

Fault-Tolerant Facility Location

Chaitanya Swamy*

David B. Shmoys*

We study a fault-tolerant generalization of the classical uncapacitated facility location problem. We want to open a subset of facilities from a given set \mathcal{F} and assign each client j in a given set \mathcal{D} to $r_j \geq 1$ *distinct* open facilities so as to minimize the sum of the facility opening costs and the client assignment costs. We also consider the fault-tolerant k -median problem where instead of facility costs we are given a number k of facilities that may be opened, and the objective is to minimize only the assignment cost. Multiple facilities provide a safeguard against failures. If the facility closest to a client ‘fails’, the other facilities assigned to it could be used to serve it, e.g., in designing a network involving the placement of caches or routers, one would like to connect a client to multiple caches or routers so as to be resistant under node or link failures. We consider the case where the distances, c_{ij} , form a metric.

THEOREM 1. *There is a 2.076-approximation algorithm for fault-tolerant facility location.*

This is currently the best known guarantee. If all requirements are equal, we give a 1.52-approximation algorithm by modifying the algorithm of [8, 13], which matches the current best guarantee for uncapacitated facility location. For the k -median version with uniform requirements, we obtain a 4-approximation algorithm.

Related work. A variety of techniques have been used for the design and analysis of approximation algorithms for metric uncapacitated facility location over the past several years. The first constant-factor approximation algorithm for this problem was due to Shmoys, Tardos & Aardal [14] who gave a 3.16-approximation algorithm, using the *filtering* technique of Lin & Vitter [11] to round the optimal solution of a linear program. After an improvement by Guha & Khuller [6], Chudak & Shmoys [4, 5] gave an LP rounding based $(1 + \frac{2}{e})$ -approximation algorithm. They used information about the structure of optimal primal and dual solutions, and combined randomized rounding and the decomposition results of [14] to get a variant that might be called *clustered randomized rounding*. Sviridenko [15] improved the ratio to 1.58. Jain & Vazirani [9] gave a combinatorial *primal-dual* 3-approximation algorithm. Mahdian, Markakis, Saberi & Vazirani [12] showed that a greedy set-cover type primal-dual algorithm has an approximation ratio of 1.861. This was improved in [8, 13] to 1.52,

which is the best known guarantee. Charikar, Guha, Tardos & Shmoys [3] gave the first constant-factor algorithm for the k -median problem based on LP rounding. This was improved in [9, 2, 8, 1] to $(3 + \epsilon)$ [1].

The fault-tolerant facility location problem was introduced by Jain & Vazirani [10] who gave an $O(\log(\max r_j))$ -approximation algorithm. Our algorithm is based on LP rounding. We consider the following LP: minimize $\sum_i f_i y_i + \sum_{j,i} c_{ij} x_{ij}$ s.t. $\sum_i x_{ij} \geq r_j$, $\forall j$, and $0 \leq x_{ij} \leq y_i \leq 1$, $\forall i, j$. The dual is: maximize $\sum_j r_j \alpha_j - \sum_i z_i$ s.t. $\alpha_j \leq c_{ij} + \beta_{ij}$, $\forall i, j$, $\sum_i \beta_{ij} \leq f_i + z_i$, $\forall i$, and $\alpha_j, \beta_{ij}, z_i \geq 0$, $\forall i, j$. Variable y_i indicates if facility i is open, and x_{ij} indicates if client j is connected to facility i . An integral solution to the LP corresponds to a solution to our problem. Guha, Meyerson & Munagala [7] round the primal LP using filtering and the decomposition technique of [14], to obtain a 3.16-approximation. Subsequent greedy improvement, as in [6], reduces the ratio to 2.47.

Our Algorithm. Our algorithm does not use filtering, but is based on the *clustered randomized rounding* technique of [5]. We exploit the *complementary slackness* conditions to bound the cost of our solution in terms of both the primal and dual optimal solutions: the primal solution is α_j -close, i.e., $x_{ij} > 0 \Rightarrow c_{ij} \leq \alpha_j$. One difficulty in using LP duality to prove an approximation ratio is the presence of the $-\sum_i z_i$ term in the dual: bounding the cost in terms of $\sum_j r_j \alpha_j$ is not enough. This problem also crops up in designing approximation algorithms for the k -median problem. In general this is not an easy problem to tackle, and for the k -median problem, the only known LP rounding algorithm [3] uses just the optimal primal LP solution. For fault-tolerant facility location, complementary slackness allows us to, in effect, get rid of the negative z_i s by a single pruning phase. Since $z_i > 0 \Rightarrow y_i = 1$, we can open all such i and charge the opening cost to the LP.

One needs to overcome several obstacles to extend [5] to the fault-tolerant setting and prove a good approximation guarantee. We now want every demand j to have r_j open facilities ‘near’ it. Our algorithm also clusters facilities around demand points, called cluster centers, and opens at least one facility in each cluster. Each facility i is opened with probability proportional to y_i . The randomization step allows us to reduce the service cost, since now there is a high probability that for any j , a facility i with $x_{ij} > 0$ is open.

We ensure that demand j has r_j nearby open

*{swamy,shmoys}@cs.cornell.edu. Cornell University, NY 14853. Research supported by NSF grant CCR-9912422.

facilities by introducing a notion of *backup requirement* which is initialized to r_j . Whenever we create a cluster we decrement the backup requirement of all j which share a facility with the cluster created. The facility opened within this cluster serves as a *backup facility* for each such j and we argue that this facility is near j . As long as the backup requirement of j is at least 1, it is a candidate for being chosen as a cluster center. So j will share facilities with r_j clusters and these provide the r_j nearby facilities. To argue feasibility, we need to *open different facilities in different clusters*. Also we would like each cluster to use up a facility weight of 1. If the facility weight within a cluster is too small, opening a facility within the cluster would incur a huge cost; if the facility weight is more than 1, then we are using up a fractional facility weight of more than 1 while opening only one facility — so a demand j may not have r_j nearby facilities. Guha et al. [7] open facilities deterministically within a cluster, possibly splitting a facility across clusters. Splitting facilities allows them to achieve a facility weight of 1, and the deterministic choice allows them to argue that a facility is opened in at most one cluster even if it is split across clusters. With randomization however, any facility within a cluster may be opened, so *clusters have to be disjoint* otherwise the same facility could be opened in different clusters. But once we require clusters to be disjoint, we cannot expect a cluster to have a facility weight of exactly 1. We tackle this problem by introducing another pruning phase where we open all facilities i with ‘large’ y_i . This allows us to *pack a substantial facility weight within a cluster* without exceeding the limit of 1.

To analyze the algorithm, we view a demand j with requirement r_j as being composed of r_j copies which have to be connected to *distinct* facilities. Instead of assigning j to the r_j nearest open facilities, we consider a suboptimal assignment scheme. Each copy of j is allotted a subset of facilities $\{i\}$, such that $x_{ij} > 0$, and a unique backup facility. A copy may only be assigned to an allotted facility or to the backup facility for that copy. For feasibility we have to ensure that a facility is allotted to at most one copy. Due to the pruning phase we can allot a substantial facility weight to each copy, which increases the probability that some facility allotted to the copy is open. If none of the allotted facilities is open, we assign the copy to the backup which we argued earlier was not too far from j . Overall this gives an approximation guarantee of 2.12. To do better, we allot facilities in a better way. We use the *pipage rounding* technique [15] to derandomize a hypothetical randomized process which allots a facility probabilistically across different copies (if necessary), so that each copy gets an equal allotment of facility weight.

This gives a 2.076-approximation algorithm.

Other Results. Mahdian et al. [12] gave a 1.861-approximation algorithm for the uniform-requirement case. We improve this to get a 1.52-approximation algorithm, by extending the algorithm of Jain, Mahdian & Saberi [8]. We analyze our algorithm using the dual fitting approach and arrive at the same factor LP as in [8], and hence obtain the same performance guarantees. Combined with greedy improvement and the analysis in [13], we get a 1.52-approximation algorithm. Using Lagrangian relaxation [9] this also gives a 4-approximation algorithm for the fault-tolerant k -median problem with uniform requirements.

References

- [1] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local search heuristics for k -median and facility location problems. *Proc. 33rd ACM STOC*, 21–29, 2001.
- [2] M. Charikar and S. Guha. Improved combinatorial algorithms for the facility location and k -median problems. *Proc. 40th IEEE FOCS*, 378–388, 1999.
- [3] M. Charikar, S. Guha, É. Tardos, and D. B. Shmoys. A constant-factor approximation algorithm for the k -median problem. *JCSS*, 2002 To appear.
- [4] F. Chudak. Improved approximation algorithms for uncapacitated facility location. *Proc. 5th IPCO*, 180–194, 1998.
- [5] F. Chudak and D. Shmoys. Improved approximation algorithms for the uncapacitated facility location problem. Submitted to *SIAM J. Comp.*.
- [6] S. Guha and S. Khuller. Greedy strikes back: Improved facility location algorithms. *J. Algorithms*, 31:228–248, 1999.
- [7] S. Guha, A. Meyerson, and K. Munagala. Improved algorithms for fault tolerant facility location. *Proc. 12th ACM-SIAM SODA*, 636–641, 2001.
- [8] K. Jain, M. Mahdian, and A. Saberi. A new greedy approach for facility location problems. *Proc. 34th ACM STOC*, 731–740, 2002.
- [9] K. Jain and V.V. Vazirani. Approximation algorithms for metric facility location and k -median problems using the primal-dual schema and Lagrangian relaxation. *JACM* 48:274–296, 2001.
- [10] K. Jain and V.V. Vazirani. An approximation algorithm for the fault tolerant metric facility location problem. *Proc. 4th APPROX*, 177–183, 2000.
- [11] J. H. Lin and J. S. Vitter. ϵ -approximations with minimum packing constraint violation. *Proc. 24th ACM STOC*, 771–782, 1992.
- [12] M. Mahdian, E. Markakis, A. Saberi, and V. Vazirani. A greedy facility location algorithm analyzed using dual-fitting. *Proc. 5th APPROX*, 127–137, 2001.
- [13] M. Mahdian, Y. Ye, and J. Zhang. Improved approximation algorithms for metric facility location. *Proc. 6th APPROX*, 229–242, 2002.
- [14] D. B. Shmoys, É. Tardos, and K. I. Aardal. Approximation algorithms for facility location problems. *Proc. 29th ACM STOC*, 265–274, 1997.
- [15] M. Sviridenko. An improved approximation algorithm for the metric uncapacitated facility location problem. *Proc. 9th IPCO*, 240–257, 2002.