

Minimum Cost Wireless Broadband Overlay Network Planning

Peng Lin, Hung Ngo, ChunMing Qiao
Department of Computer Science and Engineering
State University of New York at Buffalo
Buffalo, NY 14260
Email: {penglin,hungngo,qiao}@cse.buffalo.edu

Xin Wang
Dept. of Electrical and Computer Engineering
State University of New York at Stony Brook
Stony Brook, NY 11794
Email: xwang@ece.sunysb.edu

Ting Wang, DaYou Qian
NEC Laboratories America
Email: {ting,dqian}@nec-labs.com

Abstract

Wireless broadband networks, especially WiMAX networks, have emerged in the industry recently and many challenging research issues arise. In this paper, we proposed a heuristic clustering algorithm for minimum cost wireless broadband overlay network deployment. Moreover, we also modified and implemented two heuristic algorithms based on classic linear programming based capacitated facility location algorithms. We analyzed the theoretical worst-case performance ratio of our algorithm and our numerical results showed that our algorithm performs much better in practical network settings.

1 Introduction

Recently, wireless broadband access technology has drawn intensive attention in both the academic and industrial research institutes. The coverage of wireless high-speed broadband service is expected to expand dramatically in the near future [6]. One of the main technologies is the fast-evolving IEEE 802.16 technology, which is synonymous with WiMAX (Worldwide Inter-operability for Microwave Access).

The IEEE 802.16d standard draft [13] has specified air interfaces operating on both the licensed and licence-exempt frequencies. The two main modes defined for the MAC layers are the PMP(point-to-multipoint) mode and the Mesh mode. The former provides a centralized up-link/downlink wireless data service, similar to the cellular system, while the latter allows for a better coverage, flex-

ibility and scalability. Due to the wireless broadcasting nature of the 802.16 PHY/MAC protocols, there exists a need to maintain a network hierarchy to maximize the network efficiency and minimize co-channel interference. And this wireless broadcast nature also distinguishes a wireless broadband network design problem from a wireline broadband network design problem. In the paper, we consider the network planning problem of the following network topology: Traffic demands (routed via the 802.11 interface) are aggregated at a set of Subscriber Station(SS) nodes which are equipped with both 802.11 and 802.16 interfaces. Subsequently, the traffic demands at SS nodes will be satisfied by a set of Base Stations (BS) nodes which functions in the PMP mode [13]. At the wireless backhaul network level, these BS nodes and a set of intermediate WiMAX nodes will form a meshed backhaul network which will connect the BS nodes with IAP (Internet Access Point) nodes, one for each backhaul subnet.

The rest of the paper is organized as follows. We formally define our problem in Section 2. Existing related works are reviewed in Section 3. In Section 4, we present a heuristic algorithm for getting a near-optimal backhaul network deployment solution. Two modified algorithms, based on classic facility location algorithms are presented in Section 5. Simulation results are presented in Section 6. Finally, we conclude our work in Section 7.

2 Problem Description

In this section, we give a general description of the problem that we are interested in. We need to derive a minimum cost backhaul wireless network (BS nodes) deploy-

ment such that demands (from SS nodes) can be satisfied without violating wireless capacity constraints.

Let BS represent the potential BS nodes, let S_S represent the set of SS nodes and we are given a demand vector d_j , for $j \in S_S$, which denotes the average bandwidth request at SS node j . Let b_i denote the aggregate wireless capacity of BS node i . Let $N(i)$ denote the set of SS nodes which are within communication range for BS node i . We use $c_i(k)$ to represent the cost of picking subnet k type BS at location i (i.e., BS nodes of different types incur different costs). Note that for opening a BS at a specific location, there may be different BS costs incurred when choosing from different subnet types. The objective of the problem is to minimize the total cost of opening BS nodes which cover all traffic demands. The integral variable y_i indicates if a backhaul BS node i is chosen, $0 \leq x_{ij} \leq 1$ is the fraction of traffic from SS node j assigned to BS node i .

Our problem, named as PMP-Association Problem, can then be formulated as an mixed integer programming (MIP) problem:

$$\begin{aligned}
 & \text{minimize} && \sum_{i \in BS} c_i(k) y_i && (1) \\
 \text{subject to} &&& \sum_{i \in BS} x_{ij} \geq 1, && j \in S_S \\
 &&& x_{ij} \leq y_i, && i \in BS, j \in N(i) \\
 &&& b_i y_i - \sum_{j \in N(i)} d_j x_{ij} \geq 0, && i \in BS \\
 &&& y_i = \{0, 1\}, && i \in BS \\
 &&& 0 \leq x_{ij} \leq 1 && i \in BS, j \in S_S
 \end{aligned}$$

It should be noted that the formulation does not assume triangle inequality. Triangle inequality means that for a complete undirected graph, and for any three vertices u , v and w , it satisfies that $cost(u, v) \leq cost(u, w) + cost(v, w)$. Therefore, existing approximation algorithms for facility locations problems which assumes triangle inequality can not be directly applied here.

The PMP-Association problem we address in this paper is NP-hard. This can be easily seen since when ignoring capacity and assuming uniform cost function, we obtain a reduction from the weighted set cover problem [3, 14], which has been known to be NP-hard for a long time. Given this fact, we are interested in finding an efficient algorithm which gives a near-optimal solution in the worst case, and is computationally efficient. More specifically, we attempt to design a combinatorial algorithm which can be implemented efficiently. The advantage of such algorithm is that it can be potentially implemented as an online algorithm for incremental optimization in a dynamic traffic request situation.

3 Existing Literature

In essence, the proposed work is related to those on multi-homing and smart routing techniques [7, 12]. Nevertheless, it also introduces different features from existing models used in the other studies in that our focus is on the offline optimal cost network design algorithm in a static wireless network topology while the focus of the existing works are on adaptive routing algorithms.

In wireless ad-hoc networks, there have also been research done in clustering relaying networks to minimize the access points to wireline backbones [4, 5]. In [4], the author proposed a set of cluster algorithms by exploiting the properties of a unit disk graph [2], i.e., uniform transmission range for all network nodes, to decide the minimum set of connected subtrees while satisfying traffic demands and relay constraints. Our problem differs in that we do not assume uniform transmission range in our problem and the transmission range and capacity between any SS-BS node pair can be location dependent. In [5], authors addressed the design of a multi-hop ad-hoc network, which would be useful if we were to design the first tier network and determine how the traffic from houses can be routed and how many SS nodes are needed.

Another related problem is a weighted version *Euclidean capacitated covering by Disks problem* (or tagged as “L-balanced ρ -dominating Sets Problem”) which was shown to be NP-hard [3,10,11]. Both are similar to the general framework of set cover problems. Judit Bar-Ilan et al. has given a best-known $\ln n$ -approximation ratio algorithm which runs in polynomial time for the graph version of the problem in [3], while a heuristic algorithm is used in [10] to address the problem under the L_2 metric. A sub-exponential algorithm was proposed in [1] for near-optimal approximation. Our problem differs in that we consider a broadband access network problem where demands/capacity are location dependent and the deployment cost is subnet dependent, whereas in the other works, general uniform demand and server capacity were considered.

4 Our Proposed Methodology

In this section, we propose an algorithm to decide how to choose the minimum cost BS node set such that all demands from the SS nodes can be satisfied without violating the capacity constraints associated.

In this paper, we assume that the flow assignment can be fractional, in the sense that the traffic demand from an SS node can be divided into smaller demands and assigned to various BS nodes.

Since there does not exist an efficient approximation algorithm for the above capacitated set cover with demands, the following heuristic solution is proposed:

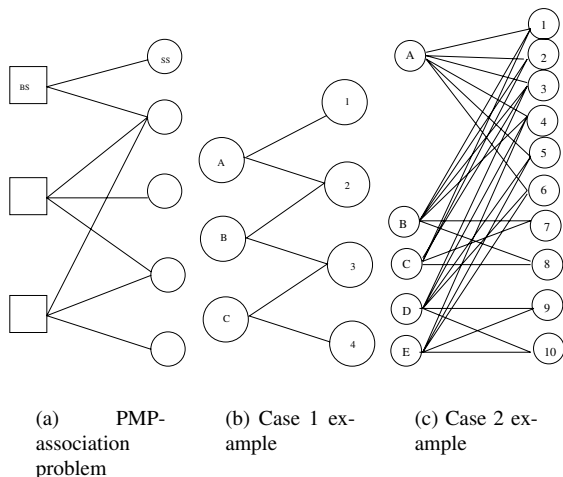


Figure 1. Illustrations for the PMP problem

First, we use a greedy approximation algorithm to determine the minimum number of locations, then at the chosen locations, find the corresponding subnet with minimum cost.

Before we describe the approximation algorithm, we notice that in practical settings, there may be cases where the actual number of BS nodes and possible cost can be further decreased using a heuristic approach without affecting the cost performance.

Below, we present two scenarios to show the possible reduction in the number of BS nodes chosen. For simplicity, we assume unit demand at each SS node. The same idea applies to general cases as well.

- Case 1: In Figure 1(b), all three nodes in the candidate BS nodes are having the same degree, it is not specified how such cases can be handled in [3], it is normally up to the maximum flow algorithm to pick arbitrarily a feasible solution. However, we notice that there is certain interesting property here: Picking node A or C first will lead to the optimal solution of two BS nodes in the greedy approach, while picking node B first will lead to a solution of three BS nodes. The intuition here is that node A and C have covered node 1 and 4 which has a degree 1, therefore, A and C have to appear in the optimal set whatsoever. If we include A and C as early as possible in the greedy approach, it will lead to a BS set which is no larger than another feasible solution using the same greedy approach.
- Case 2: In Figure 1(c), we consider an example of an optimal solution of two BS nodes. Choosing nodes B or D will lead to a better solution than any solution choosing A, as the SS nodes covered by node A can be

dominated by nodes $\{B \cup D\}$. It is not obvious that B or D should be selected first, the general observation is that among the three candidates A, B and D, which are the maximum flow solutions, B (or D) covers the nodes, e.g., nodes 7 – 10, whose degree is 2. while the minimum degree of the nodes covered by A is 3. Generally, since we have to cover all nodes eventually, B and D are more likely to be included in the final assignment. Therefore, we include them in our solution in preference to A.

In our network models, most SS nodes to be covered should have a relatively low degree number, which indicates the possible subnets an SS node has subscribed service to (as there is no relay among SS nodes). In fact, some of the SS nodes such as the boundary nodes and nodes which have subscribed to only one contiguous subnet due to policy issues have a degree of one. Therefore, it is more likely that the heuristic which picks BS nodes covering a lower degree SS node first will be able to reduce the size of the BS set.

Algorithms 1-3 describe the proposed iterative covering heuristic algorithm based on the greedy approach proposed in [3]. We extend the original algorithm by adding the filtering process and allowing the maximum flow algorithm to taking care of non-uniform demands and fractional assignments. The performance of our algorithm is given in the following theorem. The detailed proof can be found in [9].

<p>input : Facility Set BS, Demand Set SS, Associated Cost</p> <p>output: Chosen facilities, An assignment of the demands x_{ij},</p> <pre> 1 $U \leftarrow S_S$; 2 $S_B \leftarrow \emptyset$; 4 while $U \neq \emptyset$ do 5 Construct a bipartite graph $G' = (BS, U, E)$, with $E = \{(x, y) x \in BS, y \in U\}$; 6 Add two vertices S and T, connect S to every node in BS, connect every node in U to T; 7 Define capacities c on the graph: $c(S, x) \leftarrow b_x, c(x, y) \leftarrow \infty, c(y, T) \leftarrow d_y$; 8 Run procedure <i>MaxIF</i>(BS); 9 end 10 for $i \leftarrow 1$ to S_B do 11 Pick the index of a subnet such that the cost incurred when opening a BS node at location i is minimized; 12 end</pre>

Algorithm 1: Iterative Covering

Theorem 1. Given a candidate location set obtained above, our algorithm gives a minimum cost of no

- 1 Iterate through all nodes v in $BS \setminus \{S_B\}$, and construct a subgraph G' from the original graph G ;
- 2 During each iteration, denote its corresponding maximum flow value as M_v ;
- 3 Further_Refine();
- 4 Find \hat{v} such that $\gamma(S_B \cup \{\hat{v}\}) = \max_v(M_v)$, $\forall v \in BS \setminus \{S_B\}$;
- 5 $S_B \leftarrow S_B \cup \{\hat{v}\}$;

Algorithm 2: MaxIF(BS)

- 1 **if** $\gamma(S_B \cup \{\hat{v}_1\}) == \gamma(S_B \cup \{\hat{v}_2\})$ **then**
- 2 pick v_i such that v_i covers a node with the minimum degree in the assignment, for any $i \in \{1, 2\}$;
- 3 **end**

Algorithm 3: Further_Refine()

more than $\frac{\lceil (\ln n + \ln m) \rceil C_{max}}{C_{min}}$ of the optimal cost of the PMP-Association problem, where the total traffic demands is denoted as $n * m$ and we define $c_{max} = \max_{j \in BS, k \in SubnetSet} c_j(k)$, $c_{min} = \min_{j \in BS, k \in SubnetSet} c_j(k)$, where *SubnetSet* is the set of all subnets in the network.

The theoretical bound of the worst-case performance ratio is not practical enough from the application's point of view, due to the general hardness of the capacitated non-metric space nature. However, for general network settings, our simulation results produced good average-case performance, with respect to the optimal solutions.

5 Extended Capacitated Facility Location Algorithms

In [8], R. Levi et al. proposed a hard¹ capacity facility location algorithm. As discussed above, facility location-based algorithms do not apply directly to our network plan problem, since it assumes that the bipartite graph of the BS station set and the demand SS station set is fully connected and the cost of connecting any two stations satisfies metric space properties, such as triangle inequalities. Therefore, it is not suitable for general wireless networks with limited communication range. For instance, for a traditional facility location algorithm, an SS node may be connected to an BS node which is far beyond the communication range in practice.

¹In the context of facility location problems, hard capacity means that only one facility with a fixed capacity is allowed at any specific location, in contrary to the soft capacity version in which an arbitrary number of facilities are allowed at any location, equivalently increasing the facility capacity.

In order to study the impact introduced by the violation of the metric property, we examine two extended capacitated facility location algorithms based on the approximation algorithm proposed in [8] which has an approximation ratio of 5 under the metric space assumption. We give next a brief outline of the algorithm developed using the technique of capacitated facility location algorithms. Interested readers may refer to [8] for more thorough description and analysis.

input : Facility Set BS, Demand Set SS, Capacity Set $\{u_i\}$

output: Chosen facilities, An assignment of the demands x_{ij} ,

- 1 Solve the optimal primal and dual LP-relaxations and denote the solutions as (x, y) and $(\alpha, \beta, \gamma, z)$ respectively;
- 2 $F_j = \{i : x_{ij} > 0, i \in BS, j \in SS\}$;
- 3 $S \leftarrow SS$;
- 4 $C \leftarrow \emptyset$;
- 5 $f_{ij} = 1, \forall i, j$ in range;
- 6 $f_{ij} = \infty, \forall i, j$ out of range;
- 8 **while** $S \neq \emptyset$ **do**
- 9 Pick $j \in S$ such that $j = \arg \min\{\alpha_k | k \in S\}$;
- 10 $B_j = \{i \in F_j : i \notin \bigcup_{k \in C} N_k \text{ and } f_{ij} \leq \min_{k \in C} f_{ik}\}$;
- 11 $C = C \cup \{j\}$;
- 12 $S = \{j \notin C : \sum_{i \in B_j} x_{ij} \geq 0.5\}$;
- 13 Form cluster B_j ;
- 14 **end**
- 15 Assign any remaining BS to the nearest cluster and also the fractional demand served by the node;
- 16 Move all fractionally assigned demands to the corresponding cluster centers and create a single-node instance;
- 17 Solve the single-node instance using greedy packing algorithm in increasing order of $\frac{f_{ij}}{b_i}$, open all fractionally opened or fully opened BS nodes;
- 18 Run the maximum flow algorithm to satisfy as many demands as possible;

Algorithm 4: Modified Capacitated Facility Location Algorithm

The traditional capacitated facility location algorithm tries to first cluster BS nodes around a subset of SS nodes and concentrates traffic demands to the center SS nodes to form single node instances which are then solved by greedily packing low cost facilities. It is easy to see that the original algorithm does not ensure feasibility for single node instances when demands associated with the cluster are concentrated at a cluster center because there may exist BS/SS pairs that are not connected in the same cluster. Moreover,

input : Facility Set BS , Demand Set SS ,
Associated Cost

output: Chosen facilities, An assignment of the
demands x_{ij} ,

- 1 Solve the optimal primal capacitated facility location LP-relaxation and denote the solutions as (x, y) ;
- 2 Pick randomly the BS node with probability y_i , for any $i \in BS$;
- 3 Run the maximum flow algorithm to serve as many demands as possible;

Algorithm 5: Randomized Capacitated Facility Location Algorithm

there is no polynomial algorithm to convert a solution to the facility location instance to a solution to our problem, otherwise, the minimum set cover problem would be able to be approximated with constant factor². Hence, we modified the algorithm mainly in the following aspects:

- The solution of the facility location based algorithm may lead to demands left unsatisfied.
- When solving the linear programming formulation, we let the facility cost be infinite for connecting out-of-reach demand nodes.
- After deciding the set of facilities, we compute the minimum cost assignment of demands to open BS nodes based on the actual connectivity graph, namely, the assignment only happens if the SS node and the targeted BS node are within communication range.

We also consider an intuitive randomized algorithm which solves the LP-relaxation of the primal facility location problem and randomly chooses BS nodes with the probability of the corresponding fractional solution of the linear program relaxation.

We will compare in the following section the performance of these LP-based capacitated facility location algorithm variants and the tradeoffs incurred.

6 Performance Evaluation

Extensive simulation results have been done to compare the performance of each algorithm for different scenarios. Due to space constraints, more results appeared in [9]. In the following, we compared the performance results between our max-flow based algorithm and the exact MIP solution using various network topologies, demands matrices.

²The minimum set cover problem has been known to be not approximable within $(1-\epsilon) \ln n$ for any $\epsilon > 0$ unless $NP \subset DTIME(n^{\log \log n})$.

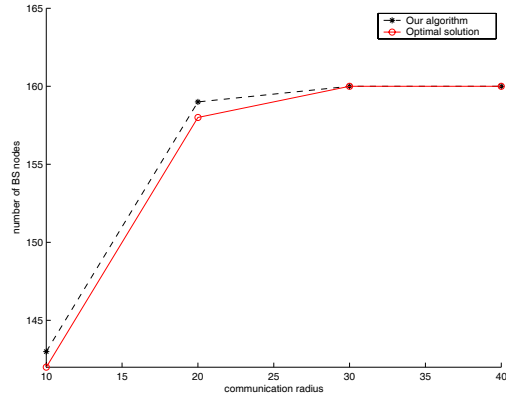


Figure 2. Total number of BS nodes chosen with varying communication radius

We also studied the tradeoffs obtained by facility location based algorithms. In the legends of the following figures, we named the exact solution to the MIP problem as optimal solution, which we computed with CPLEX package. We implemented all algorithms using LEDA packages.

For most simulations, we generated a random network topology by placing N_{BS} BS nodes in a square of 400×400 units. The communication connectivity graph model we used assumed that any BS-SS pair could communicate iff their Euclidean distance was within a communication radius, R . In the following simulations, unless specified otherwise, we adopted a network of 200 BS nodes and 800 SS nodes.

We also investigated the effects of varying communication ranges for individual wireless links. Figure 2 demonstrated that the optimal solution of the PMP-Association problem (200 BS nodes with capacity 10 Mbps and 800 SS nodes with maximum feasible demands) decreased with the increase of communication ranges. As the communication range increased, the effective area covered by a BS node was extended, and thus the number of BS nodes was reduced for the optimal solution. On the other hand, the performance of our algorithm remained almost the same with changing radius, mostly due to the fact that the increasing communication radius led to a closer-to-complete graph, therefore, optimality became harder to achieve, as the relationship between BS and SS sets became more complicated. Moreover, unlike the case in the varying demands simulation (in which BS nodes can potentially cover more SS nodes, but not the other way around), SS nodes in the optimal solution would now have more BS nodes in range, and thus had more options when deciding what BS nodes to choose to further reduce the number of BS nodes needed. This property introduces more difficulties in designing the approximation algorithm.

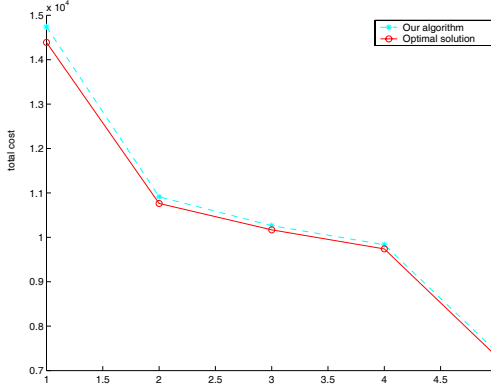


Figure 3. Total network cost with varying number of IAP nodes

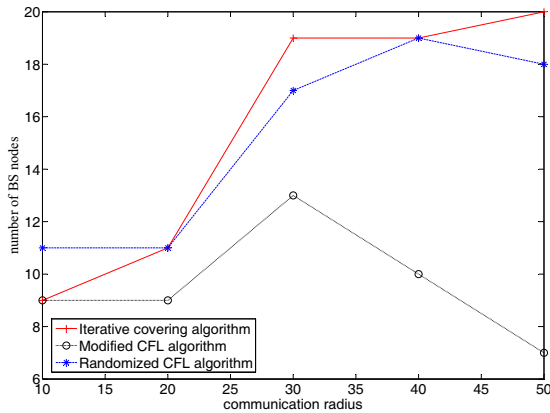


Figure 4. Total BS nodes chosen by the three algorithms

6.1 Effects of Heterogeneous BS Node Costs

In this section, we examined how the BS node cost would affect the final solution. As discussed above, we defined the BS node cost to be the distance between the BS and the nearest IAP node. In this simulation, we chose different IAP numbers (i.e., different number of subnets) and studied how the total cost of all chosen BS nodes changed. The result shown in Figure 3 indicated that the total costs for both our solution and the optimal value decreased with more IAPs, this was because more IAPs in the network potentially allowed that a selected BS be connected to a “closer” IAP and thus decreased its cost.

6.2 Extended Facility Location Algorithms

In this section, we compared the performances between tradeoffs among various algorithms with respect to the communication radius. As discussed above, the modified facility location algorithm and randomized capacitated facility location algorithm could not guarantee all demands satisfied. In Figure 5, We compared the amount of satisfied demands of the three algorithms. The iterative covering algorithm (Algorithm 1) selected base stations till all demands were satisfied. The randomized capacitated facility location algorithm used the linear programming-rounding technique and supported a reasonable amount of demands, though a fraction of the total demands were not supported. More demand requirements were violated for modified facility location algorithm. Moreover, when the communication radius increases, the difference becomes more significant, due to the fact that reducing the original problem to sub-problems in clusters would cause more infeasible assignment subject to the communication range constraints. Figure 4 illustrated the number of base stations needed for these algorithms. Among all three algorithms, the modified capacitated facility location algorithm used the least number of base stations at a cost of most unsatisfied demands, and the number of base stations needed by the other two algorithms are comparable.

Another notable observation from the figures indicated that although the modified capacitated facility location algorithm could not guarantee satisfying all user demands, it presented the best cost-effectiveness (defined as the ratio of satisfied demands and the number of deployed BS nodes) among the three candidates. When other wireless networks were available, such algorithm could be combined with other relaying techniques to provide overall coverage.

7 Conclusion

With the emerging of wireless broadband networks, many challenging issues have arisen in the area of network planning. For large scale wireless networks, maintaining cost-effectiveness is critical in improving the popularity of new technologies. In this paper, we studied and presented algorithms which attempt to find a minimum cost subset of base stations for WiMAX networks, named as the PMP-association Problem, which is shown to be NP-hard. The algorithms we proposed are based on techniques derived from set cover, capacitated facility location and LP-rounding algorithms. We analyzed the performance of our algorithms and compared them, as well as the optimal solution, in different network topologies with various demand densities, transmission ranges, capacities. The numerical results verified that our algorithms could be applied efficiently in de-

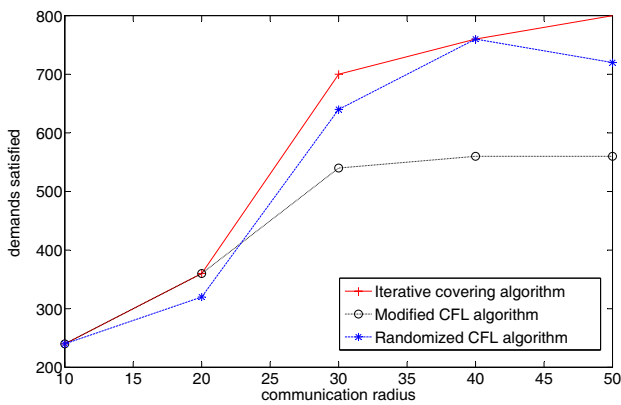


Figure 5. Total demands satisfied by the three algorithms

signing WiMAX access network deployment problems.

8 Acknowledgement

This research is sponsored in part by NSF grant 0435155.

References

- [1] Pankaj K. Agarwal and Cecilia M. Procopiuc. Exact and approximation algorithms for clustering. In *SODA '98: Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, pages 658–667. Society for Industrial and Applied Mathematics, 1998.
- [2] Suman Banerjee and Samir Khuller. A clustering scheme for hierarchical control in multi-hop wireless networks. In *INFOCOM*, pages 1028–1037, 2001.
- [3] Judit Bar-Ilan, Guy Kortsarz, and David Peleg. How to allocate network centers. *Journal of Algorithms*, 15(3):385–415, 1993.
- [4] Yigal Bejerano. Efficient integration of multihop wireless and wired networks with qos constraints. *IEEE/ACM Trans. Netw.*, 12(6):1064–1078, 2004.
- [5] Ranveer Chandra, Lili Qiu, Kamal Jain, and Mohammad Mahdian. Optimizing the placement of internet taps in wireless neighborhood networks. In *ICNP '04: Proceedings of the Network Protocols, 12th IEEE International Conference on (ICNP'04)*, pages 271–282. IEEE Computer Society, 2004.
- [6] WiMAX Forum. Business case models for fixed broadband wireless access based on WiMAX technology and the 802.16 standard. Technical report, 2004.
- [7] David K. Goldenberg, Lili Qiu, Haiyong Xie, Yang Richard Yang, and Yin Zhang. Optimizing cost and performance for multihoming. In *SIGCOMM '04: Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 79–92. ACM Press, 2004.
- [8] Retsef Levi, David B. Shmoys, and Chaitanya Swamy. Lp-based approximation algorithms for capacitated facility location. In *IPCO*, pages 206–218, 2004.
- [9] Peng Lin, Hung Ngo, ChunMing Qiao, Xin Wang, and Ting Wang. Minimum cost wireless broadband overlay network planning. *Research Technical Report*, 2005.
- [10] Robert Lupton, F. Miller Maley, and Neal E. Young. Data collection for the sloan digital sky survey - a network-flow heuristic. *Journal of Algorithms*, 27(2):339–356, 1998.
- [11] N. Megiddo and K. J. Supowit. On the complexity of some common geometric location problems. *SIAM Journal on Computing*, 13(1):182–196, 1984.
- [12] R.Dai, D.O.Stahl, and A.B.Winston. The economics of smart routing and QoS. In *the Fifth International Workshop on Networked Group Communications (NGC'03)*, 2003.
- [13] the working group for IEEE802.16 standards. IEEE P802.16-REVd/D5, Draft IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems. Technical report, IEEE, 2004.
- [14] Vijay V. Vazirani. *Approximation Algorithms*. springer-verlag, 2003.