Approximation Algorithms for Single and Multi-Commodity Connected Facility Location

Fabrizio Grandoni[1] and Thomas Rothvoß[2]

[1] University of Rome Tor Vergata, Italy, grandoni@disp.uniroma2.it
[2] MIT, USA, rothvoss@math.mit.edu

Abstract In the classical *facility location* problem we are given a set of facilities, with associated opening costs, and a set of clients. The goal is to open a subset of facilities, and to connect each client to the closest open facility, so that the total connection and opening cost is minimized. In some applications, however, open facilities need to be connected via an infrastructure. Furthermore, connecting two facilities among them is typically more expensive than connecting a client to a facility (for a given path length). This scenario motivated the study of the *connected facility location* problem (CFL). Here we are also given a parameter $M \geq 1$. A feasible solution consists of a subset of open facilities and a Steiner tree connecting them. The cost of the solution is now the opening cost, plus the connection cost, plus M times the cost of the Steiner tree.

In this paper we investigate the approximability of CFL and related problems. More precisely, we achieve the following results:
- We present a new, simple 3.19 approximation algorithm for CFL. The previous best approximation factor is 3.92 [Eisenbrand, Grandoni, Rothvoß, Schäfer-'10].
- We show that SROB, i.e. the special case of CFL where all opening costs are 0, is hard to approximate within 1.28. The previous best lower bound for SROB is 1.01, and derives trivially from Steiner tree inapproximability [Chlebík, Chlebíková-'08]. The same inapproximability result extends to other well-studied problems, such as *virtual private network* and *single-sink buy-at-bulk*.
- We introduce and study a natural multi-commodity generalization MCFL of CFL. In MCFL we are given source-sink pairs (rather than clients) that we wish to connect. A feasible solution consists of a subset of open facilities, and a forest (rather than a tree) spanning them. Source-sink connection paths can use several trees in the forest, but must enter and leave each tree at open facilities. We present the first constant approximation for MCFL.

1 Introduction

In the classical *metric facility location* problem (FL), we are given an undirected graph $G = (V, E)$, with edge costs (or weights) $c : E \to \mathbb{Q}^+$, a set of clients $C \subseteq V$, and a set of facilities $F \subseteq V$, with opening costs $o : F \to \mathbb{Q}^+$. A

feasible solution is given by a subset $F' \subseteq F$ of *open* facilities. The goal is to minimize the opening cost of F', plus the shortest path distance from each client to the closest open facility. More formally, let $c(v, u)$ denote the shortest path distance between u and v, and $c(v, U) := \min_{u \in U} c(v, u)$ for any $U \subseteq V$. Then the objective function to be minimized is $\sum_{f \in F'} o(f) + \sum_{v \in C} c(v, F')$.

In several applications one needs to connect open facilities via an infrastructure. Typically, connecting facilities among them is more expensive than connecting clients to facilities. This scenario motivated the introduction of the following problem. Let $c(E') := \sum_{e \in E'} c(e)$ for any $E' \subseteq E$, and $c(G') = c(E(G'))$ for any subgraph G' of G.

> CONNECTED FACILITY LOCATION (CFL). Given an undirected graph $G = (V, E)$, with edge costs $c : E \to \mathbb{Q}^+$, a set of clients $C \subseteq V$, a set of facilities $F \subseteq V$, with opening costs $o : F \to \mathbb{Q}^+$, and a parameter $M \geq 1$. Compute a subset $F' \subseteq F$ of open facilities, and a tree T' spanning F', in order to minimize $\sum_{f \in F'} o(f) + \sum_{v \in C} c(v, F') + M \cdot c(T')$.

CFL is well-studied in the literature [9,18,21,28]. The current best approximation for it is 3.92 [9]. A very well-studied [9,18,20,23,28] special case of CFL is the *single-sink rent-or-buy* problem (SROB), where $F = V$ and opening costs are zero. In this context, we can think of edges of T' as *bought* edges (for which we pay a fixed, large cost), and edges outside T' as *rented* edges (for which we pay a cost proportional to the number of paths using them).

Another way to interpret CFL is as follows. Clients are users who want to reach a public transportation network T' each day to get to their office. Commuting has a social cost which is shared on T', and payed on an individual basis outside T'. Here open facilities are stations at which users can access T'. This view of CFL suggests a natural multi-commodity generalization of the problem. Replace clients with origin-destination pairs, and imagine that you can construct several, possibly disconnected, transportation networks. Each network can be reached and left at stations. More formally, one can define the following problem. For a forest T' and a subset of nodes V', let $c_{V',T'}(u, v)$ be the shortest path distance between nodes u and v after adding one edge of cost zero between each pair of nodes in V' belonging to the same tree of T'.

> MULTI-COMMODITY CONNECTED FACILITY LOCATION (MCFL). Given an undirected graph $G = (V, E)$, with edge costs $c : E \to \mathbb{Q}^+$, a set of source-sink pairs[3] $P = \{(s_1, r_1), \ldots, (s_k, r_k)\}$, $s_i, r_i \in V$, a set of facilities $F \subseteq V$, with opening costs $o : F \to \mathbb{Q}^+$, and a parameter $M \geq 1$. Compute a subset $F' \subseteq F$ of open facilities, and a forest T', in order to minimize $\sum_{f \in F'} o(f) + \sum_{(s,r) \in P} c_{F',T'}(s, r) + M \cdot c(T')$.

To the best of our knowledge, MCFL was not addressed before (at least, from the point of view of approximation algorithms). However, there is a special case of the problem which is well-studied in the literature: the *multi-commodity*

---

[3] For notational convenience, we will consider P as a multi-set of pairs.

rent-or-buy problem (MROB) is the special case of MCFL where $F = V$ and opening costs are zero [2,3,12,20,24]. A solution to an MROB instance consists of a forest T' of bought edges. The cost of the solution is given by $M \cdot c(T') + \sum_{(s,r) \in P} c_{T'}(s, r)$, where $c_{T'}(u, v)$ denotes the shortest path distance between u and v, after contracting the connected components of T'. In other terms, MROB is the multi-commodity version of SROB.

1.1 Our Results and Techniques

In this paper we study the approximability of CFL and related problems. In particular, we obtain the following three main results.

(1) An Improved Approximation for CFL. We present a 3.19 approximation algorithm for CFL, improving on the previous best 3.92 approximation [9]. The approximation algorithms for CFL by Gupta, Srinivasan and Tardos [21] and Eisenbrand, Grandoni, Rothvoß, and Schäfer [9] are both based on simple random sampling steps (more details in Section 1.2). Here we present a third, simple random sampling algorithm. We first randomly sample clients, and buy a Steiner tree T over them. Then we define a facility location instance (on all the clients), where the opening cost of each facility f is increased by the cost of augmenting T to include f. This way, the modified opening cost encodes both the real opening cost, and the cost of connecting f to the other open facilities via T.

Like in [9], our technique can be extended to the connected version of some variants of facility location. However, our approach is more flexible. For example, differently from [9], it gives a constant approximation for connected facility location with hard capacities. Due to space limits, extensions will be discussed in the full version of the paper.

From the analytical point of view, we exploit the *core-detouring* technique in [9], which was already successfully applied in the analysis of CFL [9] and related problems [15]. The basic idea is bounding the cost of connecting a set of clients to a random subset of them. This bound is based on detouring connection paths through a proper connected *core* graph. We cannot directly apply (as a black box) the core-detouring theorem in [9], since we need to connect clients to facilities rather than clients among them. However, the connection scheme used in the proof of the theorem has some particular properties that we can exploit for our purposes.

(2) A Constant Approximation for MCFL. We present the first constant approximation for MCFL: the approximation factor is 16.2. Our result is based on two main ingredients. The first ingredient is a reduction to the *prize-collecting* version of facility location, where we are allowed not to connect all the clients, but we have to pay a penalty for each disconnected client. More formally:

PRIZE-COLLECTING FACILITY LOCATION (PFL). Given an undirected graph $G = (V, E)$, with edge costs $c : E \to \mathbb{Q}^+$, a set of clients $C \subseteq V$, with penalties $p : C \to \mathbb{Q}^+$, and a set of facilities $F \subseteq V$, with opening costs $o : F \to \mathbb{Q}^+$. Compute a subset $F' \subseteq F$ of open facilities and a

subset $C' \subseteq C$ of *disconnected* clients, in order to minimize $\sum_{f \in F'} o(f) + \sum_{v \in C-C'} c(v, F') + \sum_{v \in C'} p(v)$.

In a MCFL solution there might be pairs which are connected directly via a shortest path (without using edges of the forest T'): intuitively, prizes are used to get rid of those pairs. In particular, each source and sink will define a client, and each pair containing at least one disconnected client in the PFL solution will be connected directly via a shortest path.

The second ingredient is a reduction to MROB. For each residual pair, we consider the associated pair of facilities in the PFL solution. This defines an MROB instance. On this instance, we run the MROB algorithm `rand` by Fleischer, Könemann, Leonardi, and Schäfer [12] (see also [20]). Here, we crucially exploit some properties of `rand` which are implicitly proved in [12]. In particular, using a different (possibly better) approximation algorithm for MROB might lead to a worse approximation for MCFL.

(3) A Stronger Inapproximability Result for SROB. Observe that CFL is not a generalization of FL, since its definition excludes $M = 0$. However, the techniques in [16] can be adapted to prove the same 1.463-inapproximability bound for CFL as for FL (we omit the proof for lack of space).

Theorem 1. *Unless* $\mathbf{NP} \subseteq \mathbf{DTIME}(n^{O(\log n)})$, *there is no polynomial time* 1.463-*approximation for CFL.*

Here we show that SROB, a very special case of CFL, is hard to approximate within 1.278. This greatly improves over the previously known approximation hardness of 1.01, which is based on the Steiner tree hardness result in [6] combined with a trivial reduction. The same hardness result extends immediately to other well-studied generalizations of SROB, as MROB and *single-sink buy-at-bulk* (SSBB). It also applies to *virtual private network* (VPN). (See Section 1.2 for omitted definitions). We remark that CFL is not a special case of the latter problems (hence, the results in [16] do not extend to them). Our result is based on a reduction to a special case of facility location, where facility costs are uniform, and client-facility distances are either 1 or 2. We show that the latter problem is hard via a reduction to a set-cover-like problem whose hardness was proved by Guha and Khuller [16].

1.2   Related Work

CFL is well-studied in the literature. Gupta, Kleinberg, Kumar, Rastogi, and Yener [18] obtain a 10.66-approximation for this problem, based on rounding an exponential size LP. Gupta, Srinivasan and Tardos [21] describe a random sampling algorithm for CFL, leading to a 9.01-approximation. Their algorithm randomly samples clients, and then runs an (unconnected) facility location approximation algorithm on the sampled clients: the corresponding open facilities form the set of open facilities in the final CFL solution. Swamy and Kumar [28] later improved the approximation to 8.55, using a primal-dual algorithm. The

best-known result prior to our work is the 4.00-approximation by Eisenbrand, Grandoni, Rothvoß, and Schäfer [9]. They use a random-sampling approach subtly different from the one in [21]. In particular, they first solve an unconnected facility location problem on *all* the clients (not only on the sampled ones), and then randomly select a subset of the resulting (deterministic) pool of open facilities. Using the improved Steiner tree approximation algorithm by Byrka, Grandoni, Rothvoß, and Sanità [5], the approximation factor reduces to 3.92. In this paper we present a third, still simple random-sampling algorithm for CFL.

A lot of research was devoted to a special case of CFL, namely SROB. The first constant approximation for SROB is given by Karger and Minkoff [23]. Gupta et al. [18] give a 9.01-approximation algorithm. Swamy and Kumar [28] describe a primal-dual 4.55-approximation algorithm for the same problem. Gupta, Kumar, Pal, and Roughgarden [20] propose a simple random sampling algorithm which gives a 3.55-approximation. Based on a refinement of the analysis in [20] via the core-detouring technique, the current best 2.80 approximation is given in [9].

Single-sink buy-at-bulk (SSBB) is the generalization of SROB, where we are given a set of cable types, each one with a cost and a capacity. Capacity on edges has to be reserved by installing zero or more copies of each cable type. The goal is sending one unit of flow from each source node to a sink. SROB can be seen as the special case of SSBB with two cable types: one of very small capacity and unit cost per unit capacity (corresponding to rented edges) and one of fixed cost and very large capacity (corresponding to bought edges). After a long sequence of improvements [13,14,17,20,22,26,29], the current best 20.41 approximation was recently given in [15].

The *virtual private network* problem (VPN), despite its rather different formulation, is intimately related to the other mentioned problems (see, e.g., [15]). Here, we are given upper and lower bounds on the amount of traffic that each node can send and receive. A solution is given by a capacity reservation and one path for each source-sink pair. The goal is minimizing the cost of the capacity reservation so that every traffic matrix which satisfies the upper bounds can be routed along the specified paths without exceeding edge capacities. Also this problem is well-studied [7,8,18,20]. The current best approximation is 2.80 [15].

As mentioned before, MCFL was not addressed before to the best of our knowledge. However, its special case MROB is a well-known problem. A $O(\log n)$-approximation for MROB is obtained by combining the approach by Awerbuch and Azar [2] with the refined Bartal trees in [11]. The first constant approximation is given by Kumar, Gupta, and Roughgarden [24] via the primal-dual method. A better and simpler random sampling algorithm is presented by Gupta et al. [20]. Based on a similar approach, the constant was later improved to 6.83 by Becchetti, Könemann, Leonardi, and Pál [3], and eventually to 5 by Fleischer, Könemann, Leonardi, and Schäfer [12].

In [24] the term MCFL is used to define a variant of MROB, where each connection path can use at most one tree in the forest T': let us call this problem

---

Algorithm 1 Approximation algorithm for CFL.

---

1. Guess a facility r from the optimum solution. Sample each client with probability $\frac{\alpha}{M}$. Let C' be the sampled clients.
2. Compute a ρ_{st}-approximate Steiner tree T on terminals $C' \cup \{r\}$.
3. Define a FL instance with clients C, facilities F and opening costs $o'(f) := o(f) + M \cdot c(f, C' \cup \{r\})$. Compute a (λ_F, λ_C)-approximate solution $F' \subseteq F$ to this instance.
4. Augment T with shortest paths from each $f \in F'$ to T. Let T' be the augmented tree.
5. Return (F', T')

1-MROB. (A generalization of this problem, where subsets of pairs are grouped together, is discussed in [19]). The authors show that any β approximation for 1-MROB gives a 2β approximation for MROB, and present a constant approximation for the first problem. Indeed, essentially the same reduction works also in the opposite direction (giving a 10 approximation for 1-MROB based on the result in [12]). We can define the 1-MCFL problem analogously: this problem models natural scenarios (e.g., a person might want to use at most one public transportation network to commute). Using the same reduction as in [24], we obtain a $2 \cdot 16.2 = 32.4$ approximation for 1-MCFL: details are postponed to the full version of the paper.

2   An Improved Approximation for CFL

Let us consider Algorithm 1 in the figure. Here $\alpha \in (0, 1]$ is a constant to be fixed later. A *bifactor* (λ_F, λ_C)-approximation for FL is an algorithm which produces a solution to a FL instance of cost at most $\lambda_F \cdot O + \lambda_C \cdot C$, where O and C are the opening and connection cost of *any given* feasible solution. We will exploit the following result by Byrka [4].

Lemma 1. [4] *For any $\lambda_F > 1.67$, there is a $(\lambda_F, 1 + 2e^{-\lambda_F})$-approximation algorithm for FL.*

We remark that we would obtain an improved approximation also using a standard facility location algorithm. However, the approximation ratio would be slightly higher. A second tool that we need is the following Lemma which is implicitly proved in [9].

Lemma 2. [9] *Given an undirected graph $G = (V, E)$, with edge costs $c : E \to \mathbb{Q}^+$, a set of clients $C \subseteq V$, a subtree T' (core) containing a root node r, a mapping $\sigma : C \to V(T')$, and a probability $p \in (0, 1]$. Mark each client independently with probability p, and denote marked clients by C'. Let $\sigma(C') := \cup_{v \in C'} \sigma(v)$. Then $E[\sum_{v \in C} c(v, \sigma(C') \cup \{r\})] \leq \frac{0.807}{p} c(T') + \sum_{v \in C} c(v, \sigma(v))$.*

We let $OPT = (F^*, T^*)$ denote the optimal solution to the considered CFL instance, where F^* is the set of facilities and T^* the Steiner tree connecting

them. We also let $\sigma^*(v) \in F^*$ be the facility serving $v \in C$ in OPT. By $OPT = O^* + C^* + S^*$ we denote the optimal cost, where $O^* := o(F^*)$ is the opening cost, $C^* := \sum_{v \in C} c(v, F^*)$ the connection cost, and $S^* := M \cdot c(T^*)$ the Steiner cost. Let $O_{fl} := \sum_{f \in F'} o'(f) = \sum_{f \in F'} (o(f) + M \cdot c(f, C' \cup \{r\}))$ be the opening cost of the facility location solution computed in Step 3, and $C_{fl} := \sum_{v \in C} c(v, F')$ be the connection cost in the same solution. We need a few, simple intermediate results.

Lemma 3. *The cost of the returned solution is at most $M \cdot c(T) + O_{fl} + C_{fl}$.*

Proof. The connection cost in the FL and CFL solutions are the same. Recall that $o'(f) = o(f) + M \cdot c(f, C' \cup \{r\})$. Thus the modified opening costs o' pay fully for both opening F' and for augmenting T to T'.

Lemma 4. *One has $E[M \cdot c(T)] \leq \rho_{st} \cdot (S^* + \alpha \cdot C^*)$.*

Proof. A feasible Steiner tree on $C' \cup \{r\}$, of expected cost $c(T^*) + \frac{\alpha}{M} C^*$, is obtained by augmenting T^* with the shortest paths between each $v \in C'$ and $\sigma^*(v)$. Multiplying by $\rho_{st} \cdot M$ then gives the claim.

Lemma 5. *One has $E[O_{fl} + C_{fl}] \leq \lambda_F(O^* + \alpha C^*) + \lambda_C(C^* + \frac{0.807}{\alpha} S^*)$.*

Proof. We provide a FL solution, whose expected opening cost is $O^* + \alpha C^*$ and whose expected connection cost is $C^* + \frac{0.807}{\alpha} S^*$. Choose facilities $\sigma^*(C') \cup \{r\}$, with $\sigma^*(C') := \cup_{v \in C'} \sigma^*(v)$. Then the expected opening cost is

$$E\Big[\sum_{f \in \sigma^*(C') \cup \{r\}} o'(f) \Big] \leq E\Big[\sum_{f \in F^*} o(f) \Big] + M \cdot E\Big[\sum_{v \in C'} c(v, \sigma^*(v)) \Big]$$

$$= O^* + M \cdot \frac{\alpha}{M} \cdot \sum_{v \in C} c(v, \sigma^*(v)) = O^* + \alpha C^*.$$

The crucial argument here is that we need to account for the extra term $M \cdot c(v, \sigma^*(v))$ only if $v \in C'$, which happens with probability $\frac{\alpha}{M}$.

In order to bound the expected connection cost, we apply Lemma 2, with clients C, core T^*, mapping $\sigma = \sigma^*$, root r, and probability α/M:

$$E\Big[\sum_{v \in C} c(v, \sigma^*(C') \cup \{r\}) \Big] \leq \frac{0.807}{\alpha/M} c(T^*) + \sum_{v \in C} c(v, \sigma^*(v)) = \frac{0.807}{\alpha} S^* + C^*.$$

Theorem 2. *Algorithm 1 is an expected 3.19-approximation algorithm for CFL.*

Proof. Recall that $\rho_{st} \leq \ln(4) + \varepsilon$ for every fixed $\varepsilon > 0$ [5]. From Lemmas 1, 3, 4, and 5, the total expected cost of the approximate solution is upper bounded by

$$\rho_{st} \cdot (S^* + \alpha \cdot C^*) + \lambda_F(O^* + \alpha C^*) + \lambda_C\left(C^* + \frac{0.807}{\alpha} S^*\right)$$

$$= O^* \lambda_F + S^* \left(\rho_{st} + \lambda_C \frac{0.807}{\alpha}\right) + C^*(\rho_{st}\alpha + \lambda_F\alpha + \lambda_C)$$

$$\underset{\leq}{\overset{\substack{\alpha=0.539 \\ \lambda_F=2.294>1.67 \\ \lambda_C \leq 1+2e^{-\lambda_F}}}{}} 2.30 \cdot O^* + 3.19 \cdot S^* + 3.19 \cdot C^* \leq 3.19 \cdot OPT.$$

Algorithm 2 Approximation algorithm for MCFL.

---

1. Define a PFL instance on the input instance with clients $C = \{s_1, r_1, \ldots, s_k, r_k\}$ (as multiset), and $p(s) = p(r) = c(s,r)/2$ for any $(s,r) \in P$. Compute a ρ_{pfl}-approximate solution (F', C') for this problem. Let $\sigma(v) \in F'$ be the open facility which is closest to $v \in C - C'$, $P_{dir} := \{(s,r) \in P : \{s,r\} \cap C' \neq \emptyset\}$, and $P_{ind} = P - P_{dir}$.
2. Sample each pair $(\sigma(s), \sigma(r))$, with $(s,r) \in P_{ind}$, independently with probability $1/M$. Let P' be the sampled pairs of facilities.
3. Compute a 2-approximate Steiner forest T' over P' using the primal-dual algorithm in [1].
4. Output (P', T').

---

3   A Constant Approximation for MCFL

Let us consider Algorithm 2 in the figure. With a slight notational abuse, for a set of pairs $\tilde{P}$, we use $\tilde{P}$ also to denote the corresponding set of nodes. In the first step, we define and (approximately) solve a proper PFL instance, whose clients C are given by the nodes in the input pairs P. Currently, $\rho_{pfl} \leq 1.86$ [30]. Let C' be discarded clients, and $\sigma(v)$ be the facility serving $v \in C - C'$. Intuitively, the pairs P_{dir} with at least one endpoint in C' are connected directly via a shortest path. For the remaining pairs $(s,r) \in P_{ind}$, we consider the associated pairs of facilities $(\sigma(s), \sigma(r))$. The latter pairs define an MROB instance `mrob`. To this instance we essentially apply the MROB algorithm `rand` in [12]. In particular, we sample each pair independently with probability $1/M$, and compute a Steiner forest T' on the sampled pairs P' with the 2-approximation algorithm in [1]. The output solution is given by facilities P' and forest T'.

We need the following result in [12,25]. We recall that, for a set of nodes V' and a forest T', $c_{T'}$ defines distances after contracting the connected components of T', while $c_{V',T'}$ defines distances after contracting the nodes in V' belonging to same tree of T'. In particular, in general $c_{V',T'}(u,v) \geq c_{T'}(u,v)$.

Lemma 6. [12,25] *Consider an MROB instance on pairs $\tilde{P}$, with optimal cost OPT_{mrob}. Sample each pair in $\tilde{P}$ independently with probability $1/M$, and compute a 2-approximate Steiner forest T' on the sampled pairs P' with the algorithm in [1]. Then $E[M \cdot c(T') + \sum_{(s,r) \in \tilde{P}} c_{T'}(s,r)] \leq 5 \cdot OPT_{mrob}$. The same claim holds by replacing $c_{T'}$ with $c_{P',T'}$.*

The last claim of Lemma 6 is not relevant for MROB (it just comes out as a byproduct of the analysis in [12]). However, it is crucial for our analysis. In particular, since P' is a subset of facilities in our case, the connection path for $(\sigma(s), \sigma(r))$ in the MROB solution as given by Lemma 6 enters and leaves trees in T' at facilities. Henceforth, we can extend such connection path with shortest paths $(s, \sigma(s))$ and $(\sigma(r), r)$ to obtain a feasible connection path for pair (s,r).

With a notation analogous to Section 2, we let $OPT = (F^*, T^*)$ denote the optimum solution, where F^* is the set of open facilities and T^* a Steiner forest. We also let $OPT = O^* + C^* + S^*$ be the optimal cost, where O^*, C^*, and S^*

are the opening, connection, and Steiner cost, respectively. By O_{pfl}, P_{pfl}, and C_{pfl} we denote, respectively, the opening, penalty, and connection cost of the PFL solution computed in Step 1. We also let $S_{mrob} = M \cdot c(T')$ and $C_{mrob} = \sum_{(s,r) \in P_{ind}} c_{P',T'}(\sigma(s), \sigma(r))$ be the Steiner and connection cost, respectively, of the MROB solution computed in Step 3, as suggested by Lemma 6. Eventually, $APX = O_{apx} + S_{apx} + C_{apx}$ is the cost of the approximate solution, where O_{apx}, S_{apx}, and C_{apx} are the opening, Steiner, and connection cost, respectively.

Lemma 7. $APX \leq O_{pfl} + S_{mrob} + 2P_{pfl} + C_{pfl} + C_{mrob}$.

Proof. By definition, $APX = O_{apx} + S_{apx} + C_{apx}$. Trivially, $O_{apx} \leq O_{pfl}$ (we open a subset P' of the facilities F' in the PFL solution). Moreover, $S_{apx} = S_{mrob}$ by construction.

In order to prove the claim, it is then sufficient to describe connection paths of total cost $2P_{pfl} + C_{pfl} + C_{mrob}$. Let us connect all the pairs in P_{dir} directly via a shortest path. Since by definition at least one endpoint of each pair in P_{dir} belongs to the discarded clients C', $\sum_{(s,r) \in P_{dir}} c(s,r) \leq 2 P_{pfl}$. Consider now the remaining pairs P_{ind}. For each $(s,r) \in P_{ind}$, we connect s to $\sigma(s)$ and $\sigma(r)$ to r via a shortest path. Then we connect $\sigma(s)$ to $\sigma(r)$ using a shortest path with respect to $c_{P',T'}$. Observe that this is a feasible connection path for (s,r). The total cost of these paths is $\sum_{(s,r) \in P_{ind}} (c(s, \sigma(s)) + c_{P',T'}(\sigma(s), \sigma(r)) + c(\sigma(r), r)) = C_{mrob} + \sum_{v \in P_{ind}} c(v, \sigma(v)) = C_{mrob} + C_{pfl}$. The claim follows.

Lemma 8. $O_{pfl} + C_{pfl} + P_{pfl} \leq \rho_{pfl}(O^* + C^*)$.

Proof. It is sufficient to show that there exists a PFL solution of cost at most $O^* + C^*$. Let $P_{dir}^* \subseteq P$ be the pairs whose connection path in $OPT = (F^*, T^*)$ does not use any edge of T^*, and $P_{ind}^* = P - P_{dir}^*$. By C_{dir}^* (resp., C_{ind}^*) we denote the connection cost of OPT restricted to P_{dir}^* (resp., P_{ind}^*).

Consider the PFL solution (F^*, P_{dir}^*). This solution has opening cost O^* and penalty cost $\sum_{(s,r) \in P_{dir}^*} (p(s) + p(r)) = \sum_{(s,r) \in P_{dir}^*} 2\frac{c(s,r)}{2} = C_{dir}^*$. Moreover, its connection cost is $\sum_{(s,r) \in P_{ind}^*} (c(s, F^*) + c(r, F^*)) \leq \sum_{(s,r) \in P_{ind}^*} c_{F^*,T^*}(s,r) = C_{ind}^*$. Altogether, the cost of this PFL solution is upper bounded by $O^* + C_{dir}^* + C_{ind}^* = O^* + C^*$.

Lemma 9. $E[S_{mrob} + C_{mrob}] \leq 5(C_{pfl} + C^* + S^*)$.

Proof. Let $\tilde{P} := \{(\sigma(s), \sigma(r)) : (s,r) \in P_{ind}\}$ (considered as a multiset). The triple $(\tilde{P}, P', T')$ satisfies the conditions of Lemma 6. Hence, $E[S_{mrob} + C_{mrob}] = E[M \cdot c(T') + \sum_{(s',r') \in \tilde{P}} c_{P',T'}(s',r')] \leq 5 OPT_{mrob}$, where OPT_{mrob} is the optimum solution to the MROB instance mrob induced by pairs $\tilde{P}$.

A bound on OPT_{mrob} is given by the following feasible solution to mrob. Buy the edges of the optimal forest T^* of OPT. This costs S^*. For each pair $(\sigma(s), \sigma(r)) \in \tilde{P}$, connect $\sigma(s)$ to s and $\sigma(r)$ to r via a shortest path, and then connect s to r via the connection path between s and r in OPT. The cost of this solution is $\sum_{(s,r) \in P_{ind}} (c(s, \sigma(s)) + c(r, \sigma(r)) + c_{F^*,T^*}(s,r)) \leq C_{pfl} + C^*$. Then $OPT_{mrob} \leq C_{pfl} + C^* + S^*$. The claim follows.

Theorem 3. *Algorithm 2 is an expected* 16.2*-approximation algorithm for MCFL.*

Proof. One has

$$
\begin{aligned}
E[APX] \;&\overset{\text{Lem 7}}{\leq}\; O_{pfl} + 2P_{pfl} + C_{pfl} + E[S_{mrob} + C_{mrob}] \\
&\overset{\text{Lem 9}}{\leq}\; O_{pfl} + 2P_{pfl} + C_{pfl} + 5\left(C_{pfl} + C^* + S^*\right) \\
&\leq 6\left(O_{pfl} + P_{pfl} + C_{pfl}\right) + 5\left(C^* + S^*\right) \\
&\overset{\text{Lem 8}}{\leq}\; 6\,\rho_{pfl}(O^* + C^*) + 5(C^* + S^*) \\
&\leq (6\,\rho_{pfl} + 5)(O^* + C^* + S^*) \leq 16.2\,OPT.
\end{aligned}
$$

4   On the Approximability of SROB

Recall that SROB is the special case of CFL where every node is a facility with opening cost zero. Without loss of generality, we can assume that we are also given a root node $r \in V$ which belongs to the tree T^* in the optimum solution. In this section we show that SROB cannot be approximated within a factor of 1.278, unless $\mathbf{NP} \subseteq \mathbf{DTIME}(n^{O(\log \log n)})$. This heavily improves over the previously known approximation hardness of 1.01 (due to hardness of Steiner tree [6]).

As an intermediate step, we consider a reduction to the *uniform facility location* problem (UnifFL), i.e. the special case of metric facility location where all facilities have uniform opening cost o. Indeed, we consider an even more restrictive case. For a set N of non-negative numbers, let N-UnifFL denote the special case of UnifFL where, for any client $v \in C$ and facility $f \in F$, $c(v, f) \in N$. Given a solution $F' \subseteq F$, we let $\sigma(v)$ denote the facility in F' which is closest to client $v \in C$. We also say that v is *assigned* to $\sigma(v)$ and that $\sigma(v)$ *serves* v.

Guha and Khuller [16] showed that, unless $\mathbf{NP} \subseteq \mathbf{DTIME}(n^{O(\log \log n)})$, $\{1, 3\}$-UnifFL cannot be approximated within a factor of 1.463. While this case seems hard to reduce to SROB, we are able to prove a similar reduction for $\{1, 2\}$-UnifFL.

Lemma 10. *Given an α-approximation algorithm for SROB, there is an α-approximation algorithm for $\{1, 2\}$-UnifFL.*

Proof. Consider a given $\{1, 2\}$-UnifFL instance on clients C and facilities F. First suppose that the uniform opening cost is $o \geq 1$. We define an SROB instance as follows. Consider the complete graph G on nodes $C \cup F \cup \{r\}$, with clients C, root r, and $M = o \geq 1$. Edges (r, f) and (v, f), with $f \in F$, $v \in C$, and $c(v, f) = 1$, have unit costs. All other edges have cost 2. Let the *degree* $d(f) = |\{v \in C : c(v, f) = 1\}|$ of a facility $f \in F$ be the number of clients at distance 1 from that facility. We remark that, if $d(f) \leq M$ for all $f \in F$, then there is an optimum solution of $\{1, 2\}$-UnifFL where only one facility f^* is opened. In fact, suppose $f \neq f^*$ is opened as well, where f serves $x' \leq M$ clients at distance 1 and x'' clients at distance 2. By closing f and assigning its clients to f^*, one

saves at least $(o + x' + 2x'') - (2x' + 2x'') = M - x' \geq 0$. The best solution with one open facility can be computed in polynomial time. So we can assume without loss of generality that there is at least one facility f with $d(f) > M$.

Observe that any solution F' to an $\{1,2\}$-UnifFL instance induces an SROB solution of the same cost. In fact, it is sufficient to consider the tree T' induced by edges $\{r, f\}$, $f \in F'$: this solution costs $|F'| \cdot o + \sum_{v \in C} c(v, F')$. Hence, the cost of an α-approximate solution to SROB costs at most α times the cost of the optimum solution to $\{1,2\}$-UnifFL.

Thus, it is sufficient to show that any feasible solution to SROB can be turned in polynomial time into a solution to $\{1,2\}$-UnifFL of not larger cost. Consider any such solution T' to SROB. Suppose that $T' = \emptyset$. Then, adding edge $\{r, f\}$, with f being the maximum degree facility (recall that $d(f) > M$), can only decrease the cost. Next assume $T' \neq \emptyset$. Suppose there is any client v connected either to r or to another client in C. This connection costs 2, thus reconnecting v to any node in $F \cap V(T')$ can only make the solution cheaper. Suppose now that T' contains one edge $\{v, f\}$ with $v \in C$ and $f \in F$, but not edge $\{r, f\}$. Then replacing $\{v, f\}$ by $\{r, f\}$ leaves T' connected and can only reduce the cost. Finally we may still have edges $\{v, f\}$ and $\{r, f\}$, with $v \in C$ and $f \in F$. Then deleting $\{v, f\}$ again can only decrease the cost, since $M \geq 1$ and no other client (but v) is connected to v.

At the end of the process, T' only contains edges of type $\{r, f\}$, $f \in F$. Let $F' := \{f \in F : \{r, f\} \in T'\}$. F' induces a feasible solution to $\{1,2\}$-UnifFL of cost equal to the cost of the (modified) SROB solution. The claim follows.

A similar proof holds for the case $o < 1$, by letting $M = 1$ and setting edge costs $\{r, f\}$ to o.

We need the following result by Guha and Khuller [16].

Lemma 11. [16] *Suppose we have a set cover instance* $(\{1, \ldots, n\}, \{S_1, \ldots, S_m\})$ *with unit cost for sets and optimal cost* $OPT_{sc} = k$. *If there is a polynomial time algorithm that can pick* βk *sets (for any constant* $\beta > 0$) *and cover* $c' \cdot n$ *elements, where* $c' > c_\beta = 1 - e^{-\beta}$, *then* **NP** $\subseteq$ **DTIME**$(n^{O(\log \log n)})$.

To see why Lemma 11 holds, suppose for the sake of contradiction that such an algorithm does exist. Then we can apply it iteratively to a set cover instance. Let $\alpha_t n$ be the number of covered elements at iteration t, $t = 1, \ldots, T$ (in particular, $\sum_{t=1}^{T} \alpha_t = 1$). At iteration t the algorithm uses only $\beta_t k \leq \delta \cdot \ln(\frac{1}{1-\alpha_t})$ many sets where $\delta < 1$ is a constant. Then we obtain a solution with $\sum_{t=1}^{T} \beta_t k \leq \sum_{t=1}^{T} \delta \ln(\frac{1}{1-\alpha_t})k \leq \delta \cdot \ln(n)k$ sets, contradicting the hardness result of [10].

Lemma 12. *There is no* α-*approximation algorithm with* $\alpha \leq 1.278$ *for* $\{1,2\}$-*UnifFL, unless* **NP** $\subseteq$ **DTIME**$(n^{O(\log \log n)})$.

Proof. Suppose for the sake of contradiction that we have an α-approximation algorithm for $\{1,2\}$-UnifFL and $\alpha \leq 1.278$. Consider a set cover instance with n elements, sets $S_1, \ldots, S_m$ of unit cost, and optimal value $OPT_{sc} = k$. Define a $\{1,2\}$-UnifFL instance as follows. Introduce a facility f for each set S_f and a

client v for each element v. Set $c(v, f) = 1$ if $v \in S_f$, and $c(v, f) = 2$ otherwise. Let $o = \gamma \frac{n}{k}$ be the uniform cost. Here $0 < \gamma < 1$ is a constant, that we will determine later. Let OPT_{ufl} be the optimum solution to this instance. By opening the k facilities which correspond to the k sets in the optimum set cover solution, we obtain that $OPT_{ufl} \leq k \cdot o + 1 \cdot n = k \cdot \gamma \frac{n}{k} + n = (1 + \gamma)n$.

Using the α-approximation algorithm we get a $\{1, 2\}$-UnifFL solution F' of cost at most $\alpha(1 + \gamma)n$. Let $c \in [0, 1]$ be the fraction of demands, whose service costs are 1 (the others are served at cost 2). Define $\beta := |F'|/k > 0$. Then

$$(\beta\gamma + 2 - c)n = \underbrace{\beta k}_{=|F'|} \cdot o + \underbrace{cn + 2 \cdot (1 - c)n}_{\text{connection cost}} \leq \alpha(1 + \gamma)n.$$

This can be rearranged to

$$c \geq \beta \cdot \gamma + 2 - \alpha(1 + \gamma) \overset{\alpha \leq 1.278, \gamma := 0.278}{\geq} 0.278\beta + 0.3667 > 1 - e^{-\beta},$$

contradicting Lemma 11.

Theorem 4. *There cannot be a factor* 1.278-*approximation algorithm for SROB, MROB, CFL, MCFL, SSBB, and VPN, unless* $\mathbf{NP} \subseteq \mathbf{DTIME}(n^{O(\log \log n)})$.

Proof. The claim for SROB follows from Lemmas 10 and 12. The same result trivially extends to MROB, CFL, MCFL, and SSBB (which are all generalizations of SROB). The hardness for VPN follows from the fact, that any SROB instance is equivalent to a VPN instance with a single receiver of capacity M (see, e.g., [15,27]).

References

1. A. Agrawal, P. Klein, and R. Ravi. When trees collide: an approximation algorithm for the generalized Steiner problem on networks. *SIAM Journal on Computing*, 24: 440–456, 1995.
2. B. Awerbuch and Y. Azar. Buy-at-bulk network design. In *FOCS*, 542–547, 1997.
3. L. Becchetti, J. Könemann, S. Leonardi, and M. Pál. Sharing the cost more efficiently: improved approximation for multicommodity rent-or-buy. In *SODA*, 375–384, 2005.
4. J. Byrka. An Optimal Bifactor Approximation Algorithm for the Metric Uncapacitated Facility Location Problem. In *APPROX*, 29–43, 2007.
5. J. Byrka, F. Grandoni, T. Rothvoß, and L. Sanità. An improved LP-based approximation for Steiner tree. In *STOC*, 583–592, 2010.
6. M. Chlebík and J. Chlebíková. The Steiner tree problem on graphs: Inapproximability results. *Theoretical Computer Science*, 406(3):207–214, 2008.
7. F. Eisenbrand and F. Grandoni. An improved approximation algorithm for virtual private network design. In *SODA*, 928–932, 2005.
8. F. Eisenbrand, F. Grandoni, G. Oriolo, and M. Skutella. New approaches for virtual private network design. *SIAM Journal on Computing*, 37(3): 706–721, 2007.

9. F. Eisenbrand, F. Grandoni, T. Rothvoß, and G. Schäfer. Connected facility location via random facility sampling and core detouring. *Journal of Computer and System Sciences*, 76: 709-726, 2010.
10. U. Feige. A Threshold of ln n for Approximating Set Cover. *Journal of the ACM*, 45(4): 1998.
11. J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *Journal of Computer and System Sciences*, 69(3): 485-497, 2004.
12. L. Fleischer, J. Könemann, S. Leonardi, and G. Schäfer. Simple cost sharing schemes for multicommodity rent-or-buy and stochastic steiner tree. In *STOC*, 663–670, 2006.
13. N. Garg, R. Khandekar, G. Konjevod, R. Ravi, F. Salman, and A. Sinha. On the integrality gap of a natural formulation of the single-sink buy-at-bulk network design problem. In *IPCO*, 170–184, 2001.
14. F. Grandoni and G. F. Italiano. Improved approximation for single-sink buy-at-bulk. In *ISAAC*, 111–120, 2006.
15. F. Grandoni and T. Rothvoß. Network design via core detouring for problems without a core. In *ICALP*, 490–502, 2010.
16. S. Guha and S. Khuller. Greedy Strikes Back: Improved Facility Location Algorithms. *Journal of Algorithms*, 31(1): 228–248, 1999.
17. S. Guha, A. Meyerson, and K. Munagala. A constant factor approximation for the single sink edge installation problem. *SIAM Journal on Computing*, 38(6): 2426-2442, 2009.
18. A. Gupta, J. Kleinberg, A. Kumar, R. Rastogi, and B. Yener. Provisioning a virtual private network: a network design problem for multicommodity flow. In *STOC*, 389–398, 2001.
19. A. Gupta and A. Kumar. A constant-factor approximation for stochastic Steiner forest. In *STOC*, 659–668, 2009.
20. A. Gupta, A. Kumar, M. Pal, and T. Roughgarden. Approximation via cost-sharing: simpler and better approximation algorithms for network design. *Journal of the ACM*, 54(3): 11, 2007.
21. A. Gupta, A. Srinivasan, and E. Tardos. Cost-sharing mechanisms for network design. In *APPROX*, 139–150, 2004.
22. R. Jothi and B. Raghavachari. Improved approximation algorithms for the single-sink buy-at-bulk network design problems. In *SWAT*, 336–348, 2004.
23. D. R. Karger and M. Minkoff. Building Steiner trees with incomplete global knowledge. In *FOCS*, 613–623, 2000.
24. A. Kumar, A. Gupta, and T. Roughgarden. A constant-factor approximation algorithm for the multicommodity rent-or-buy problem. In *FOCS*, 333–342, 2002.
25. S. Leonardi. Private communication. 2008.
26. A. Meyerson, K. Munagala, and S. Plotkin. Cost-distance: two metric network design. In *FOCS*, 624–630, 2000.
27. T. Rothvoß, and L. Sanità. On the Complexity of the Asymmetric VPN Problem. In *APPROX*, 326–338, 2009.
28. C. Swamy and A. Kumar. Primal–dual algorithms for connected facility location problems. *Algorithmica*, 40(4):245–269, 2004.
29. K. Talwar. The single-sink buy-at-bulk LP has constant integrality gap. In *IPCO*, 475–486, 2002.
30. G. Xu and J. Xu. An improved approximation algorithm for uncapacitated facility location problem with penalties. *Journal of Combinatorial Optimization*, 17(4):424–436, 2009.