Multi-Commodity
Connected Facility Location

Fabrizio Grandoni*

*Dipartimento di Informatica, Sistemi e Produzione, Università di Roma Tor Vergata
via del Politecnico 1, 00133, Roma, Italy. Email: grandoni@disp.uniroma2.it

Abstract

In this paper we define the multi-commodity connected facility location (MCFL) problem, a natural generalization of the well-known multi-commodity rent-or-buy (MROB) and connected facility location (CFL) network design problems. Like in MROB, we wish to send one unit of flow from a set of sources to a set of sinks. This flow is supported by renting or buying edges. For each bought edge e, we pay $M \geq 1$ times the cost of e, and we are then free to route an unbounded amount of flow along e. For each rented edge e, we pay its cost times the amount of flow routed along e. However, differently from MROB and similarly to CFL, the rented paths cannot reach and leave the connected components of bought edges at any node. This must happen only at specific nodes, the open facilities, for which we have to pay an opening cost. MCFL models, e.g., a transportation system consisting of several disconnected networks (the bought edges). Users can switch from one network to the other on foot or by bike (rented edges), entering and leaving each network at stations (facilities).

We present the first constant approximation algorithm for MCFL. Our algorithm is based on two main ingredients. We use an approximation algorithm for the price-collecting facility location problem to identify pairs that can be cheaply connected with rented edges only. The remaining pairs and the facilities serving them are used to define an MROB instance, on which we run a proper approximation algorithm.

Keywords: *approximation algorithms, network design, connected facility location, multi-commodity rent-or-buy.*

1 Introduction

In this paper we define a new network design problem, *multi-commodity connected facility location* (MCFL), which generalizes in a natural way two well-known NP-hard problems: *multi-commodity rent-or-buy* (MROB) and *connected facility location* (CFL). We present the first constant approximation algorithm for MCFL.

Informally, in the MROB problem we are given a weighted graph and a set of source-sink pairs. We wish to send one unit of flow from each source to the corresponding sink. In order to support this flow, we can either rent or buy edges. Buying one edge e costs M times the cost $c(e)$ of e, but then we are allowed to route an unbounded amount of flow along e. Renting one edge e costs $c(e)$ times the number of connection paths using it. More formally, we are given an undirected graph $G = (V, E)$, with edge costs $c : E \to \mathbb{R}^+$, a set of source-sink pairs $\mathcal{P} = \{(s_1, r_1), \ldots, (s_k, r_k)\}$, $(s_i, r_i) \in V \times V$, and a parameter $M \geq 1$. With a slight notational abuse, given a set $\mathcal{P}$ of pairs, we will use $\mathcal{P}$ also to denote the corresponding set of elements $\{x : x \in (s, r) \in \mathcal{P}\}$. The meaning will be clear from the context. The aim is finding a forest $\mathcal{S}' \subseteq E$ of *bought* edges which minimizes the sum of the *Steiner cost* $M \cdot c(\mathcal{S}') = M \cdot \sum_{e \in \mathcal{S}'} c(e)$ and *connection cost* $\sum_{(s,r) \in \mathcal{P}} \ell_{\mathcal{S}'}(s, r)$. Here $\ell_{\mathcal{S}'}$ denotes the distance function obtained by contracting the components of $\mathcal{S}'$. (We use ℓ to denote the distance function with respect to edge costs $c(e)$ in the original graph).

We can think of bought edges as a transportation system, consisting of several disconnected networks. People each day use one or more networks, switching from one network to the other on foot or by bike. Switching has a given social cost per kilometer, as well as running the transportation system. (The parameter M models the relation between the two costs). However, the switching cost is paid on a per-person base, while the running cost is independent from the number of persons using each network.

MROB fails to model the fact that transportation networks are typically entered and left only at given points along the network. For example, reaching a railroad is not sufficient to get a train: we have to arrive at the closest train station. Moreover, the running cost of train stations is hardly captured by the extension of the railroad (which is considered in the Steiner cost).

In view of these limitations, we defined the following generalization of MROB. Let us introduce a set $\mathcal{F} \subseteq V$ of *facilities* (the train stations in the scenario above), with opening costs $f : \mathcal{F} \to \mathbb{R}^+$. A feasible solution to the *multi-commodity connected facility location* problem (MCFL) is given by a subset $\mathcal{F}' \subseteq \mathcal{F}$ of *open* facilities, and a forest $\mathcal{S}' \subseteq E$ of *bought* edges. The cost of $(\mathcal{F}', \mathcal{S}')$, that we wish to minimize, is the sum of the *opening cost* $\sum_{v \in \mathcal{F}'} f(v)$, *Steiner cost* $M \cdot c(\mathcal{S}') = M \cdot \sum_{e \in \mathcal{S}'} c(e)$, and *connection cost* $\sum_{(s,r) \in \mathcal{P}} \ell_{\mathcal{F}',\mathcal{S}'}(s, r)$. Here $\ell_{\mathcal{F}',\mathcal{S}'}$ denotes the distance function obtained by contracting the open facilities which belong to the *same connected component* of $\mathcal{S}'$.

Hence MROB is the special case of MCFL where $\mathcal{F} = V$ (i.e., we can open a facility at any node) and the opening cost of the facilities is zero. A solution to MROB can be specified by giving the forest $\mathcal{S}'$ only, since we can implicitly assume that all the nodes spanned by $\mathcal{S}'$ are open facilities.

MCFL can be also considered as a generalization of *connected facility location* (CFL), which is the special case where all the pairs share a common sink r, which is a facility of opening cost zero. (Sometimes a slightly different, but equivalent, definition of CFL is given in the literature). Under this assumption, the optimum solution consists of a set of facilities $\mathcal{F}' \ni r$, and a Steiner tree $\mathcal{S}'$ spanning $\mathcal{F}'$.

Our results and techniques. We present a 17-approximation algorithm for $MCFL$. Our approximation algorithm is based on the following two main ingredients.

The first ingredient is a reduction to the price-collecting facility location (PFL) problem. Informally, PFL is a variant of facility location, where we are not forced to connect all the clients to some facility, but we have to pay a penalty for each unconnected client. More precisely, given G, c, $\mathcal{F}$, and f like in MCFL, and a set of clients $\mathcal{D} \subseteq V$, with *penalties* $p : \mathcal{D} \to \mathbb{R}^+$. We wish to find a subset $\mathcal{F}' \subseteq \mathcal{F}$ of open facilities and a subset $\mathcal{D}' \subseteq \mathcal{D}$ of *discarded* clients which minimize the sum of the *penalty cost* $\sum_{v \in \mathcal{D}'} p(v)$, opening cost $\sum_{v \in \mathcal{F}'} f(v)$, and connection cost $\sum_{v \in \mathcal{D} \setminus \mathcal{D}'} \ell(v, \mathcal{F}')$. Here $\ell(v, \mathcal{F}')$ denotes the distance between v and the closest facility $\sigma(v) \in \mathcal{F}'$.

We define a PFL instance on the top of the input MCFL instance by letting the endpoints of $\mathcal{P}$ be the clients $\mathcal{D}$, and setting the penalty of each client $v \in (s, r) \in \mathcal{P}$ to $\ell(s, r)/2$. We compute a solution $(\mathcal{F}', \mathcal{D}')$ to this PFL instance using a ρ_{pfl} approximation algorithm. Currently, $\rho_{pfl} \leq 2$ [10]. Intuitively, we can afford to connect the pairs $\mathcal{P}_{dis}$ with at least one endpoint among the discarded clients $\mathcal{D}'$ directly via a shortest path (without using bought edges). This costs at most twice the penalty cost P_{pfl} of the PFL solution computed.

At this point the second ingredient comes into play. Fleischer, Könemann, Leonardi, and Schäfer [4] (see also [7]) gave the following MROB algorithm `randMROB`. Sample each pair independently with probability $1/M$, and compute a Steiner forest $\mathcal{S}'$ over the sampled pairs $\mathcal{P}'$ via the 2-approximate primal-dual Steiner forest algorithm by Agrawal, Klein, and Ravi [1]. Let $\mathcal{P}_{sel} = \mathcal{P} \setminus \mathcal{P}_{dis}$. Recall that $\sigma(v)$ is the facility serving v in the PFL solution. The pairs $(\sigma(s), \sigma(r))$, $(s, r) \in \mathcal{P}_{sel}$, define an MROB instance. On this instance we run `randMROB`. Then we open the random set of facilities $\mathcal{P}'$, and buy the edges of the forest $\mathcal{S}'$.

It is easy to see that the opening cost of the final solution is at most the opening cost O_{pfl} of the PFL solution. In order to bound the connection cost of the pairs in $\mathcal{P}_{sel}$ and the Steiner cost, we use the following theorem in [4].

Theorem 1 **[4]** *Algorithm* `randMROB` *computes a solution of expected Steiner cost* $E[c(\mathcal{S}')] \leq 2\,opt_{mrob}$ *and expected connection cost* $E[\sum_{(s,r)\in\mathcal{P}} \ell_{\mathcal{S}'}(s,r)] \leq 3\,opt_{mrob}$*, where* opt_{mrob} *is the cost of the optimum solution. The same claim holds if* $\ell_{\mathcal{S}'}$ *is replaced by* $\ell_{\mathcal{P}',\mathcal{S}'}$*.*

The last claim of Theorem 1 is not relevant for MROB (it just comes out as a byproduct of the analysis in [4]). However, it is crucial for our analysis. Intuitively, there is a cheap, with respect to $\ell_{\mathcal{P}',\mathcal{S}'}$, connection path between each pair $(\sigma(s), \sigma(t))$, $(s,r) \in \mathcal{P}_{sel}$. By extending this path with the connection paths between the pairs $(s, \sigma(s))$ and $(r, \sigma(r))$ in the PFL solution we obtain a convenient connection path between s and r (see also Figure 1).

Previous and Related Work. A lot of work has been devoted to MROB [2, 4, 5, 6, 7, 8, 9, 11]. The current best 5-approximation for the problem is due to Fleischer, Könemann, Leonardi, and Schäfer [4], who exploit the basic random-sampling approach by Gupta, Kumar, Pal, and Roughgarden [7]. As mentioned before, their algorithm and analysis is crucial for our purposes.

Gupta, Kleinberg, Kumar, Rastogi, and Yener [6] obtain a 10.66-approximation algorithm for CFL, based on rounding an exponential size LP. Swamy and Kumar [11] later improved this result to 8.55, using a primal-dual algorithm. The current best 4-approximation for the problem is due to Eisenbrand, Grandoni, Rothvoß, and Schäfer [3]. Their result is based on the computation of an approximate (unconnected) facility location solution. The open facilities are then sampled in a way similar to the approach used here to sample the facility pairs.

The *single-sink rent-or-buy* problem (SROB) is a well-studied special case of MROB, where all the pairs share a common sink r. This can be interpreted also as a special case of CFL, where $\mathcal{F} = V$ and the opening costs are zero. Following a series of results [6, 9, 11], the current best 2.92-approximation for SROB is due to Eisenbrand et al. [3], who exploit the random-sampling framework by Gupta, Kumar, and Roughgarden [9].

2 Algorithm and Analysis

Recall that the PFL instance induced by the input MCFL instance has $\mathcal{P}$ and $\mathcal{F}$ as set of clients and facilities, respectively. Moreover, the penalty of $v \in (s,r) \in \mathcal{P}$ is $p(v) = \ell(s,r)/2$. Our algorithm `randMCFL` for MCFL works as follows.

(1) Compute a ρ_{pfl}*-approximate solution* $(\mathcal{F}', \mathcal{D}')$ *to the PFL instance induced by the input instance. Let* $\sigma(v) \in \mathcal{F}'$ *be the facility serving* $v \in \mathcal{P} \setminus \mathcal{D}'$*.*

(2) Sample each pair $(\sigma(s), \sigma(r))$*, with* $\{s,r\} \cap \mathcal{D}' = \emptyset$*, independently with probability* $\frac{1}{M}$*. Let* $\mathcal{P}'$ *denote the marked pairs of facilities. Compute a* 2*-approximate Steiner forest* $\mathcal{S}'$ *over* $\mathcal{P}'$ *using the* 2*-approximate primal-dual algorithm in [1].*

(3) Output $APX = (\mathcal{P}', \mathcal{S}')$*.*

We introduce some notation. We will denote by $OPT = (\mathcal{F}^*, \mathcal{S}^*)$ an optimum solution, and by *opt* its cost. We also let O^*, S^*, and C^* be the opening, Steiner, and connection cost of OPT, respectively. By O_{pfl}, C_{pfl}, and P_{pfl} we denote the opening, connection, and penalty cost, respectively, of the PFL solution computed at Step (1). Let $\mathcal{P}_{dis} \subseteq \mathcal{P}$ be the pairs with at least one (discarded) endpoint in $\mathcal{D}'$, and $\mathcal{P}_{sel} = \mathcal{P} \setminus \mathcal{P}_{dis}$. We also let $S_{mrob} = c(\mathcal{S}')$ and $C_{mrob} = \sum_{(s,r)\in\mathcal{P}_{sel}} \ell_{\mathcal{F}',\mathcal{S}'}(\sigma(s), \sigma(r))$.

We start with a simple bound on the cost *apx* of APX.

Lemma 1 $apx \leq O_{pfl} + S_{mrob} + 2P_{pfl} + C_{pfl} + C_{mrob}$.

Proof. Trivially the opening cost of APX is at most O_{pfl}, since the algorithm opens a subset of the facilities in $\mathcal{F}'$. Moreover, since we buy only the edges of $\mathcal{S}'$, the Steiner cost of APX is exactly $S_{mrob} = c(\mathcal{S}')$.

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 $\mathcal{P}_{dis}$ directly via a shortest path. Since by definition at

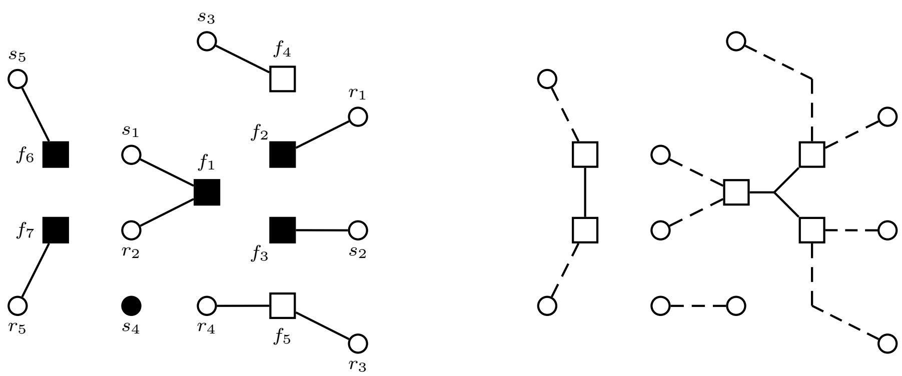

Figure 1: Example of execution of `randMCFL`. On the left, the PFL solution: black nodes indicate discarded clients $\mathcal{D}'$ and open facilities $\mathcal{P}'$. On the right, a solution with the claimed guarantee: full edges are bought. Observe that s_4 is discarded, hence the pair (s_4, r_4) is directly connected via a shortest path. Moreover, the pair (s_3, r_3), whose facility pair $(\sigma(s_3), \sigma(r_3)) = (f_4, f_5)$ is not sampled, is connected using the path $s_3, f_4, f_2, f_3, f_5, r_3$, which uses some bought edges between f_2 and f_3. In the example each connection path passes through at most one component of bought edges: this is not always the case.

least one endpoint of each pair in $\mathcal{P}_{dis}$ belongs to the discarded clients $\mathcal{D}'$, $\sum_{(s,r)\in\mathcal{P}_{dis}} \ell(s,r) \leq 2\, P_{pfl}$. Consider now the remaining pairs $\mathcal{P}_{sel}$. For each $(s,r) \in \mathcal{P}_{sel}$, 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 $\ell_{\mathcal{P}',\mathcal{S}'}$. Observe that this is a feasible connection path for (s,r). The total cost of these paths is

$$\sum_{(s,r)\in\mathcal{P}_{sel}} (\ell_{\mathcal{P}',\mathcal{S}'}(\sigma(s),\sigma(r)) + \ell(s,\sigma(s)) + \ell(r,\sigma(r))) = C_{mrob} + \sum_{v\in\mathcal{P}_{sel}} \ell(v,\sigma(v)) = C_{mrob} + C_{pfl}.$$

The claim follows. □

Lemma 2 $O_{pfl} + C_{pfl} + P_{pfl} \leq \rho_{pfl}(O^* + C^*)$.

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

Consider the PFL solution $(\mathcal{F}^*, \mathcal{P}^*_{dir})$. This solution has opening cost O^* and penalty cost $\sum_{(s,r)\in\mathcal{P}^*_{dir}}(p(s)+p(r)) = \sum_{(s,r)\in\mathcal{P}^*_{dir}} 2\frac{\ell(s,r)}{2} = C^*_{dir}$. Moreover, its connection cost is $\sum_{(s,r)\in\mathcal{P}^*_{dir}}(\ell(s,\mathcal{F}^*) + \ell(r,\mathcal{F}^*)) \leq \sum_{(s,r)\in\mathcal{P}^*_{dir}} \ell_{\mathcal{F}^*,\mathcal{S}^*}(s,r) = C^*_{ind}$. Altogether, the cost of this PFL solution is upper bounded by $O^* + C^*_{dir} + C^*_{ind} = O^* + C^*$. □

Lemma 3 $E[S_{mrob} + C_{mrob}] \leq 5(C_{pfl} + C^* + S^*)$.

Proof. The pairs $\{(\sigma(s),\sigma(r)) : (s,r) \in \mathcal{P}_{sel}\}$ define an MROB instance `mrob`. Let opt_{mrob} be the optimal cost of this instance. Each such pair is independently sampled with probability $1/M$. Moreover, on the sampled pairs we run the primal-dual 2-approximation algorithm in [1]. Hence the values of S_{mrob} and C_{mrob} satisfy the conditions of Theorem 1. In particular, $E[S_{mrob} + C_{mrob}] = E[c(\mathcal{S}') + \sum_{(s,r)\in\mathcal{P}_{sel}} \ell_{\mathcal{P}',\mathcal{S}'}(s,r)] \leq 5\, opt_{mrob}$.

It remains to bound opt_{mrob}. Consider the following solution to `mrob`. Buy the edges of the optimal forest $\mathcal{S}^*$ of OPT. This costs S^*. For each pair $(\sigma(s),\sigma(r))$, connect $\sigma(s)$ to s and r to $\sigma(r)$ via a shortest path, and then connect s to r via the connection path in OPT. The cost of this solution is $\sum_{(s,r)\in\mathcal{P}_{sel}}(\ell(s,\sigma(s)) + \ell(r,\sigma(r)) + \ell_{\mathcal{F}^*,\mathcal{S}^*}(s,r)) \leq C_{pfl} + C^*$. Then $opt_{mrob} \leq C_{pfl} + C^* + S^*$. The claim follows. □

Theorem 2 *Algorithm* `randMCFL` *is a* 17 *expected approximation algorithm for* $MCFL$*.*

Proof. Trivially, the algorithm computes a feasible solution in polynomial time. By Lemmas 1, 2, and 3:

$$\begin{aligned} E[apx] &\leq O_{pfl} + 2P_{pfl} + C_{pfl} + E[S_{mrob} + C_{mrob}] \leq O_{pfl} + 2P_{pfl} + C_{pfl} + 5\,(C_{pfl} + C^* + S^*) \\ &\leq 6\,(O_{pfl} + P_{pfl} + C_{pfl}) + 5\,(C^* + S^*) \leq 6\,\rho_{pfl}(O^* + C^*) + 5(C^* + S^*) \\ &\leq (6\,\rho_{pfl} + 5)(O^* + C^* + S^*) \leq 17\,opt. \end{aligned}$$

□

References

[1] A. Agrawal, P. Klein, and R. Ravi. When trees collide: an approximation algorithm for the generalized Steiner problem on networks. *SIAM J. Comput.*, 24:440–456, 1995.

[2] L. Becchetti, J. Könemann, S. Leonardi, and M. Pál. Sharing the cost more efficiently: improved approximation for multicommodity rent-or-buy. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 375–384, 2005.

[3] F. Eisenbrand, F. Grandoni, T. Rothvoß, and G. Schäfer. Approximating connected facility location problems via random facility sampling and core detouring. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1174–1183, 2008.

[4] L. Fleischer, J. Könemann, S. Leonardi, and G. Schäfer. Simple cost sharing schemes for multicommodity rent-or-buy and stochastic steiner tree. In *ACM Symposium on the Theory of Computing (STOC)*, pages 663–670, 2006.

[5] S. Guha, A. Meyerson, and K. Munagala. A constant factor approximation for the single sink edge installation problem. In *ACM Symposium on the Theory of Computing (STOC)*, pages 383–388, 2001.

[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 *ACM Symposium on the Theory of Computing (STOC)*, pages 389–398, 2001.

[7] A. Gupta, A. Kumar, M. Pal, and T. Roughgarden. Approximation via cost-sharing: a simple approximation algorithm for the multicommodity rent-or-buy problem. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 606–617, 2003.

[8] A. Gupta, A. Kumar, and T. Roughgarden. A constant-factor approximation algorithm for the multicommodity. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 333–344, 2002.

[9] A. Gupta, A. Kumar, and T. Roughgarden. Simpler and better approximation algorithms for network design. In *ACM Symposium on the Theory of Computing (STOC)*, pages 365–372, 2003.

[10] K. Jain, M. Mahdian, E. Markakis, A. Saberi, and V. Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *Journal of the Association for Computing Machinery*, 50:795–824, 2003.

[11] C. Swamy and A. Kumar. Primal–dual algorithms for connected facility location problems. *Algorithmica*, 40(4):245–269, 2004.