf{E}[C] \leq 2 \sum_{i=1}^K \frac{\delta_K}{\delta_i} C_i^* + 0.807 S^* + C_U$.*

Proof. We apply the core connection game described in [8] with clients D_K , core T_{core}^* , mapping $\sigma = \sigma^*$, $w(e) = c(e)$, and probability $\frac{\sigma_K}{M}$. This yields

$$\mathbf{E}\left[\sum_{j \in D_K} \sigma_K \cdot l(j, I \cup \{r\})\right] \leq 2 \frac{\sigma_K}{u_K} \sum_{j \in D} l(j, F^*) + \sigma_K \frac{0.807}{\sigma_K/M} \cdot \frac{S^*}{M} + C_U .$$

With $\sum_{j \in D} l(j, F^*) \leq \sum_{i=1}^K \frac{u_i}{\sigma_i} C_i^*$, we obtain the claimed bound. \square

For the Steiner tree T'_K computed in Step 6 we obtain the following bound.

Lemma 3.6 $\mathbf{E}[T'_K] \leq \frac{\rho_{ST}}{M}(S^* + \sum_{i=1}^K \frac{\delta_K}{\delta_i} C_i^*) + \frac{1}{M}C_U$

Proof. We construct a feasible Steiner tree on the marked clients in D'_K by augmenting the optimal Steiner tree T_{core}^* by the shortest paths from each client in D'_K to T_{core}^* . This tree has expected cost at most

$$\frac{1}{M}S^* + \mathbf{E}\left[\sum_{j \in D'_K} l(j, F^*)\right] = \frac{1}{M}S^* + \frac{\sigma_K}{u_K \cdot M} \sum_{j \in D} l(j, F^*).$$

Thus the expected cost of a ρ_{ST} -approximate Steiner tree on D'_K is at most $\frac{\rho_{ST}}{M}(S^* + \sum_{i=1}^K \frac{\delta_K}{\delta_i} C_i^*)$. Furthermore, the expected cost of connecting each client $j \in D'_K$ to $\sigma_U(j) \in F$ is at most $\frac{\sigma_K}{M} \sum_{j \in D'_K} l(j, F_U) = \frac{1}{M}C_U$. Altogether, we obtain the claimed bound. \square

Together, Lemmas 3.2–3.6 imply our main result.

Theorem 3.7 *Algorithm 1 is a constant approximation for BBConFL.*

References

- [1] S. Guha, A. Meyerson, K. Munagala. A constant factor approximation for the single sink edge installation problems. In Proc. of STOC 2001, pages 383-388.
- [2] K. Talwar. The single-sink buy-at-bulk LP has constant integrality gap. In Proc. of IPCO 2002, pages 475-486.
- [3] A. Gupta, A. Kumar, and T. Roughgarden. Simpler and better approximation algorithms for network design. In Proc. of STOC 2003, pages 365-372.
- [4] R. Jothi, B. Raghavachari. Improved approximation algorithms for the single-sink buy-at-bulk network design problem. In Proc. of SWAT 2004, pages 336-348.
- [5] F. Grandoni and T. Rothvoß. Network design via core detouring for problems without a core. In Proc. of ICALP 2010, pages 490-502.
- [6] A. Gupta, J. Kleinberg, A. Kumar, R. Rastogi, and B. Yener. Provisioning a virtual private network: A network design problem for multicommodity flow. In Proc. of STOC 2001, pages 389-398.
- [7] C. Swamy and A. Kumar: Primal-dual algorithms for connected facility location problems. *Algorithmica* 40, pages 245-269, 2004.
- [8] F. Eisenbrand, F. Grandoni, T. Rothvoß, G. Schäfer. Connected facility location via random facility sampling and core detouring. *Journal of Computer and System Sciences*, 76: 709-726, 2010.
- [9] F. Grandoni, T. Rothvoß. Approximation algorithms for single and multi-commodity connected facility location. In Proc. of IPCO 2011.
- [10] R. Ravi, A. Sinha. Integrated logistics: approximation algorithms combining facility location and network design. In Proc. of IPCO 2002.