

Effective Algorithm CFL Based on Steiner Tree and UFL Problem

HONG-ZHEN ZHENG¹, DIAN-HUI CHU², DE-CHEN ZHAN³

^{1,2}*Department of Computer Science & Technology, Harbin Institute of Technology at Weihai, , China, 264209*

³*Department of Computer Science & Technology, Harbin Institute of Technology, Harbin, China, 1530001,*

Summary

In the field of approximation algorithms, a lot of earlier work on facility location problems and network design problems have sought to address these two questions independently. In this paper we present an integrated study of the overall Problem and study the problem in an integrated way that one to exploit the saving that may result from making both decision in a coordinated way to reduce the total cost of location and transportation. We provide approximation algorithms for some simple versions of the integrated problem to cover the gap.

We present a $\rho_{ST} + \rho_{UFL}$ approximation algorithm for the capacitated facility location problem (CFL in short), where ρ_P is any approximation factor achievable for the problem P. We do this by carefully combining solutions to appropriately set up Steiner tree and the uncapacitated facility location problems (UFL in short) that capture two natural lower bounds for our problem. With the current best approximation factors, this is a 3.07-approximation algorithm. Again, with the current best results, this gap is less than 5. For the case where clients have arbitrary demands and the entire demand for a client must be served by the same facility, we provide a $\rho_{ST} + 2\rho_{UFL}$ approximation, which is currently at most 4.59.

Key words:

UFL, Steiner tree, CFL, network design

1. Introduction

A ubiquitous problem faced by every corporation which manufactures and sells products to a geographically spread-out market is the following: Where should the factories be built, and how should the finished goods be transported to the markets, so as to minimize costs?

Consider the following scenario: A multinational corporation wishes to enter a promising new geographic market, characterized by demand at each city. It has identified potential locations of its manufacturing facilities, and the associated costs. Suppose the shipping of the good is to be outsourced to be a transport company. This transport company has only one type of truck, with a large

capacity. For each truck, the transport company charges at a fixed rate per mile, and offers no discount in case the truck is not utilized to full capacity. The overall logistics problem facing the corporation is to decide on the location of its manufacturing facilities, and a shipping plan of the finished goods to each city, so that the total demand at each city is met and the total cost is minimized. Assume for the sake of simplicity that facilities have no capacity limitations.

If the transport company charged in proportion to the amount shipped instead of the discrete number of trucks used, the problem becomes the uncapacitated facility location problem (UFL). If the facility location costs were not an issue, the problem becomes a single sink edge installation problem [1]. If several facilities are open, they can all be identified into a single sink node. We call this the capacitated facility location problem (CFL for short).

A variant of this problem is the median version. There are no facility location costs, but we are not allowed to open more than p facilities. We call this the capacitated p -median (CpM) problem. We note the assumption that the edge-lengths obey the triangle inequality is without loss of generality. These problems generalize known NP-hard problems, and hence are NP-hard. We present polynomial time approximation algorithms with constant approximation ratios for these problems.

The uncapacitated facility location (UFL in short) problem has been the subject of much recent activity. Charikar [2] gave the first constant factor approximation algorithm for the p -median problem with metric costs. Korupulu [3] provide a improved approximation for local search technique and was deep improved by Arya [4] to a factor of $3+\epsilon$, which is the best known at present. Hassin [5] provide a constant factor approximation for the single sink single cable installation problem. We use their method as a subroutine. A constant factor approximation for the multiple cable single sink edge installation problem was first provided by Guha [6]. Cable installation problems have also received a lot of attention in the recent [1,5,6,7].

2. CFL Problem definition

The capacitated-cable facility location problem(CFL) is defined as follows:

Given an undirected graph $G=(V,E)$, there is a weight function on the edges, $c: E \rightarrow \mathbb{R}^+$, which satisfies the triangle inequality. The clients consist of a subset of nodes, $D \subset V$. The set of potential facilities, $F \subset V$, is also part of the input. Each potential facility $j \in F$ has a facility opening cost of f_j and given an integer $u > 0$, which is the capacity of the cable type available to us.

Each client has a demand of one unit, which needs to be serviced by routing one unit of flow from it to some open facility. On any edge, we are only allowed to install integral amounts of the cable. If we install k_e copies of the cable on edge e , we can route uk_e units of flow through it, and it costs $c_e k_e$. Hence our total cost is the cost of all cables installed plus the cost of all the facilities we have opened. The objective of CFL is to open facilities and install cables connecting clients to open facilities such that no capacity constraint, all clients are served and the total cost is minimized.

3. Lower bounds to CFL

We give two lemmas which provide lower bounds to an optimal solution of CFL

Lemmas 1 Consider a UFL instance defined as follow: The set of clients and potential facilities remain the same as in the CFL instance, but all edges e , we set the edge cost to be c_e/u . Then the cost of an optimal solution to this UFL instance is a lower bound on the optimal solution to CFL.

Proof: Consider the optimal solution CFL. In the UFL instance, open all facilities which were opened by CFL. Every client in CFL is able to send one unit of flow to an open facility. Construct these flow paths: for each client, assign it to the facility it is assigned to in CFL. The cost of this assignment is at most $1/u$ of that of the flow path used by this client in the CFL solution, by triangle inequality. This constitutes a feasible solution to UFL, of cost no more than that of the CFL solution. Hence an optimal UFL solution has cost at most that of the optimal CFL solution. \square

Lemmas 2: Consider the graph $G'=(V \cup \{R\}, E \cup E')$ where $E' = \{(i,r)\} : i \in F\}$ and $c_{ir} = f_i$. Define the set of terminal to be $R = D \cup \{r\}$. Then the cost of a minimum Steiner tree in G' is a lower bound on the optimal solution of CFL.

Proof: Consider the optimal solution to CFL. The set of edges in the CFL solution, along with the edges (j,r) such that facility j is opened in the CFL solution, constitutes a Steiner tree in G' of the same cost as the CFL solution. Dropping all but one copy of edges which have

multiplicity more than 1 in the CFL solution only reduces the cost. Hence an optimal Steiner tree must cost no more than the optimal CFL solution. \square

4. Algorithm CFL

We use above two lemmas to build our solution and use a flow rerouting algorithm introduced by Hassin[1] to efficiently construct our solution.

On the one hand, we merge the two solution to obtain a feasible solution of cost no more than the sum of these two approximate solution. We adapt a re-routing algorithm described in [1]. We first open all facilities identified by the earlier two phases. If such a subtree has at most u clients, this subtree along with the facility it is attached to is a feasible solution, without adding any additional copies of the cable.

On the other hand, a subtree that has more than u clients is not feasible right away, since more cables have to be installed along the tree to route all the demand in this overloaded subtree. This is where we use the UFL solution. We clump the demands in these overloaded subtrees into subtrees which are disjoint with respect to edge capacities such that each new subtree has exactly u clients. The fact that such a clumping is possible was proved in [1]. We describe it in detail in Algorithm CFL as follow:

Algorithm CFL include UFL phase, Steiner tree phase, Merge phase and Prune phase.

(1)UFL phase:

- 1) Convert into UFL instance by changing edge cost to c_e/u .
- 2) Solve UFL(approximately).
- 3) Let F_1 denote the facilities opened.
- 4) For a client j , let $\phi(j)$ be its assigned Facility

(2)Steiner tree phase

- 1) Create a new root node r .
- 2) For every $i \in F$, add an edge (i,r) with cost f_i
- 3) Define the terminal set $R = D \cup \{r\}$.
- 4) Solve(approximately) the Steiner tree problem.
- 5) Let T denote this tree
- 6) Orient all edges to point towards the root along T .
- 7) Let F_2 be the set of facilities from which there are edges to r in T .
- 8) Let T_i be the subtree of T rooted at i , for all $i \in T$.

(3) Merge phase

- 1) Open all facilities in $F_1 \cup F_2$.
- 2) For all $i \in F_2$

- 3) Let D_i be the set of clients in T_i .
 - 4) Install cable on all edges in T_i .
 - 5) While $|D_i| > u$
 - 6) Let V' be the set of nodes at which the incoming demand on each edge is less than u , but the total demand is at least u .
 - 7) For all $v \in V'$
 - 8) For every child w of v , let T_w be the subtree rooted at w .
 - 9) Let (j_w, i_w) be the nearest client-facility pair. in T_w , that is $c_{j_w, i_w} = \min_{j \in T_w, i \in F} c_{ij}$
 - 10) Pick the cheapest $|D_v / u|$ such pairs.
 - 11) Install one cable on each such picked pair (j_w, i_w) .
 - 12) Route all demand in T_w to i_w via j_w .
 - 13) Route remaining demand to either some picked pair or to w , in such a way that all newly installed cables are saturated. This means that the total remaining demand to v is less than u .
 - 14) Remove all satisfied demands from D_i .
- (4) Prune phase:
- 1) Remove all cables on which flow is zero
 - 2) Close all facilities which serve no demand.

We prove algorithm CFL in lemmas 3.

Lemmas3: The solution produced by algorithm CFL is feasible for CFL.

Proof: In the demand routing phase, client demands from a subtree that is not fully served in an iteration may be reassigned in a later iteration. In particular, let's say part of the subtree's demand is routed to a picked client(j_w^1) in a sibling subtree using upward flow on its parent arc. In the next iteration of the while loop(merge phrase), suppose one of the unsatisfied clients(j_w^2) in this subtree is part of a picked pair. Now, flow from sibling subtrees in this iteration may be routed into it using a downward flow on the same parent arc. However, by standard flow cancellation arguments, no cable is used in both an upward and a downward direction. This flow cancellation implicitly reassigns the clients from the subtree initially assigned to j_w^1 to j_w^2 . and instead redirects the appropriate demand from elsewhere headed for j_w^2 in the second iteration to j_w^1 .

□

The flow cancellation only reduces flow in the upward direction. If any cable has an upward flow, this flow has value at most $u-1$, and this may potentially be cancelled by

downward flow when a client in the subtree below it is part of a picked pair. Downward flow is assigned to any cable at most once, and the quantity of flow assigned is at most $u-1$. After such an assignment, all the clients in this subtree are deleted from further consideration.

For each such clump, we use the UFL solution to select the client which is closest to an open facility in the UFL solution, and install one cable from this client to its nearest open facility. The idea is that since each client can pay a $1/u$ fraction of the cable cost to the facility assigned to it by the UFL phase, u such clients in a clump can together pay for one full cable from a client to an open facility if this distance is the cheapest distance among these u clients. In order to achieve this, we need to re-route flow the $u-1$ other clients to our selected client in a clump. However, this rerouting takes place along the original Steiner tree solution at an extra cost since the subtrees obtained in the clumping are disjoint with respect to edge capacities. We finally prune the solution by getting rid of unused facilities and cables.

We have argued that both the underlying UFL instance and the associated Steiner tree problem are lower bounds for Algorithm CFL instance. Hence the facilities opened by these two phases can be paid for by these two lower bounds.

The cables purchased by the Steiner tree phase can be paid for by the Steiner tree lower bound. Each cable has exactly u demand flowing through it. Each of the terminals which use this cable were assigned a facility whose distance is at least the length of the cable in the UFL phase. Hence we can charge the cost of this cable to the cost in the UFL solution.

We use ρ_{ST} and ρ_{UFL} to denote the currently best known approximation ratios for the Steiner tree and UFL problems respectively. We have the following theorem.

Theorem: Algorithm CFL is a $\rho_{ST} + \rho_{UFL}$ approximation algorithm for CFL

Proof: This follows from lemmas 1 to 3.

□

The current best approximation algorithm for the Steiner tree problem is the one by Robins and Zelikovsky[8], which achieves an approximation factor of 1.55. The algorithm of Mahdian, Ye and Zhang[9] is the current best approximation for UFL, with a performance ratio of 1.52. With these values for ρ_{ST} and ρ_{UFL} , Theorem Algorithm CFL give a 3.07 approximation.

Algorithm CFL for Non-uniform demands, generalizes directly to the case at the clients, provide we are allowed to split the demand at each client to different facilities. If the demands are unsplittable, Hassin[1] showed how their (single sink) problem can be solved in the unsplittable demand case with a slight increase in the approximation ratio. Clients which have more than u demand can be sent directly to their nearest facilities, incurring an additional

factor of at most 2. To assign the remaining clients, we proceed as before.

We now demands to lie between u and $2u$, and now use the UFL bound at most twice. Hence the approximation ratio for this problem is at most $\rho_{ST} + 2\rho_{UFL} = 4.59$.

5. Hardness and relation to other problems

If there is only a single potential facility ($|F| = 1$) and u is infinity, then the problem reduces to the Steiner tree problem. If there is a single facility an $1 < u < \infty$, CFL is the single-sink, single-cable edge installation problem. If $u = 1$ but $|F| > 1$, CFL is the the uncapacitated facility location problem. All these problems have been studied in the past, and all three are known to be MAX-SNP-hard. Hence CFL is also MAX-SNP-hard, meaning that there is a constant $c > 1$ such that it is impossible to approximate CFL better than c , unless $P = NP$.

6. Conclusion

We present a $\rho_{ST} + \rho_{UFL}$ approximation algorithm for CFL, where ρ_P is any approximation factor achievable for the problem P. We do this by carefully combining solutions to appropriately set up Steiner tree and UFL problems that capture two natural lower bounds for our problem. With the current best approximation factors, this is a 3.07-approximation algorithm. Again, with the current best results, this gap is less than 5. For the case where clients have arbitrary demands and the entire demand for a client must be served by the same facility, we provide a $\rho_{ST} + 2\rho_{UFL}$ approximation, which is currently at most 4.59.

This research is at an early stage of study. There still remain several problems to be solved in future such as :

- (i) finding a integer programming formulation of CFL
- (ii) described above extend to providing a constant factor rounding algorithm for the linear relaxation fo the IP formulation.

Acknowledgment

we would like to acknowledge the financial support by is supported by National Doctor Specialized scientific research Foundation .

References

- [1] R.Hassin,R.Ravi and F.S.Salman. Approximation algorithms for a capacitated network design problem.In proceedings of the 3rd *International Conference on Approximation Algorithms for Combinatorial Optimization*,167-176.2000.
- [2] M.Charikar,S.Guha,D.shmoys and E.Tardos.A constant-factor approximation algorithm for the k-median problem. In Proceedings of the 31st Annual ACM Symposium on Theory of Computing,1-10,1999.
- [3]M.Korupulu,C.Plaxton and R.Rajaraman. Analysis of a local search heuristic for facility location problems. In Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms.1-1-.1988.
- [4]V.Arya,N.Garg,R.Khandekar,V.Pandit.Local search heuristic for k-median and facility location problems.In Proceedings of the 33rd Annual ACM Symposium on Theory of Computing,21-29,2001.
- [5] S.Guha,A.Meyerson and K.Munagala. Improved combinatorial algorithms for single sink edge installation problem, In Proceedings of the 33rd Annual ACM Symposium on Theory of Computing, 383-388,2001.
- [6] A.Goel and D.Estrin .Simultaneous optimization for concave cost: single sink aggregation or single source buy-at-bulk. In Proceedings of the 14th Annual ACM-SIAM Symposium on discrete Algorithms, 499-505,2003.
- [7] A.Gupta,A.Kumar. Approximations via cost-sharing. In proceedings of the 44th Annual IEEE Symposium on foundations of Computer Science,606-615,2003.
- [8] G.Robins and A.Zelikovsky. Improved Steiner tree approximation in graphs. In processing of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms,770-779,2000.
- [9] M.Madian, Y.Ye and Zhang. A 1.52 approximation algorithm for the uncapacitated facility location problem. In Proceedings of the 5th International Conference on Approximation Algorithms for Combinatorial Optimization,229-242,2002.